Association in Contingency Tables
An Informationtheoretic Approach

Von der Fakultät für Mathematik, Informatik und Naturwissenschaften der RWTH Aachen University zur Erlangung des akademischen Grades eines Doktors der Naturwissenschaften genehmigte Dissertation

vorgelegt von

Master of Science
Michael Espendiller

aus

Leonberg

Berichter: Univ.-Prof. Dr. Maria Kateri
Univ.-Prof. Dr. rer. nat. Udo Kamps

Tag der mündlichen Prüfung: 24. Februar 2017

Mathematics is the classification and study of all possible patterns.

Walter Warwick Sawyer

Acknowledgement

Special thanks goes to all the colleagues of the institute of statistics at the RWTH Aachen University for the support and the most interesting scientific discussions.

Especially, I thank Julian Górny and Benjamin Laumen for all the hints improving my LaTeX-files and for the general Ph.D. students talks. Christian Staerk and Christian Kohl have my gratitude for the many useful scientific talks and comments.

I thank Dr. Wolfgang Herff and his high level of self-organization and precision, which I could – partly – inherit. His suggestions for literature were always helpful and he always had an ear for open problems.

I would like to thank Mia Müller-Platz as my office colleague for all the debates about statistical science and her always open and critical second opinion. I am convinced that her work in Item Response Theory will be a valuable part of the statistical science with great practical impact.

Prof. Dr. Udo Kamps has my gratitude for the fast referring of this work and his interesting lectures in insurance mathematics.

I foremost want to thank my advisor Prof. Dr. Maria Kateri. It has been an honour to be her first Ph.D. student at her Chair of Statistics and Stochastic Modelling in Germany. She has introduced me to Categorical Data Analysis and related fields and taught me new ways within the statistical science. I appreciate her always open door as well as her investment of time and ideas during her supervision.

Thanks for all the advice and things I could learn. I appreciate all the cherry years.

Michael Espendiller
RWTH Aachen University
February 2017

Contents

Introduction

Categorical Data Analysis (CDA) is part of the statistical science dealing with categorical response variables, i.e. such variables that can be valued with a finite number of possible outcomes (categories). Such data occur naturally in many scientific fields such as social sciences, economics, and biomedical sciences.

Categorical data and their analysis was one of the very first emphasis of statistical science with pioneering works by Pearson [1900, 1904, 1922], Fisher [1922] and Yule [1900, 1903, 1912]. The standard reference in the late 20th century was the book *Discrete Multivariate Analysis* by Bishop, Fienberg and Holland from 1975 (with reprint Bishop et al. [1975]). Agresti's book *Categorical Data Analysis* from 1996 (in its third edition Agresti [2013]) is one of the main references. Although this topic is very classical, it is remarkable that elementary issues in CDA, like the confidence interval (CI) of binomial success probabilities, do still attract the interest of researchers (cf. Brown et al. [2001] and the discussion of this paper).

This Ph.D. thesis deals with one of the fundamental problems of CDA, namely association between categorical variables and its measurement. When several variables are involved, the connections and dependencies between them is of interest. A sensitive analysis of the data can be provided by applying models. Although models give a broad tool box and high flexibility for data analysis, their application can be rather complex for practitioners, requiring model fitting procedures and advanced understanding of statistical software. Association measures provide a more compact identification and overall quantification of underlying association. They are easy to understand and interpret. Therefore such measures are an important tool in CDA. As one of the most fundamental questions, association measurement was a focus of the above referred literature in the first part of the 20th century.

The most basic tools of categorical data analysis, which are linked to measuring association, are introduced in Chapter 1. Two dichotomous variables are cross-classified in a 2×2 contingency table. The very first and still up to date measure of association for this data class, introduced by Yule [1900], is the odds ratio. The sample size in 2×2 tables can often be small. Consider for example a medical case-control study of a rare disease. In instances with small sample sizes, the asymptotic Wald confidence interval for the odds ratio can lead to degenerate coverage probabilities (Agresti [1999]). Inferential problems are also caused by the presence of sampling zeros. In this case, the log-odds ratio estimates can be infinite, while their asymptotic variance is also infinite.

In Chapter 2 the standard approach of adding a continuity correction to overcome these problems is revisited and studied under a general set-up. The predominant type of correction, which adds a value of $c = 0.5$ to each cell (Anscombe [1956], Haldane [1956], Gart [1962]), is not optimal, since it does not always solve the problem of degenerate coverage probabilities of the Wald confidence intervals for the odds ratios. Alternative CIs can be obtained by inverting tests based on the score, likelihood ratio or Cressie-Read (Cressie and Read [1984]) statistics. Inverting the score test instead of the Wald test could already improve the quality of the binomial confidence intervals (Brown et al. [2001]) and therefore, such an approach is promising for the odds ratios as well. As far as the author knows, no extensive evaluation study in terms of coverage probability and mean length has been done so far comparing the different confidence intervals and correction techniques applied to the odds ratio. Such an analysis is performed in Chapter 2 and the behaviour of various alternative approaches is comparatively commented.

Another approach to solve the problem of infinite estimates of the odds ratios can be obtained based on the Information Theory introduced by Shannon [1948] using entropies and divergences. Most statistical tools (test, models, estimation, etc.) are based on an underlying concept of divergence between observed sample proportions and the estimated probabilities under the assumed structure. For example, the maximum likelihood estimator (MLE) for categorical random variables can also be obtained by minimizing the Kullback-Leibler divergence between the observed and expected frequencies under the assumed model. In addition, the association model for contingency tables is under certain conditions the closest model to the independence model in terms of the Kullback-Leibler divergence. An important family of divergences is the ϕ-divergence, which was independently introduced by Csiszár [1963] and Ali and Silvey [1966]. Using the ϕ-divergence instead of the Kullback-Leibler divergence, the MLEs, statistical models and tests can be generalised. The basic background on ϕ-divergence and its application to CDA is also introduced in Chapter 1.

Under an information theoretic set-up, the odds ratio is connected to the Kullback-Leibler divergence. Considering the ϕ-divergence, alternative association measures are derived for 2×2 contingency tables in Chapter 3. Their properties are studied and asymptotic inference is developed. For some members of this family the estimated association measures remain finite in the presence of a sampling zero. For a subset of these members the estimators of these measures have finite variance as well, making continuity correction dispensable in such cases. Special attention is given to the power divergence, which is a parametric family. The role of its parameter λ, regarding the asymptotic confidence interval's coverage probability and average relative length is further discussed. In special probability table structures, for which the performance of the asymptotic Wald confidence intervals for the classical log-odds ratio is poor, the measure corresponding to $\lambda = 1/3$ is suggested as an alternative.

When two cross-classified categorical variables have I and J categories, respectively, the data are summarized in an $I \times J$ contingency table. The classical log-odds ratio can be generalised to $I \times J$ tables in different ways. This can be done by considering 2×2 local subtables or 2×2 tables formed by merging categories if the corresponding classification variable is ordinal. The association structure of an $I \times J$ contingency table is completely captured by a minimal set of such $(I-1)(J-1)$ generalised odds ratios. Often, the association is described by a scalar measure. There exist plenty of such measures like Phi, Cramér's V, Goodman and Kruskal's γ or Pearson's Contingency Coefficient. Such measures are a simple tool for providing information on the strength (and direction) of the underlying association in a direct and simple way. Even though these measures are a simple and easily interpretable tool to present the association structure, they can be misleading when the association structure is more complex and cannot be exactly described with a single parameter.

Models are more accurate but also more complex compared to scalar measures and are therefore able to approach more complex association structures. Generalised odds ratios have been connected to parameters of models and facilitated their interpretation. For example, local odds ratios, which are defined using adjacent cells forming 2×2 tables, there are directly connected to the interpretation parameters of loglinear models. Furthermore, the simplest and most parsimonious association model for two-way tables, the uniform association model, assumes a common value for all local odds ratios of the table. This common value can be used as a scalar association measure. Less parsimonious association models impose other types of constraints on the local odds ratios. For example, the row-effect association model, applicable when the column classification variable is ordinal, assumes the same local odds ratio between adjacent rows. This row odds ratio values can be regarded as multidimensional measure of association. These models can be extended to *generalised association models* by setting the corresponding restrictions on other types of generalised odds ratios than the local ones. In Chapter 4, this close connection between generalised association models and generalised odds ratios is used to introduce new measures of association, that can approach more complex association structures. These model-based measures can be estimated by MLE methods based on the estimation procedure of the corresponding model, which requires iterative algorithms. Clayton [1974] and Beh and Davy [2004] introduced closed-form estimators as alternative to MLE for the uniform and linear-by-linear association model, respectively. In this thesis, these non-iterative estimators are extended for generalised association models. Selected examples are analysed in detail while the proposed closed-form estimators are compared to the MLE via an extensive simulation study.

The generalised association models can be extended via the ϕ-divergence. Kateri and Papaioannou [1995] introduced the ϕ-scaled local odds ratios and linked them to the parameters of the ϕ-association model, which for the Kullback-Leibler divergence reduce to the local log-odds ratios and the standard association models. The other types of generalised odds ratios for two-way tables (see Chapter 1) are analogously extended to the generalised ϕ-scaled odds ratios in Chapter 5. Their asymptotic behaviour, connections to classical dependence concepts, and other properties are analysed. It is shown that under certain conditions the new measures uniquely define the joint distribution when marginals are known and thus encode the association within the data.

Model-based measures can capture more complex association structures without complicating the interpretability. Additionally, non-iterative estimations provide a good alternative, which gives values close to MLE. In Chapter 6, the generalized ϕ-scaled odds ratios (Chapter 5) are used to define new generalised ϕ-scaled association models and corresponding multidimensional model-based ϕ-scaled measures of association. Changing the scale on which association is measured can lead to more adequate association measures. The effect of such a scale change is analysed for the power divergence. Closed-form estimators of ϕ-scaled measures based on ϕ-scaled generalised uniform or row-effect association models are analysed and compared to MLE. Based on simulation studies, suggestions for application of the newly presented methods are given.

In the final Chapter 7, square tables with commensurable classification variables are discussed. Such special structures of contingency tables are of interest e.g. in social sciences to consider the social mobility of a population to value the permeability of educational and economical systems. Such tables can be analysed with models of (quasi-)symmetry and marginal homogeneity. The tests of McNemar [1947] for 2×2 tables and its generalisation to $I \times I$ tables (Bowker [1948]) are classical approaches to test for symmetry. Using the ϕ-divergence, one can define a family of directed asymmetry measures arising from ϕ-quasi-symmetry models (Kateri and Agresti [2006]; Kateri and Papaioannou [1997]), which then lead to ϕ-scaled tests for symmetry, the ϕ-McNemar and ϕ-Bowker tests. For the Pearson divergence, the ϕ-based symmetry tests coincide with the ordinary McNemar and Bowker test. An evaluation study regarding type I error rates based on the power divergence is conducted for the new ϕ-McNemar test, which shows that the ordinary McNemar test is the best test with respect to type I error rates in the power divergence family.

This thesis introduces new measures and models for two-way contingency tables based on the ϕ-divergence. A scale change for the standard measure of association in 2×2 tables, the log-odds ratio, can improve the compatibility of the measure with sampling zeros and can – in some set-ups – increase the quality of the corresponding Wald confidence intervals regarding coverage probability and average relative length. New non-scalar measures of association are introduced, which are directly linked to the model parameters of generalised association models. Closed-form estimators for these measures are provided and extensively studied on log- and ϕ-scale. In most cases they provide a good estimate, very close to the MLE. The generalised ϕ-linear model is introduced during this process. The possible scale change in generalized association models increases their flexibility in capturing association structures. In the special case of $I \times I$ table with commensurable classification variables, a multidimensional asymmetry measure on ϕ-scale is introduced, which is linked to the corresponding ϕ-scaled (quasi-)symmetry model. Generalized ϕ-scaled versions for the symmetry tests of McNemar and Bowker are introduced and studied.

The structure of this thesis is sketched in Figure 1, where the chapters and their interrelations are listed.

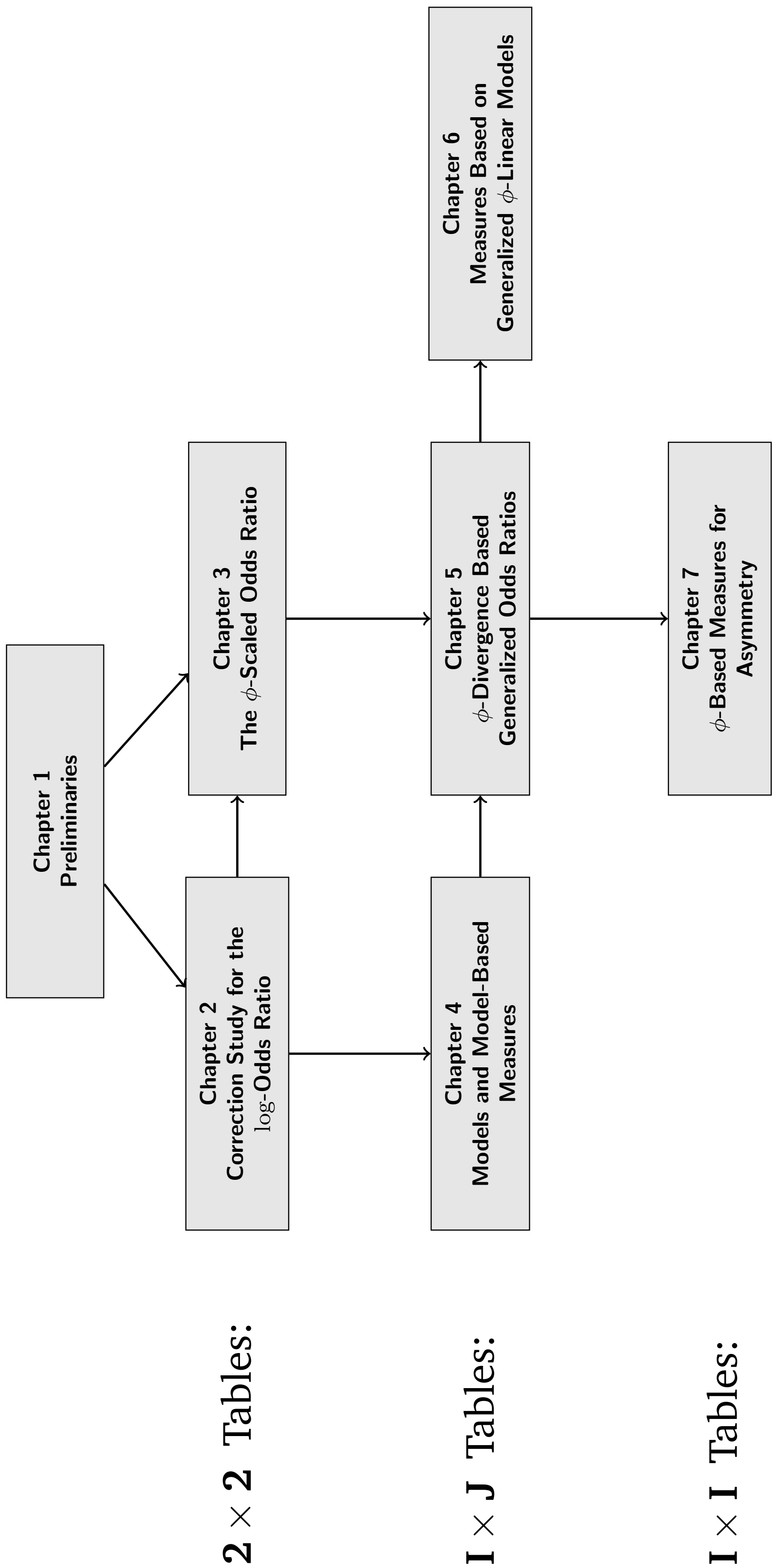


Figure 1: Chapter Structure

Chapter 1

Preliminaries

The basic preliminaries required for this work are presented. After giving the fundamental definitions for categorical data variables in Section 1.1 the notations used during this work are introduced. In Section 1.2 the Poisson, multinomial, and product multinomial distributions used as underlying random mechanisms in contingency tables are presented. Independence between two categorical classification variables is discussed in Section 1.3. MLEs and basic asymptotic results for multinomial random variables are presented in Section 1.4. In Section 1.5 goodness-of-fit tests for contingency tables are introduced. Measures of association are debated in Section 1.6, while the most important measure in 2×2 contingency tables, the odds ratio, is reviewed in Section 1.7. Additionally, the relative risk and its properties are briefly summarized in Section 1.8, before presenting the gneralized odds ratios for $I \times J$ tables in Section 1.9. Categorical data models like loglinear models, association models, HLP models and symmetry models are also introduced and their connection to generalised odds ratios is also discussed. To generalise known tools in categorical data analysis, this work uses an information-theoretic approach based on the ϕ-divergence introduced in Section 1.10. Well-known members of the ϕ-divergence family are the Pearson, Kullback-Leibler, and power divergences. In Section 1.11 models based on ϕ-divergence are presented, which are closely connected to the new information-theoretic measures studied in this work.

1.1 Basics

A random variable which can be measured on a scale consisting of a set of categories is called *categorical*. Categorical variables are measured mostly on nominal or ordinal scales. If the categories inherit a natural ordering, they are called *ordinal*, otherwise they are called *nominal*. By ignoring the ordering, every ordinal variable can be analysed by methods for nominal variables. Binary variables have only two outcome categories and can be viewed as nominal or ordinal. Much attention is paid to the study of interaction between variables: If a variable Y is influenced by a variable X but not vice versa, Y is called *response* while X is called *explanatory*. If the influence is not one-sided, X and Y are considered both as responses and treated in a symmetric manner. If all included variables are categorical, the data can be represented in a contingency table (Section

1.2). The term contingency table was first introduced by Pearson [1904]. Representative references for categorical data analysis are Agresti [2007, 2013], Bishop et al. [1975] and Tutz [2000]. For a focus on ordinal data see Agresti [2010].

A matrix with $I \in \mathbb{N}$ rows and $J \in \mathbb{N}$ columns with elements $n_{ij}, i = 1, ..., I, j = 1, ..., J$ is denoted $[n_{ij}] \in \mathbb{R}^{I \times J}$ where superscript T denotes its transpose, $[n_{ij}]^T = [n_{ji}] \in \mathbb{R}^{J \times I}$. For two matrices $[a_{ij}], [b_{ij}] \in \mathbb{R}^{I \times J}$ and a constant $k \in \mathbb{R}$ define the sum and scalar multiplication as component-wise operations, i.e. the matrix $[c_{ij}] = k[a_{ij}] + [b_{ij}]$ is the matrix with components $c_{ij} = ka_{ij} + b_{ij}$. The row and column marginals of the matrix are denoted by $n_{i+} = \sum_{j=1}^{J} n_{ij}$ and $n_{+j} = \sum_{i=1}^{I} n_{ij}$, respectively. In the following, a matrix $[n_{ij}]$ will be expanded row-wise in vector form $\mathbf{n} = (n_{ij}) = (n_{11}, ..., n_{1J}, n_{21}, ..., n_{IJ})$. Let $\Delta_{I,J} \subset \mathbb{R}^{IJ}$ be the open $I \cdot J$-simplex defined by

$$\Delta_{I,J} := \{\boldsymbol{\pi} = (\pi_{ij}) \in (0,1)^{IJ} \mid \sum_{i,j=1}^{I,J} \pi_{ij} = 1\}.$$

Let $\Delta_{I \times J} := \{\boldsymbol{\pi} = [\pi_{ij}] \in (0,1)^{I \times J} \mid \sum_{i,j=1}^{I,J} \pi_{ij} = 1\} \subset \mathbb{R}^{I \times J}$ denote the open $I \cdot J$-simplex with elements in matrix form.

The elements of Δ_{IJ} (and $\Delta_{I \times J}$) correspond to discrete non-degenerate finite probability vectors (matrices), i.e. the elemental probabilities are non-zero.

1.2 Contingency Tables

Let X and Y be two categorical variables with categories $\{1, ..., I\}$ and $\{1, ..., J\}$, respectively. A contingency table $\mathbf{n} = (n_{ij})$ is a realisation of a random variable $\mathbf{N} = (N_{ij})$, where the elements N_{ij} count the number of outcomes $(X, Y) = (i, j)$. The underlying distribution of $\mathbf{N}$ is called sampling scheme. The three most common sampling schemes are the Poisson, the multinomial, and the product multinomial distribution.

Let $\boldsymbol{\gamma} = (\gamma_{ij})$ be a vector of real numbers $\gamma_{ij} > 0$ and assume, that $N_{ij} \sim \mathcal{P}(\gamma_{ij}), i = 1, ..., I, j = 1, ..., J$ are independent Poisson distributed random variables. For $\mathbf{n} = (n_{ij}) \in \mathbb{N}_0^{IJ}$, the probability mass function of the *Poisson sampling scheme* is

$$p(\mathbf{n}) = \prod_{i,j=1}^{I,J} \frac{\exp(-\gamma_{ij})\gamma_{ij}^{n_{ij}}}{n_{ij}!}. \tag{1.1}$$

A Poisson distributed random variable is denoted by $\mathbf{N} \sim \mathcal{P}(\boldsymbol{\gamma})$. It holds $\mathbb{E}(N_{ij}) = \gamma_{ij} = \mathbb{V}\text{ar }(N_{ij})$ and the maximum likelihood estimator (MLE) of $\boldsymbol{\gamma}$ is $\hat{\boldsymbol{\gamma}} = \mathbf{N}$. In the Poisson sampling scheme, the total sample size $N = \sum_{ij=1}^{I,J} N_{ij}$ of the contingency table is not fixed. Hence, it is the underlying sampling scheme in studies before the sample size is known.

Often, the study design fixes the total sample size to $N = n \in \mathbb{N}$. Conditioning on the total sample size, the Poisson sampling scheme becomes equivalent to the multinomial sampling scheme. Let $\mathbf{N} \sim \mathcal{P}(\boldsymbol{\gamma})$ follow the Poisson sampling scheme with probability mass function (1.1) and let $\mathbf{n} = (n_{ij})$ be a realisation with $\sum_{i,j=1}^{I,J} n_{ij} = n$. Then using the fact that $\sum_{i,j=1}^{I,J} N_{ij} \sim \mathcal{P}(\sum_{i,j=1}^{I,J} \gamma_{ij})$ it holds

$$\mathbf{P}\left(\mathbf{N} = \mathbf{n} \,\middle|\, \sum_{i,j=1}^{I,J} N_{ij} = n\right) = \frac{\mathbf{P}(\mathbf{N} = \mathbf{n})}{\mathbf{P}\left(\sum_{i,j=1}^{I,J} N_{ij} = n\right)} = \frac{n!}{\prod_{i,j=1}^{I,J} n_{ij}!} \prod_{i,j=1}^{I,J} \pi_{ij}^{n_{ij}},$$

where $\pi_{ij} := \frac{\gamma_{ij}}{\sum_{i,j=1} \gamma_{ij}}$, which is the probability mass function of the *multinomial sampling scheme*. Since all $\gamma_{ij} > 0, i = 1, ..., I, j = 1, ..., J$ it holds $\boldsymbol{\pi} = (\pi_{ij}) \in \delta_{IJ}$. For a realization $\mathbf{n} = (n_{ij})$ with $\sum_{,j=1}^{I,J} n_{ij} = n$, a multinomial distributed random variable with sample size n and multinomial probability vector $\boldsymbol{\pi}$ is denoted by $\mathbf{N} \sim \mathcal{M}(n, \boldsymbol{\pi})$. It holds for the expected value $\mathbb{E}\mathbf{N} = n\boldsymbol{\pi}$ and for the variance-covariance matrix $\mathbb{V}\text{ar } \mathbf{N} = n(\text{diag}(\boldsymbol{\pi}) - \boldsymbol{\pi}^T\boldsymbol{\pi})$, where $\text{diag}(\boldsymbol{\pi})$ denotes the diagonal matrix with entries (π_{ij}). The MLE of $\boldsymbol{\pi}$ is $\hat{\boldsymbol{\pi}} = \mathbf{N}/n$.

When the variable X is explanatory the row marginal counts $(n_{1+}, ..., n_{I+})$ are often fixed by design. In this case, the underlying sampling scheme is based on independent multinomials in each row (product multinomial), i.e. $\mathbf{N}_i = (N_{i1}, ..., N_{iJ}) \sim \mathcal{M}(n, \boldsymbol{\pi}_i)$ with $\boldsymbol{\pi}_i = (\pi_{1|i}, ..., \pi_{J|i})$ the probability vector for the i-th row. It is straightforward to verify that the multinomial sampling scheme $\mathbf{N} = (N_{ij}) \sim \mathcal{M}(n, \boldsymbol{\pi})$ by conditioning on the row marginal counts is equivalent to the product multinomial one with $\boldsymbol{\pi}_i = (\pi_{1|i}, ..., \pi_{J|i})$, where $\pi_{j|i} = \pi_{ij}/\pi_{i+}, i = 1, ..., I, j = 1, ..., J$

The MLE of $\boldsymbol{\pi}_i$ in the product multinomial sampling scheme is $\hat{\boldsymbol{\pi}}_i = (N_{i1}, ..., N_{iJ})/n_{i+}, i = 1, ..., I$. Notice that for the Poisson (conditional on n), multinomial and product multinomial sampling schemes share the same kernel of the likelihood function and hence lead to the same MLEs. If the number of categories in the column classification variable is two ($J = 2$), the product multinomial sampling scheme reduces to the product binomial sampling scheme $N_{i1} \sim \mathcal{B}(n_i, \pi_{1|i}), N_{i2} = n_{i+} - N_{i1}, i = 1, ..., I$. In this case, $\pi_i = \pi_{1|i}, i = 1, ..., I$, can be used as a simplified notation.

The underlying random mechanism has to be chosen depending on the study design. For example, a case-control study using an explanatory variable with two categories ($I = 2$) fixes the number of cases (n_{1+}) and the number if controls (n_{2+}), which is modelled using the product multinomial sampling scheme ($J > 2$) or product binomial sampling scheme ($J = 2$), respectively.

Up to now, only contingency tables with two classification variables X and Y (two-way tables) have been considered. If a data set consists of more categorical variables, they can be cross-classified to form a multi-way table. For example, a three-way contingency table for variables X, Y and Z with I, J and K categories, respectively, is an element $(n_{ijk}) \in \mathbb{N}_0^{IJK}$. The previously presented sampling schemes and their relations easily expand to multi-way tables.

Next, the following two data sets will be used as examples throughout this thesis:

Example 1.2.1. The larynx data (Table 1.1, (a)) refers to a retrospective cohort study on treating cancer of the larynx and cross-classifies the binary response of controlling the illness for at least two years after the treatment (radiation therapy or surgery) vs. the not controlling it after two years [Mendenhall et al., 1984].

Example 1.2.2. The prednisolone data (Table 1.1, (b)) summarizes the effect of prednisolone on severe hypercalcaemia in women with metastatic breast cancer by comparing a treatment to a placebo group. The effect is measured with respect to the normalisation of their serum-ionized calcium level [Kristensen et al., 1992].

Table 1.1: Data sets for Example 1.2.1 and 1.2.2.

(a) Larynx data

	Cancer controlled	
	yes	no
Surgery	21	2
Radiation	15	3

(b) Prednisolone data

	Normalisation	
	yes	no
Prednsl.	7	8
Placebo	0	15

In both data sets, the row marginal counts were fixed by experimental design: In the larynx cancer study, 23 patients were assigned to receive the surgical intervention, and 18 to the radiation therapy; in the prednisolone study, 15 patients were assigned to each of the treatment and the placebo group. Thus, the row classification variables can be regarded as explanatory variables and the underlying sampling scheme is therefore the product binomial sampling scheme.

1.3 Independence in Two-Way Contingency Tables

The simplest form of "association" between two categorical variables X and Y is independence, that is variable X has no influence on the outcome of variable Y and vice versa. Mathematically, X and Y are independent when

$$\mathbf{P}(X = i, Y = j) = \mathbf{P}(X = i)\,\mathbf{P}(Y = j), \qquad i = 1, ..., I, j = 1, ..., J.$$

In terms of the multinomial probability vector $\boldsymbol{\pi} = (\pi_{ij}) \in \Delta_{I,J}$, independence is expressed by

$$\pi_{ij} = \pi_{i+}\pi_{+j}, \qquad i = 1, ..., I, j = 1, ..., J. \tag{1.2}$$

If Y is a response variable and X is explanatory, independence can be expressed in terms of the conditional row probabilities of the multinomial probability matrix $\boldsymbol{\pi} = [\pi_{ij}]$ as

$$\mathbf{P}(Y = j \mid X = i) = \mathbf{P}(Y = j), \qquad i = 1, ..., I, j = 1, ..., J, \tag{1.3}$$

which becomes $\pi_{j|i} = \pi_{+j}, i = 1, ..., I, j = 1, ..., J$ in terms of the product multinomial probabilities $\boldsymbol{\pi}_i = (\pi_{1|i}, ..., \pi_{J|i}), i = 1, ..., I$. Therefore, under independence $\pi_{j|1} = ... = \pi_{j|I}$ for all $j = 1, ..., J$, so that the probability structure in every row has the same profile. This is called *homogeneity* of the conditional distributions. Test for independence in two-way contingency tables will be introduced in Section 1.5.

1.4 Estimation and Asymptotics

In this section standard results for estimation and asymptotic behaviour for multinomial random variables will be presented.

Lemma 1.4.1. Let $\mathbf{N} \sim \mathcal{M}(n, \boldsymbol{\pi})$ be multinomial distributed with sample size $n \in \mathbb{N}$ and multinomial probability vector $\boldsymbol{\pi} \in \Delta_{I,J}$. Let $\hat{\boldsymbol{\pi}} = \mathbf{N}/n$ be the MLE of $\boldsymbol{\pi}$. Then $\hat{\boldsymbol{\pi}} \xrightarrow{P} \boldsymbol{\pi}$ for $n \to \infty$.

Lemma 1.4.1 holds for MLEs, as long as they fulfil special regularity conditions (cf. Cox and Hinkley [1974, Section 9.2, pp. 287]), which are fulfilled for multinomial distributed random variables. In addition, using the central limit theorem [Rao, 1973, p. 128], $\hat{\boldsymbol{\pi}}$ is asymptotic normal distributed:

Lemma 1.4.2. (cf. [Agresti, 2013, p. 590, Section 16.1])

For $n \in \mathbb{N}$ and $\boldsymbol{\pi} = (\pi_{ij}) \in \Delta_{I,J}$, let $\mathbf{N} \sim \mathcal{M}(n, \boldsymbol{\pi})$ be multinomial distributed and let $\hat{\boldsymbol{\pi}} = \mathbf{N}/n$ be the MLE of $\boldsymbol{\pi}$. It holds

$$\sqrt{n}(\hat{\boldsymbol{\pi}} - \boldsymbol{\pi}) \xrightarrow{d} \mathcal{N}(0, \operatorname{diag}(\boldsymbol{\pi}) - \boldsymbol{\pi}^T\boldsymbol{\pi}) \qquad \text{for } n \to \infty. \tag{1.4}$$

Throughout this work, asymptotic distributions of random variables of the form $g(\hat{\boldsymbol{\pi}})$ for a function $g : \mathbb{R}^{IJ} \to \mathbb{R}$ are derived by the delta method.

Theorem 1.4.3. (cf. Agresti [2013, pp. 587, Section 16.1])

Let $(\mathbf{X}_n)_n$ be a sequence of multivariate random vectors that is asymptotically normal distributed with mean $\boldsymbol{\mu} = (\mu_1, ..., \mu_J) \in \mathbb{R}^J$ and covariance matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{J\times J}$, e.g.

$$\sqrt{n}(\mathbf{X}_n - \boldsymbol{\mu}) \xrightarrow{d} \mathcal{N}(0, \boldsymbol{\Sigma}) \qquad \text{for } n \to \infty.$$

Let $g : \mathbb{R}^J \to \mathbb{R}^q, \mathbf{t} = (t_1, ..., t_J) \in \mathbb{R}^J$ be a vector function such that all partial derivatives exist and are continuous. Let $\mathbf{D}$ be the Jacobian matrix at point $\boldsymbol{\mu}$:

$$\mathbf{D} := \frac{\partial g}{\partial \boldsymbol{\mu}} := \frac{\partial g(\mathbf{t})}{\partial \mathbf{t}}|_{\mathbf{t}=\boldsymbol{\mu}} = \left(\frac{\partial g_i(\mathbf{t})}{\partial t_j}|_{\mathbf{t}=\boldsymbol{\mu}}\right)_{i=1,...,q,j=1,...,J}.$$

If $\mathbf{D} \neq \mathbf{0}$, the sequence of function values $(g(\mathbf{X}_n))_n$ inherits an asymptotic normal distribution:

$$\sqrt{n}(g(\mathbf{X}_n) - g(\boldsymbol{\mu})) \xrightarrow{d} \mathcal{N}(0, \mathbf{D}^T\boldsymbol{\Sigma}\mathbf{D}) \qquad \text{for } n \to \infty. \tag{1.5}$$

As a direct consequence of Lemma 1.4.2 and Theorem 1.4.3, the asymptotic distribution of functions of multinomials can be calculated:

Corollary 1.4.4. Let $\boldsymbol{\pi} = (\pi_{ij}) \in \Delta_{I,J}$, $\mathbf{N} \sim \mathcal{M}(n, \boldsymbol{\pi})$ and let $\hat{\boldsymbol{\pi}} = \mathbf{N}/n$ be the MLE of $\boldsymbol{\pi}$. Let $g : \mathbb{R}^{IJ} \to \mathbb{R}, \boldsymbol{\pi} \mapsto g(\boldsymbol{\pi})$ be a continuously differentiable function such that $\mathbf{D} = \frac{\partial g}{\partial \boldsymbol{\pi}} \neq \mathbf{0}$. Then $g(\hat{\boldsymbol{\pi}})$ obtains an asymptotic normal distribution with mean $g(\boldsymbol{\pi})$ and variance

$$\mathbb{V}\text{ar } g(\hat{\boldsymbol{\pi}}) = \sum_{i,j=1}^{I,J} \pi_{ij} \left(\frac{\partial g(\boldsymbol{\pi})}{\partial \pi_{ij}}\right)^2 - \left(\sum_{i,j=1}^{I,J} \pi_{ij} \frac{\partial g(\boldsymbol{\pi})}{\partial \pi_{ij}}\right)^2.$$

There clearly is an extension of Corollary 1.4.4 for vector-valued functions $g : \mathbb{R}^{IJ} \to \mathbb{R}^k, \boldsymbol{\pi} \mapsto g(\boldsymbol{\pi})$, where the asymptotic variance becomes $\mathbf{D}^T\boldsymbol{\Sigma}\mathbf{D}$ following Theorem 1.4.3.

1.5 Goodness-of-fit Tests

The significance of statistical models is evaluated by goodness-of-fit tests. The associated test statistics measure the divergence of the expected values under the model to the observed values. The standard asymptotic significance tests are the Wald, score and likelihood ratio test (see [Kateri, 2014, p. 9]), which will be presented next.

Let $\varphi \in \mathcal{P} \subseteq \mathbb{R}$ be a scalar parameter of a parametric family of distributions with MLE $\hat{\varphi}$ and standard error estimated $SE(\hat{\varphi})$. Let $\varphi_0 \in \mathcal{P}$. Wald [1943] introduced a test for testing $H_0 : \varphi = \varphi_0$ against $H_1 : \varphi \neq \varphi_0$, the *Wald test*. Its test statistic has the form

$$W = \frac{\hat{\varphi} - \varphi_0}{SE(\hat{\varphi})} \tag{1.6}$$

and is asymptotic standard normal distributed under H_0 given certain regularity conditions (cf. Serfling [1980, p. 144]), which are satisfied in the set-up of this work. Let L be the likelihood function. Another test, the *score test*, for testing H_0 against H_1 was introduced by Rao [1948]. Let $u(\varphi_0) = \partial \log L(\varphi)/\partial\varphi|_{\varphi=\varphi_0}$ be the *score function*. The *score test statistic* is defined by

$$S^2 = \left(\frac{u(\varphi_0)}{SE(u(\varphi_0))}\right)^2 = \frac{(\partial \log L(\varphi)/\partial\varphi|_{\varphi=\varphi_0})^2}{-\mathbb{E}(\partial^2 \log L(\varphi)/\partial^2\varphi|_{\varphi=\varphi_0})}, \tag{1.7}$$

and is asymptotic χ^2_1 distributed under H_0. The third very popular test statistic is the *Likelihood Ratio* (LR) test, given by

$$G^2 = 2(\log L(\hat{\varphi}) - \log L(\varphi_0)), \tag{1.8}$$

which is again asymptotic χ^2_1 distributed (Wilks [1938]). Therefore, all test statistic W^2, S^2 and G^2 from (1.6) to (1.8) are asymptotic equivalent. The previous test statistics for vector parameters $\boldsymbol{\varphi} = (\varphi_1, ..., \varphi_K) \in \mathbb{R}^K$ are relevant in the context of models for contingency tables and are presented next.

Let $\boldsymbol{\pi}, \boldsymbol{\pi}_0 \in \Delta_{I,J}$ be two multinomial probability vectors for an $I \times J$ table. Let $\mathbf{N} = (N_{ij}) \sim \mathcal{M}(n, \boldsymbol{\pi})$ be a multinomial distributed random variable and consider the null hypothesis

$$H_0 : \boldsymbol{\pi} = \boldsymbol{\pi}_0. \tag{1.9}$$

Often, the probability $\boldsymbol{\pi}_0$ in the null hypothesis (1.9) is not fixed, but depends on a parameter vector of size $K < IJ - 1$:

$$\boldsymbol{\pi}_0 = \boldsymbol{\pi}_0(\boldsymbol{\varphi}), \qquad \boldsymbol{\varphi} = (\varphi_1, ..., \varphi_K) \in \mathbb{R}^K.$$

The MLE $\hat{\boldsymbol{\varphi}}$ of $\boldsymbol{\varphi}$ determines the MLE $\hat{\boldsymbol{\pi}}_0 = \boldsymbol{\pi}_0(\hat{\boldsymbol{\varphi}})$ by the invariance property of MLEs (cf. Casella and Berger [2002, Theorem 7.2.10, p. 320]), and hence the expected cell frequencies under H_0 are $\hat{\boldsymbol{\mu}} = \mathbb{E}\mathbf{N} = n\hat{\boldsymbol{\pi}}_0$. The very first test statistic to test goodness-of-fit for model (1.9) was introduced by Pearson [1900]. *Pearson's X^2 test statistic* is defined by

$$X^2 = \sum_{i,j=1}^{I,J} \frac{(n_{ij} - \hat{\mu}_{ij})^2}{\hat{\mu}_{ij}}, \qquad \hat{\boldsymbol{\mu}} = (\hat{\mu}_{ij}), \tag{1.10}$$

which then is asymptotic χ^2_{df} distributed with $df = IJ-1-K$ degrees of freedom. Notice that the model $\boldsymbol{\pi} = \boldsymbol{\pi}(\varphi)$ needs to fulfil special requirements (cf. Agresti [2013, p. 592, Section 16.2]). In the case of the independence hypothesis (1.2),

$$\varphi = (\pi_{1+}, ..., \pi_{I-1+}, \pi_{+1}, ..., \pi_{+J-1}), \qquad K = (I-1) + (J-1)$$

with $\hat{\mu}_{ij} = n_{i+}n_{+j}/n$ and (1.10) is χ^2_{df} distributed with $df = IJ - 1 - K = (I-1)(J-1)$ degrees of freedom. Independence is rejected with significance level $\alpha \in (0,1)$ if $X^2 > \chi^2_{df,\alpha}$, where $\chi^2_{df,\alpha}$ is the $(1-\alpha)$-quantile of a χ^2_{df} distribution.

Notice that Pearson's X^2 is the score statistic (1.7) for the multinomial sampling scheme. The LR statistic (1.8) becomes

$$G^2 = 2\sum_{i,j=1}^{I,J} n_{ij} \log\left(\frac{n_{ij}}{\hat{\mu}_{ij}}\right), \qquad \hat{\boldsymbol{\mu}} = (\hat{\mu}_{ij}). \tag{1.11}$$

Cressie and Read [1984] introduced a generalised parametric family of test statistics for $I \times J$ tables, which takes the form

$$CR_\lambda = \frac{2}{\lambda(\lambda+1)} \sum_{i,j=1}^{I,J} n_{ij} \left(\left(\frac{n_{ij}}{\hat{\mu}_{ij}}\right)^\lambda - 1\right), \qquad \lambda \neq -1, 0, \tag{1.12}$$

for $\boldsymbol{\mu} = (\hat{\mu}_{ij})$. For $\lambda = 1$, (1.12) becomes the Pearson statistic (1.10). For $\lambda \to 0$, CR_λ becomes the likelihood ratio test statistic (1.11). Cressie and Read [1984] suggested the use of $\lambda = 2/3$, which is used in the following and referred to as *Cressie-Read* (CR) statistic. Again, CR_λ from (1.12) and G^2 from (1.11) have the same asymptotic distribution as X^2 from (1.10).

1.6 Measures of Association

A main focus of CDA is the study of association between categorical variables. The categorical variables are independent if there is no association at all. If the categorical variables are not independent, one would want to explore the strength of association between the variables, and - in case of ordinal variables - the direction of association. Measures of association are a useful tool to help quantify such an association.

The concept of measures of association is very old with roots in 2×2 tables given by Yule [1900, 1912]. It finds applications in classical problems of CDA in sociology, economics, medicine and other scientific fields and and is still used today.

In 2×2 tables, the direction of association can only be one-dimensional (positive and negative). A change of the category $X = 1$ to $X = 2$ will either increase or decrease the success probability of Y. But for $I \times J$ tables, association can display more complex structures. For example, consider the cross-classification of several treatment doses (X) and their effect on the patients' health status (Y). While an increase of a too low dose will improve the health status, an increase of a high dose will worsen it due to the increasing side effects. Thus, the association of X on Y can be positive up to a level of X, say

$X = i$, and negative afterwards. Therefore, association in non-binary tables requires often a multi-dimensional description.

Since measures of association aim for a simple data analysis, they describe the association in a single number. Clearly, much information is lost during this process and the extend of this compromise is unclear. As Cramér [1924] pointed out, "*... there is no absolutely general measure of the degree of dependence.*" Hence, a mathematical definition of association does highly depend on the aim of the data analysis itself. However, the reduction process also gives a great flexibility to the concept of association, which is reflected by the many different measures of association developed during the last centuries. Goodman and Kruskal [1954] share this point of view as they write:

> "*Our belief is that each scientific area that has use for measures of association should, after appropriate argument and trail, settle down on these measures most useful for its needs.*"
>
> - Leo A. Goodman and William H. Kruskal, 1954

Very pioneering work for measuring association has been done by Goodman and Kruskal [1954, 1959, 1963, 1972]. Following their suggestions, Costner [1965] give a sociological view on measures of association. He prefers the interpretation of measures as proportional reduction in error (PRE) measures, which measures the quality of a measure in terms of predictability. He also gives an overview for PRE measures as measures of association. Another approach is the use of Pearson's statistic X^2 (1.10) for the independence hypothesis ($\hat{\mu}_{ij} = n_{i+}n_{+j}/n$). The X^2 statistic can be used to measure the divergence of an observation from independence, giving a geometrical interpretation of association. The most popular measures based on Pearson's statistic are the following:

$$\begin{aligned}
\phi^2 &= X^2/n && \text{(Phi)} \\
V &= \sqrt{\frac{X^2/n}{\min(I,J)-1}} && \text{(Cramér's } V\text{)} \\
C &= \sqrt{\frac{X^2}{n+X^2}} && \text{(Contingency Coefficient).}
\end{aligned} \tag{1.13}$$

Notice that Pearson's X^2 statistic is sensitive to sample size. Thus for large sample sizes it can make a weak relationship significant. Therefore, association measures based on X^2 often include n to correct this effect. The contingency coefficient for $I \times J$ tables is bounded by $C_{\max} = \left(\frac{(I-1)(J-1)}{IJ}\right)^{-1/4}$. Thus a standardized contingency coefficient can be calculated by $C_s := C/C_{\max}$, which lies in the interval $[0,1]$ and can therefore be used to value the strength of the association between X and Y.

Boundedness of measures is a very useful property for interpretation. As Davenport and El-Sanhurry [1991] and Warrens [2008] state, it is desirable for measures of association to lie in the values range $[-1,1]$, where 0 corresponds to independence, -1 corresponds to perfect negative and 1 corresponds to perfect positive association. Measures lying in $[0,1]$, like C_s, do not give the direction of association, but can be used to describe the strength of association.

Let $\boldsymbol{\pi} = (\pi_{ij}) \in \Delta_{2,2}$ be a multinomial probability for a 2×2 contingency table. A very fundamental measure of association in this case is the odds ratio $\theta = \pi_{11}\pi_{22}/(\pi_{12}\pi_{21}) \in$

$[0, \infty]$, which - due to its importance - will be considered in detail in Section 1.7. Since the odds ratio has the disadvantage of unboundedness, Yule [1900, 1912] introduced two measures based on θ with value range $[-1, 1]$:

$$Q = \frac{\theta - 1}{\theta + 1} \qquad \text{Yule [1900]},$$
$$Y = \frac{\sqrt{\theta} - 1}{\sqrt{\theta} + 1} \qquad \text{Yule [1912]}.$$

In comparison to measures based on Pearson's X^2, Q and Y also give the direction of association aside from the just the strength of association. A value of $Q, Y < 0$ indicates a negative association, while $Q, Y > 0$ indicates positive association.

Acock and Stavig [1979] generalised Cramér's V (1.13) by replacing the Pearson's test statistic X^2 (1.10) under independence ($\hat{\mu}_{ij} = n_{i+}n_{+j}/n$) with other well-known statistics like the LR statistic G^2 (1.11), i.e. by defining

$$V^* = \sqrt{\frac{G^2/n}{\min(I, J) - 1}}.$$

Further generalisations are possible by replacing the test statistic in V^* by the Cressie-Read statistic (1.12), which gives a family of measures $V^*_\lambda = \sqrt{\frac{CR_\lambda/n}{\min(I,J)-1}}$, controlled by a real parameter λ.

1.7 The Odds Ratio

Let $Y \in \{1, 2\}$ be a binary random variable with success probability $\mathbf{P}(Y = 1) = \pi \in (0, 1)$. Very often, the odds of success

$$\text{odds} = \frac{\pi}{1 - \pi},$$

which relates the success probability to the failure probability, is used for interpretation. An odds of four means, that a success is four times as likely as a failure. Using $\pi = \text{odds}/(1 + \text{odds})$, this is equivalent to $\pi = 0.2$.

Odds are very useful when comparing two independent responses. Let $X \in \{1, 2\}$ be an explanatory variable and $\pi_i \in (0, 1), i = 1, 2$, be the success probability of $Y = 1$ under the condition $X = i$, i.e. $\mathbf{P}(Y = 1 \mid X = i) = \pi_i$. Let $\text{odds}_i = \pi_i/(1 - \pi_i)$ be the odds in category $X = i$. The most fundamental measure of association for 2×2 tables is the *odds ratio*, which is defined as

$$\theta_{\mathcal{B}} = \frac{\pi_1}{1 - \pi_1} \Bigg/ \frac{\pi_2}{1 - \pi_2} \tag{1.14}$$

The odds ratio can be expressed in terms of multinomial probabilities $\boldsymbol{\pi} = (\pi_{ij}) \in \Delta_{2,2}$ for a 2×2 table:

$$\theta = \frac{\pi_{11}/\pi_{12}}{\pi_{21}/\pi_{22}} = \frac{\pi_{11}\pi_{22}}{\pi_{12}\pi_{21}}. \tag{1.15}$$

Assume that the success probabilities $\pi_i, i = 1, 2$ come from conditioning on the row marginal probabilities, i.e. $\pi_i = \pi_{ij}/\pi_{i+}, i = 1, 2$ (cf. Section 1.2). It is easy to verify that the odds ratio is sampling scheme invariant:

$$\theta_{\mathcal{B}} = \frac{\pi_1}{1-\pi_1} \cdot \frac{1-\pi_2}{\pi_2} = \frac{\pi_{11}/\pi_{1+}}{\pi_{12}/\pi_{1+}} \cdot \frac{\pi_{22}/\pi_{2+}}{\pi_{21}/\pi_{2+}} = \frac{\pi_{11}/\pi_{12}}{\pi_{21}/\pi_{22}} = \frac{\pi_{11}\pi_{22}}{\pi_{12}\pi_{21}} = \theta. \tag{1.16}$$

The same results holds for the Poisson sampling scheme with parameter $\boldsymbol{\gamma} = (\gamma_{ij}), i, j = 1, 2$, when $\pi_{ij} = \gamma_{ij}/\sum_{i,j=1}^{I,J} \gamma_{ij}$. Then it holds $\theta = \gamma_{11}\gamma_{22}/(\gamma_{12}\gamma_{21})$. Therefore, θ and $\theta_{\mathcal{B}}$ will not be distinguished in the following.

The odds ratio was first defined and studied by Yule [1900, 1912]. An odds ratio $\theta = 1$ corresponds to independence between the classification variables X and Y. An odds ratio $\theta > 1$ ($\theta < 1$) indicates positive (negative) association and the strength of association increases with the size of $|\theta - 1|$. Since $\theta \in [0, 1)$ represents negative association and $\theta \in (1, \infty]$ represents positive association, the value range is skew. To overcome this problem, the odds ratio is often transformed to log-scale, leading to the log-*odds ratio* $\log\theta$, which takes values in $[-\infty, \infty]$, where $\log(0) = -\infty$ in the case of $\pi_{11} = 0$ or $\pi_{22} = 0$. Here, the sign of $\log\theta$ distinguish the direction of association (positive or negative), while $\log\theta = 0$ is equivalent to independence between the classification variables X and Y. The term $|\log\theta|$ can be used to value the strength of association.

Let $\mathbf{N} = (N_{ij}) \sim \mathcal{M}(n, \boldsymbol{\pi})$. By the invariance property of MLEs ([Casella and Berger, 2002, Theorem 7.2.10, p. 320]), the MLE of $\log\theta$ is

$$\log\hat{\theta}(\mathbf{N}) = \log\left(\frac{N_{11}N_{22}}{N_{12}N_{21}}\right) \tag{1.17}$$

with estimate $\log\hat{\theta}(\mathbf{n}) = \log\left(\frac{n_{11}n_{22}}{n_{12}n_{21}}\right)$, which is the same under the Poisson or product binomial sampling scheme. The log-odds ratio takes values in $[-\infty, \infty]$, while in presence of sampling zeros, $\log\hat{\theta}(\mathbf{n}) = \infty$ if $n_{12} = 0$ or $n_{21} = 0$. The problem with sampling zeros is further commented in Section 1.7.2.

The log-odds ratio estimate for 2×2 contingency tables is calculated using the R function `log.OR` (see Appendix B.1).

The basic properties of the log-odds ratio are given in the following proposition:

Proposition 1.7.1. Let $\boldsymbol{\pi} = (\pi_{ij}) \in \Delta_{2,2}$ be a multinomial probability. The log-odds ratio $\log\theta$, defined in (1.15), exhibits the following properties:

(i) $\log\theta = 0$ if and only if X and Y are independent.

(ii) For fixed row and column marginals, $\log\theta$ is increasing in π_{11}.

(iii) By altering the entries of a single column (row) keeping the column (row) marginals fixed, $\log\theta$ is monotonic in the row (column) marginals difference.

(iv) $\log\theta$ is invariant under table rotation.

(v) $\log\theta$ changes sign when rows or columns are interchanged (Antisymmetry)

(vi) $\log\theta$ is inversion invariant, e.g. invariant under changing row and column entries.

(vii) $\log\theta$ is scaling invariant, i.e. for a matrix $\mathbf{n} = [n_{ij}] \in \mathbb{N}_0^{2\times 2}$ and numbers $0 \neq k_1, k_2 \in \mathbb{R}$, it holds $\log\hat{\theta}(\mathbf{n}) = \log\hat{\theta}(\mathbf{Dn})$, where $\mathbf{D} = \mathrm{diag}(k_1, k_2)$ is the 2×2 diagonal matrix with entries k_1 and k_2.

Proof. Properties (i) - (vi) are a special case ($\phi(x) = \phi_{KL}(x) = x\log x - x + 1$) of Proposition 3.1.3. Property (vii) follows directly by

$$\log\hat{\theta}(\mathbf{Dn}) = \log\left(\frac{k_1 n_{11} k_2 n_{22}}{k_1 n_{12} k_2 n_{21}}\right) = \log\left(\frac{n_{11}n_{22}}{n_{12}n_{21}}\right) = \log\hat{\theta}(\mathbf{n}).$$

□

It is well-known that the odds ratio encodes the association of 2×2 tables and, with knowledge of the marginals, specifies the joint distribution $\boldsymbol{\pi} \in \Delta_{2,2}$ completely such that another parametrization of $\boldsymbol{\pi}$ is the triple $\{\pi_{1+}, \pi_{+1}, \theta\}$ (cf. Kateri [2014, p. 43]). For a given odds ratio value $\theta_0 \in (0, \infty)$, let $\tilde{\boldsymbol{\pi}}(\theta_0, \pi_{1+}, \pi_{+1}) = (\tilde{\pi}_{ij}(\theta_0, \pi_{1+}, \pi_{+1})) \in \Delta_{2,2}$ be the multinomial probability vector with marginals π_{1+} and π_{+1} such that

$$\frac{\tilde{\pi}_{11}(\theta_0, \pi_{1+}, \pi_{+1})\,\tilde{\pi}_{22}(\theta_0, \pi_{1+}, \pi_{+1})}{\tilde{\pi}_{12}(\theta_0, \pi_{1+}, \pi_{+1})\,\tilde{\pi}_{21}(\theta_0, \pi_{1+}, \pi_{+1})} = \theta_0. \tag{1.18}$$

Problem (1.18) is a quadratic equation. For $\theta_0 \neq 1$ set

$$A_1 = -\frac{1 - (1-\theta_0)(\pi_{1+} + \pi_{+1})}{2(1-\theta_0)} \qquad \text{and} \qquad B_1 = \sqrt{A_1^2 + \frac{\theta_0 \pi_{1+}\pi_{+1}}{1-\theta_0}}.$$

The solution of (1.18) becomes

$$\tilde{\pi}_{11}(\theta_0, \pi_{1+}, \pi_{+1}) = \begin{cases} A_1 + B_1 & \text{if } \theta_0 < 1 \\ \pi_{1+}\pi_{+1} & \text{if } \theta_0 = 1 \\ A_1 - B_1 & \text{if } \theta_0 < 1. \end{cases} \tag{1.19}$$

The analytical solution to the quadratic equation (1.18) for the product binomial sampling scheme with fixed marginal counts $n_i, i = 1, 2$ and $n_{+j}, j = 1, 2$, and for $\theta_0 \neq 1$, the solution is

$$\begin{aligned} \tilde{\pi}_1 = \tilde{\pi}_1(\theta_0, n_1, n_2, n_{+1}) &= \frac{\tilde{\pi}_2\theta_0}{1 + \tilde{\pi}_2(\theta_0 - 1)}, \\ \tilde{\pi}_2 = \tilde{\pi}_2(\theta_0, n_1, n_2, n_{+1}) &= \frac{-B_2 + \sqrt{B_2^2 - 4A_2C_2}}{2A_2}, \end{aligned} \tag{1.20}$$

where

$$A_2 = n_2(\theta_0 - 1), \qquad B_2 = n_1\theta_0 + n_2 - n_{+1}(\theta_0 - 1), \qquad C_2 = -n_{+1}.$$

The case $\theta_0 = 1$ corresponds to independence ($\tilde{\pi}_1(1, n_1, n_2, n_{+1}) = \tilde{\pi}_2(1, n_1, n_2, n_{+1}) = \frac{n_{+1}}{n_1 + n_2}$) (Miettinen and Nurminen [1985, p. 218]).

Another very useful property of the log-odds ratio is its asymptotic normality, which enables construction of asymptotic tests and confidence intervals for $\log\theta$.

Lemma 1.7.2. **(i)** Let $\boldsymbol{\pi} = (\pi_{ij}) \in \Delta_{2,2}$ be a multinomial probability vector, $n \in \mathbb{N}$ a sample size and let $\mathbf{N} = (N_{ij}) \sim \mathcal{M}(n, \boldsymbol{\pi})$. Let $\hat{\theta}$ be the MLE of the odds ratio given in (1.17). Then $\log \hat{\theta}$ obtains asymptotic normality

$$\sqrt{n}(\log \hat{\theta} - \log \theta) \xrightarrow{d} \mathcal{N}(0, \sigma^2(\boldsymbol{\pi})), \qquad \text{for } n \to \infty, \tag{1.21}$$

where

$$\sigma^2(\boldsymbol{\pi}) = \frac{1}{\pi_{11}} + \frac{1}{\pi_{12}} + \frac{1}{\pi_{21}} + \frac{1}{\pi_{22}}.$$

(ii) Let $\pi_i \in (0,1)$ and $n_i \in \mathbb{N}, i = 1,2$. Let N_{i1} be independent binomial distributed random variable $\mathcal{B}(n_i, \pi_i)$, $i = 1,2$. Let $\hat{\theta}_{\mathcal{B}}$ be the MLE of the odds ratio for the product binomial sampling scheme, also given in (1.17). Then, for $n_1, n_2 \to \infty$ with $\lim\limits_{n_1,n_2\to\infty} \frac{n_2}{n_1} = r$, $\log \hat{\theta}_{\mathcal{B}}$ obtains asymptotic normality

$$\sqrt{n_1}\left(\log \hat{\theta}_{\mathcal{B}} - \log \theta_{\mathcal{B}}\right) \xrightarrow{d} \mathcal{N}(0, \sigma^2_{\mathcal{B}}(\pi_1, \pi_2, r)) \qquad \text{for } n \to \infty, \tag{1.22}$$

where

$$\sigma^2_{\mathcal{B}}(\pi_1, \pi_2, r) = \frac{1}{\pi_1(1-\pi_1)} + \frac{1}{r\pi_2(1-\pi_2)}. \tag{1.23}$$

Proof. Part (i) is a special case of Theorem 3.2.1 and part (ii) is a special case of Theorem 3.2.2 when using $\phi(x) = \phi_{KL}(x) = x \log x - x + 1$. The result can also be derived using the delta method (Theorem 1.4.3). □

In the multinomial and product binomial sampling scheme, the variance estimate of $\log \hat{\theta}$ resp. $\log \hat{\theta}_{\mathcal{B}}$ is

$$\hat{\sigma}^2 := \sigma^2(\hat{\boldsymbol{\pi}}) = \frac{1}{n_{11}} + \frac{1}{n_{12}} + \frac{1}{n_{21}} + \frac{1}{n_{22}} = \sigma^2_{\mathcal{B}}(\hat{\pi}_1, \hat{\pi}_2, r) =: \hat{\sigma}^2_{\mathcal{B}}.$$

Thus, construction of asymptotic tests and confidence intervals for the log-odds ratio does not depend on the sampling scheme.

1.7.1 Tests and Confidence Intervals for the log-Odds Ratio

Using the asymptotic normality of the log-odds ratio $\log \theta$ (Lemma 1.23), the corresponding Wald test (1.6) is well-known for testing the hypothesis $H_0 : \theta = \theta_0$ against $H_1 : \theta \neq \theta_0$ in a 2×2 contingency table $\mathbf{n} = (n_{ij})$ (cf. Agresti [2013]). It becomes

$$W^2 = (\log \hat{\theta} - \log \theta_0)^2 \left(\frac{1}{n_{11}} + \frac{1}{n_{12}} + \frac{1}{n_{21}} + \frac{1}{n_{22}}\right)^{-1}. \tag{1.24}$$

Now, in addition fix the marginal counts n_{1+} and n_{+1}, and sample size n of $\mathbf{n} = (n_{ij})$. A score test for the odds ratio θ can be constructed using the parametrization of a multinomial probability vector $\boldsymbol{\pi} \in \Delta_{2,2}$ as $\{\theta, \pi_{1+}, \pi_{+1}\}$. For fixed $\theta_0 \in (0, \infty)$ and fixed marginal probabilities $\pi_{1+} = n_{1+}/n$ and $\pi_{+1} = n_{+1}/n$, let $\tilde{\pi}_{ij}(\theta_0, \pi_{1+}, \pi_{+1})$ be the multinomial probability as given in (1.19), leading to an odds ratio θ_0. In the multinomial set-up, the score statistic S^2 becomes Pearson's test statistic (1.10). It has the form

$$S^2 = \sum_{i,j=1}^{2} \frac{(n_{ij} - n\tilde{\pi}_{ij})^2}{n\tilde{\pi}_{ij}}, \qquad \tilde{\pi}_{ij} = \tilde{\pi}_{ij}(\theta_0, \pi_{1+}, \pi_{+1}), \quad i,j = 1,2. \tag{1.25}$$

Score CIs for odds ratio have already been introduced by Cornfield [1956] and are considered in Miettinen and Nurminen [1985]. Fixing the marginals is justifiable as the marginals in a contingency table are ancillary statistics. With marginals $n_i, i = 1,2$, (1.25) gives - up to a correction factor $\frac{n_1+n_2-1}{n_1+n_2}$, which makes the variance estimate less biased - the same test statistic as introduced by Miettinen and Nurminen [1985] for the binomial sampling scheme. With success probabilities π_i, $i = 1,2$, it is algebraically equivalent to

$$S^2 = \frac{(n_1(\hat{\pi}_1 - \tilde{\pi}_1))^2}{n_1\tilde{\pi}_1(1-\tilde{\pi}_1)} + \frac{(n_2(\hat{\pi}_2 - \tilde{\pi}_2))^2}{n_2\tilde{\pi}_2(1-\tilde{\pi}_2)}, \tag{1.26}$$

where $\tilde{\pi}_i = \tilde{\pi}_i(\theta_0, n_1, n_2, n_{+1}), i = 1,2$ is the solution given in (1.20) and $\hat{\pi}_i = N_{ij}/n_{i+}$ is the MLE of $\pi_i, i = 1,2$.

Since $n\tilde{\pi}_{ij}(\theta_0, \pi_{1+}, \pi_{+1}), i,j = 1,2$ are the expected cell frequencies of the 2×2 table under fixed marginals for given odds ratio θ_0, the likelihood ratio (LR) test statistic (1.11) takes the form

$$G^2 = 2\sum_{i,j=1}^{2} n_{ij} \log\left(\frac{n_{ij}}{n\tilde{\pi}_{ij}}\right), \qquad \tilde{\pi}_{ij} = \tilde{\pi}_{ij}(\theta_0, \pi_{1+}, \pi_{+1}), \quad i,j = 1,2. \tag{1.27}$$

As seen in Section 1.5, the Pearson statistic (1.25) and the LR statistic (1.27) are special members of the Cressie-Read statistic (1.28) for $\lambda = 1$ and $\lambda \to 0$, respectively. The CR statistic is

$$CR_\lambda = \frac{2}{\lambda(\lambda+1)} \sum_{i,j=1}^{2} n_{ij}\left(\left(\frac{n_{ij}}{n\tilde{\pi}_{ij}}\right)^\lambda - 1\right), \quad \tilde{\pi}_{ij} = \tilde{\pi}_{ij}(\theta_0, \pi_{1+}, \pi_{+1}), \quad \lambda \neq -1, 0. \tag{1.28}$$

Under $H_0 : \theta = \theta_0$ all test statistics, W^2, S^2, G^2 and CR_λ are asymptotic χ^2_1 distributed.

Remark 1.7.3. The values of the test statistics S^2, G^2 and CR_λ for 2×2 tables in the product binomial sampling scheme is calculated using the R functions `OR.score.test`, `OR.LR.test` and `OR.CR.test`, respectively, which is found in the digital attachment of this work (see Appendix B.1). In addition, a program (`OR.Pearson.test`) for calculating the value of the Pearson test statistic (1.25) in the product binomial sampling has been written, which coincides with `OR.score.test` as expected.

Confidence intervals at level $\alpha \in (0,1)$ can easily be constructed by inverting the test statistic (cf. Casella and Berger [2002, Theorem 9.2.2, pp. 421]), that is, the CI is the set of θ_0 such that the test statistic does not reject the hypothesis $H_0 : \theta = \theta_0$ at level α. In the case of the Wald test statistic, inversion is well-known and is done analytical. The two-sided $(1-\alpha)$% Wald CI for the odd ratio is

$$\exp\left(\left[\log\hat{\theta} \pm u_{1-\alpha/2}\sqrt{\left(\frac{1}{n_{11}} + \frac{1}{n_{12}} + \frac{1}{n_{21}} + \frac{1}{n_{22}}\right)}\right]\right), \tag{1.29}$$

where $u_{1-\alpha/2}$ is the $(1-\alpha/2)$-quantile of a standard normal distribution. For the score test, the Likelihood Ratio test and the Cressie-Read test, inversion has to be done numerically to get the sets

$$\{\theta_0 \in (0, \infty) \mid T \leq \chi^2_{1,\alpha}\}, \qquad T \in \{S^2, G^2, CR_\lambda\}, \tag{1.30}$$

where $\chi^2_{1,\alpha}$ is the $(1-\alpha)$-quantile of a χ^2_1 distribution.

Remark 1.7.4. The confidence set (1.30) based on inverting the test statistics S^2, G^2 and CR_λ for 2×2 tables in the product binomial sampling scheme is calculated using the R functions `get.score.CI`, `get.LR.CI` and `get.CR.CI` which is found in the digital attachment of this work (see Appendix B.1). In addition, a program (`get.Pearson.CI`) for calculating the Pearson confidence interval by inverting the Pearson test statistic (1.25) in the product binomial sampling has been written, which coincides with the outcome of `get.score.CI`. The score CI using the correction factor $\frac{n_1+n_2-1}{n_1+n_2}$ as given in Miettinen and Nurminen [1985] is calculated using `get.Pearson.CI2` or using the `orscoreci` of the **PropCIs**-package in R.

Remark 1.7.5. Score CIs (1.25) for the odds ratio seem to perform better in comparison to the Wald CI (1.24) and is the spotlight of newer research (Agresti [2003, 2011]; Agresti et al. [2008]; Agresti and Min [2002]).

Inversion of Cressie-Read's test statistic (1.28) has already been done by Bedrick [1987] to get CIs for the ratio of two binomial proportions (π_1/π_2). Application of (1.28) in the context of odds ratio is new and will be studied extensively in Chapter 2 of this work.

1.7.2 Examples and Problems with Sampling Zeros

For the larynx data (Example 1.2.1), the log-odds ratio is estimated as $\log\hat{\theta} = 0.74$ (or $\hat{\theta} = 2.1$) with an standard error estimated of $SE = 0.97$. The odds to belong in the cancer controlled no category are 2.1 times greater for the radiation therapy than the odds for the surgery therapy. The Wald confidence interval for the log-odds ratio at $\alpha = 5\%$ is $[\log\hat{\theta} \pm 1.96SE] = [-1.17, 2.65]$. Since 0 is included in the Wald confidence interval, the Wald test does not reject the hypothesis of independence for the larynx data. For Example 1.2.2, since the prednisolone data includes a sampling zero, the log-odds ratio estimate becomes $\log\hat{\theta} = \infty$.

The asymptotic CIs (Section 1.7.1) for both examples are presented in Table 1.2. Recall that both data sets have fixed row marginals (product binomial sampling), while Example 1.2.2 possesses a sampling zero. Thus, the Wald CI (1.24) is not defined, since $\hat{\sigma}_{\mathcal{B}}(\hat{\pi}_1, \hat{\pi}_2, r) = \infty$. Score, CR and LR CIs have a length of infinity and are calculated using the R function `calc.log.CI`, `get.score.CI`, `get.LR.CI` and `get.CR.LR` given in the attachment (see Appendix B.1).

All CIs for Example 1.2.1 include $\theta = 1$, indicating that the hypothesis of independence cannot be rejected at significance level $\alpha = 5\%$, while this is not the case for Example 1.2.2.

Focusing on 2×2 contingency tables, the bibliography on inference for the odds ratio is enormous rich, with part of it concentrating on its estimation and the study of the

Table 1.2: CIs ($\alpha = 5\%$) of θ for Example 1.2.1 (left) and Example 1.2.2 (right).

Larynx Data

Type	CI
Wald	[0.312, 14.152]
Score	[0.366, 11.859]
CR ($\lambda = 2/3$)	[0.311, 17.477]
LR	[0.346, 13.211]

Prednisolone Data

Type	CI
Wald	-
Score	$[2.923, \infty)$
CR ($\lambda = 2/3$)	$[3.293, \infty)$
LR	$[5.114, \infty)$

behavior of corresponding confidence intervals when a zero cell count occurs, causing the asymptotic CIs to be either not defined or to obtain a length of infinity. A missing upper (or lower) bound for the CIs of θ is problematic in the data analysis, as they suggest an association strength up to infinity. One way to overcome this problem caused by the presence of sampling zeros is to use the continuity correction by adding $c = 0.5$ to each cell (Anscombe [1956]; Gart [1962]; Haldane [1956]). The corrected estimate for Example 1.2.2 becomes $\log \hat{\theta}_c = 3.31$ with $SE_c = 1.52$ and the corresponding 95% Wald CI is $\exp([0.33, 6.29]) = [1.39, 539.83]$ for θ, supporting significant positive association. There are more correction techniques (cf. Table 2.2) and the effect of continuity corrections on CIs of the log-odds ratio will be studied extensively in Chapter 2 of this work.

1.8 Relative Risk

Next to the odds ratio, another measure for comparing two independent binomial probabilities in 2×2 tables is the relative risk, defined as the ratio of success probabilities

$$R = \pi_1/\pi_2. \tag{1.31}$$

When $\mathcal{R} = 1$, the success probabilities are equal ($\pi_1 = \pi_2$). Higher success probabilities in the first row category are equivalent to $\mathcal{R} \in (1, \infty)$, while $\mathcal{R} \in (0, 1)$ indicates a higher success probability in the second row. Therefore, similar to the odds ratio (Section 1.7), the value scale of $\mathcal{R}$ is skew. To overcome this problem, $\mathcal{R}$ is often transformed to the log-relative risk $\log \mathcal{R} \in (-\infty, \infty)$. In addition, the log-relative risk obtains an asymptotic normal distribution, which will later be proofed in a more general context (Theorem 3.36).

There is a simple connection between the odds ratio and the relative risk:

$$\theta = \mathcal{R} \cdot \frac{1 - \pi_2}{1 - \pi_1}. \tag{1.32}$$

Often, when π_1 and π_2 are small, the approximation $\theta \approx \mathcal{R}$ is used. Cornfield [1951] suggested the use of this approximation, when the relative risk itself cannot be estimated. Take for example Table 1.3, which shows the two different study designs to determine if smoking is a carcinogen. In design (a), an investigation by selecting groups with and without a characteristic (smoking) and observing over time if they develop the disease (lung cancer) is conducted. Such a study is expensive and time-consuming. Often these kind of studies are reversed (design (b)), selection groups with and without the disease and checking them for the characteristic.

Table 1.3: Study designs.

(a)

	Lung cancer yes	Lung cancer no
Smoker	π_1	$1-\pi_1$
Non-smoker	π_2	$1-\pi_2$

(b)

	Lung cancer yes	Lung cancer no
Smoker	π_1	π_2
Non-smoker	$1-\pi_1$	$1-\pi_2$

In study design (b), the relative risk is just the odds of developing lung cancer, which is not the interest of the study. But the odds ratio estimates are equal in both study designs. Thus, in such a case with π_1 and π_2 close to one, the approximation $\theta \approx \mathcal{R}$ following from (1.32) is very useful, as it approximates the relative risk in such sampling schemes.

1.9 Generalized Odds Ratios for $I \times J$ Tables

The odds ratio θ can be extended to $I \times J$ tables and its generalisations occur naturally as an important tool for interpretation in contingency table analysis. Let $X \in \{1, ..., I\}$ and $Y \in \{1, ..., J\}$ be two classification variables and $\boldsymbol{\pi} = (\pi_{ij}) \in \Delta_{I,J}$ be a multinomial probability vector with $\mathbf{P}(X = i, Y = j) = \pi_{ij}$. Let i_1, i_2 be two row categories and j_1, j_2 be two column categories. The four cells $(i_a, j_b), a, b = 1, 2$, forming a 2×2 subtable, can be used to form an odds ratio

$$\theta_{i_1 j_1}^{i_2 j_2} = \frac{\pi_{i_1 j_1}\pi_{i_2 j_2}}{\pi_{i_2 j_1}\pi_{i_1 j_2}}, \tag{1.33}$$

which measures the association between the categories $X \in \{i_1, i_2\}$ and $Y \in \{j_1, j_2\}$. Let $\pi_{j|i} = \mathbf{P}(Y = j \mid X = i) = \pi_{ij}/\pi_{i+}, i = 1, ..., I, j = 1, ..., J$ be the conditional probabilities under conditioning of the row marginals in the product multinomial sampling scheme. Then (1.33) can be expressed as

$$\theta_{i_1 j_1}^{i_2 j_2} = \frac{\pi_{j_1|i_1}\pi_{j_2|i_2}}{\pi_{j_1|i_2}\pi_{j_2|i_1}} = \frac{\pi_{j_1|i_1}/\pi_{j_2|i_1}}{\pi_{j_1|i_2}/\pi_{j_2|i_2}},$$

and $\theta_{i_1 j_1}^{i_2 j_2}$ describes the change of the ratio $\pi_{j_1|i}/\pi_{j_2|i}$, when the row category $i = i_1$ is changed to $i = i_2$. The most natural choice of subtables is the comparison between two adjacent categories ($i_2 = i_1 + 1, j_2 = j_1 + 1$). Then there are $(I-1)(J-1)$ 2×2 subtables corresponding to the *local odds ratios*

$$\theta_{ij}^L = \theta_{ij}^L(\boldsymbol{\pi}) = \frac{\pi_{ij}\pi_{i+1j+1}}{\pi_{i+1j}\pi_{ij+1}}, \qquad i = 1, ..., I-1, j = 1, ..., J-1. \tag{1.34}$$

Any odds ratio (1.33) can be expressed in terms of the local odds ratio:

$$\theta_{i_1 j_1}^{i_2 j_2} = \prod_{i_1 \le i < i_2} \prod_{j_1 \le j < j_2} \theta_{ij}^L. \tag{1.35}$$

Remark 1.9.1. As Agresti [2013, Section 2.4.1, p. 54] points out, the conversion of the probabilities $\boldsymbol{\pi} = (\pi_{ij})$ into the local odds ratios (1.34) does not discard information. As the marginals $\{\pi_{i+}, \pi_{+j} \mid i = 1, ..., I, j = 1, ..., J\}$ and the odds ratios $\{\theta_{ij}^L \mid i = 1, ..., I, j = 1, ..., J\}$ uniquely specify $\boldsymbol{\pi} = (\pi_{ij})$. Therefore, a minimal set of $(I-1)(J-1)$ local odds ratios is sufficient to describe the association in an $I \times J$ table.

Further parameter sets are given by the generalised odds ratios, including the local odds ratio (cf. [Kateri, 2014, Section 2.2.5]). When one underlying classification variable is ordinal, association can also be measured by comparing events being up to or above a specific category. When using this cumulative option, generalised odds ratios can be defined by merging subsets of cells to other 2×2 subtables. Fusing the cells greater than the selected category of $Y = j$ leads to the *continuation odds ratio*:

$$\theta_{ij}^{CO} = \theta_{ij}^{CO}(\boldsymbol{\pi}) = \frac{\pi_{j|i} \sum_{b>j} \pi_{b|i+1}}{\pi_{j+1|i} \sum_{b>j} \pi_{b|i}} = \frac{\pi_{ij} \sum_{b>j} \pi_{i+1b}}{\pi_{ij+1} \sum_{b>j} \pi_{ib}}, \qquad i = 1, ..., I-1, j = 1, ..., J-1. \tag{1.36}$$

Merging all cells left and right of the selected category of $Y = j$ leads to the *cumulative odds ratio*:

$$\theta_{ij}^{C} = \theta_{ij}^{C}(\boldsymbol{\pi}) = \frac{\sum_{b\leq j} \pi_{b|i} \sum_{b>j} \pi_{b|i+1}}{\sum_{b\leq j} \pi_{b|i+1} \sum_{b>j} \pi_{b|i}} = \frac{\sum_{b\leq j} \pi_{ib} \sum_{b>j} \pi_{i+1b}}{\sum_{b\leq j} \pi_{i+1b} \sum_{b>j} \pi_{ib}}, \; i = 1, ..., I-1, j = 1, ..., J-1. \tag{1.37}$$

When both classification variables are ordinal, the whole table can be collapsed into dichotomies. This leads to the *global odds ratio*:

$$\theta_{ij}^{G} = \theta_{ij}^{G}(\boldsymbol{\pi}) = \frac{\sum_{a\leq i} \sum_{b\leq j} \pi_{ab} \sum_{a>i} \sum_{b>j} \pi_{ab}}{\sum_{a\leq i} \sum_{b>j} \pi_{ab} \sum_{a<i} \sum_{b\leq j} \pi_{ab}}, \qquad i = 1, ..., I-1, j = 1, ..., J-1. \tag{1.38}$$

Notice that the global odds ratio cannot be described using the conditional probabilities $\pi_{j|i}$ in the product multinomial sampling scheme. But it holds

$$\theta_{ij}^{G} = \frac{\mathbf{P}(Y \leq j|X \leq i)\mathbf{P}(Y > j|X > i)}{\mathbf{P}(Y \leq j|X > i)\mathbf{P}(Y > j|X \leq i)}, \qquad i = 1, ..., I-1, j = 1, ..., J-1.$$

Write $\boldsymbol{\theta}^h = \boldsymbol{\theta}^h(\boldsymbol{\pi}) = (\theta_{11}^h, ..., \theta_{I-1J-1}^h)(\boldsymbol{\pi}), h = L, CO, C, G$. The odds ratios $\boldsymbol{\theta}^L, \boldsymbol{\theta}^{CO}, \boldsymbol{\theta}^C, \boldsymbol{\theta}^G$ are called *generalised odds ratios*. They can be estimated in terms of the expected cell frequencies. Let $\mathbf{N} = (N_{ij})$ follow the multinomial sampling scheme $\mathbf{N} \sim \mathcal{M}(n, \boldsymbol{\pi})$. Then the generalised odds ratios are estimated by replacing the probabilities $\boldsymbol{\pi}$ by the sample proportions $\hat{\boldsymbol{\pi}} = \mathbf{N}/n$ in formulas (1.34) to (1.38), i.e. $\hat{\boldsymbol{\theta}}^h = \boldsymbol{\theta}^h(\hat{\boldsymbol{\pi}})$.

Interpretation is analoguous to that of θ for the 2×2 case, using the collapsed categories: A global odds ratio $\theta_{ij}^G > 1$ indicates, that the odds to ly in the categories $Y \leq j$ instead of lying in $Y > j$ are higher for the categories $X \leq i$ than for $X > i$. Independence is a special case. For all generalized odds ratios it holds that

$$X \text{ and } Y \text{ are independent} \quad \Leftrightarrow \quad \theta_{ij}^h = 1 \text{ for all } i = 1, ..., I-1\,, j = 1, ..., J-1, \tag{1.39}$$

for all $h = L, CO, C, G$.

1.9.1 Transformation Properties

The behaviour of the generalised odds ratios under table transformation can easily be shown and is given by the following proposition:

Proposition 1.9.2. Let $\boldsymbol{\pi}$ be an $I \times J$ multinomial probability.

1. It holds $\boldsymbol{\theta}^L(\boldsymbol{\pi}^T) = (\boldsymbol{\theta}^L(\boldsymbol{\pi}))^T$ and $\boldsymbol{\theta}^G(\boldsymbol{\pi}^T) = (\boldsymbol{\theta}^G(\boldsymbol{\pi}))^T$.
2. Let $\boldsymbol{\theta}_i$ be the row of generalised odds ratios between the rows i and $i+1$. Let $\boldsymbol{\theta}_{i_1}^{i_2}$ be the generalised odds ratio between i_1 and i_2. Clearly $\boldsymbol{\theta}_i = \boldsymbol{\theta}_i^{i+1}$ and $\boldsymbol{\theta}_i^{i+2} = \boldsymbol{\theta}_i \cdot \boldsymbol{\theta}_{i+1}$. Changing the rows i_1 and i_2 gives the following transformation for all $h = L, CO, C$:
$$(\boldsymbol{\theta}_1, ..., \boldsymbol{\theta}_{I-1})^T \mapsto (\boldsymbol{\theta}_1, ..., \boldsymbol{\theta}_{i_1-2}, \boldsymbol{\theta}_{i_1-1}^{i_2}, \boldsymbol{\theta}_{i_2}^{i_1+1}, \boldsymbol{\theta}_{i_1+1}, ..., \boldsymbol{\theta}_{i_2-2}, \boldsymbol{\theta}_{i_2-1}^{i_2}, \boldsymbol{\theta}_{i_1}^{i_2+1}, \boldsymbol{\theta}_{i_2+1}, ..., \boldsymbol{\theta}_{I-1})^T.$$
A similar results holds under column change for $h = L$.
3. Let $F_R, F_C : \mathbb{R}^{I\times J} \to \mathbb{R}^{I\times J}$ be the functions that reverse the order of the rows $(\mathbf{x}_1, ..., \mathbf{x}_I)$ resp. the columns $(\mathbf{x}'_1, ..., \mathbf{x}'_J)$:
$$F_R(\mathbf{x}_1, ..., \mathbf{x}_I) = (\mathbf{x}_I, ..., \mathbf{x}_1) \qquad \text{and} \qquad F_C(\mathbf{x}'_1, ..., \mathbf{x}'_J) = (\mathbf{x}'_J, ..., \mathbf{x}'_1).$$
It holds
$$\boldsymbol{\theta}^h(F_R(\boldsymbol{\pi})) = -F_R(\boldsymbol{\theta}^h(\boldsymbol{\pi})) \qquad \text{and} \qquad \boldsymbol{\theta}^h(F_C(\boldsymbol{\pi})) = -F_C(\boldsymbol{\theta}^h(\boldsymbol{\pi}))$$
for $h = L, CO, C, G$.
4. It holds
$$\boldsymbol{\theta}^h(F_R(F_C(\boldsymbol{\pi}))) = F_R(F_C((\boldsymbol{\theta}^h(\boldsymbol{\pi})))$$
for $h = L, CO, C, G$.

Since $h = C, CO$ require ordinal response variable in the column and $h = G$ requires both variables to be ordinal, not all transformations are useful. These have been excluded from Proposition 1.9.2. For example, there is no natural symmetry for $h = CO, C$, but transformation leads to new odds ratios with continuation (cumulation) in the columns instead of the rows.

1.9.2 Dependence Concepts

As Lehmann [1966] points out, positive (negative) association occurs if higher values of one variable correspond to higher (lower) values of the other variable. Lehmann applied the theory of stochastic orders in this situation. It turns out, that there is a general relation of the dependence concepts to the different generalised log-odds ratios (Douglas et al. [1990]).

Definition 1.9.3. Let $\mathbf{P}$ be the joint distribution of (X, Y). Then, X and Y are

(i) positive likelihood-ratio dependent[1] if $\mathbf{P}(X = x, Y = y')\mathbf{P}(X = x', Y = y) \leq \mathbf{P}(X = x, Y = y)\mathbf{P}(X = x', Y = y')$ for all $x < x'$ and $y < y'$.

(ii) positive hazard rate ordered[2] if $\mathbf{P}(X \geq x, Y \geq y) \geq \mathbf{P}(X \geq y, Y \geq x)$ for all $x \leq y$.

(iii) positive regression dependent[3] if $\mathbf{P}(Y \leq y \mid X = x)$ is non-increasing in x.

(iv) positive quadrant dependent[4] if $\mathbf{P}(X \leq x, Y \leq y) \geq \mathbf{P}(X \leq x)\mathbf{P}(Y \leq y)$ for all x, y.

Lemma 1.9.4. (Douglas et al. [1990, Theorem 4.1, p. 198])

For the generalised log-odds ratio (1.34) to (1.38) the following relation holds:

$$\begin{aligned}
\log \boldsymbol{\theta}^{L} \geq 0 \quad &\Leftrightarrow \quad X \text{ and } Y \text{are positive likelihood-ratio dependent,} \\
\log \boldsymbol{\theta}^{CO} \geq 0 \quad &\Leftrightarrow \quad X \text{ and } Y \text{are positive hazard rate ordered,} \\
\log \boldsymbol{\theta}^{C} \geq 0 \quad &\Leftrightarrow \quad X \text{ and } Y \text{are positive regression dependent,} \\
\log \boldsymbol{\theta}^{G} \geq 0 \quad &\Leftrightarrow \quad X \text{ and } Y \text{are positive quadrant dependent.}
\end{aligned}$$

There is also a hierarchy of the generalised odds ratios.

Lemma 1.9.5. (Douglas et al. [1990, Figure 3.1, p. 197]) It holds:

$$\log \boldsymbol{\theta}^{L} \geq 0 \quad \Rightarrow \quad \log \boldsymbol{\theta}^{CO} \geq 0 \quad \Rightarrow \quad \log \boldsymbol{\theta}^{C} \geq 0 \quad \Rightarrow \quad \log \boldsymbol{\theta}^{G} \geq 0 \quad \Rightarrow$$

1.9.3 Asymptotic Behaviour

In order to present the asympotitc behaviour for the generalised odds ratios in a unified manner, it is useful to introduce a notation for the sums used to define (1.34) to (1.38). Let $\mathbf{x} = [x_{ij}] \in \mathbb{R}^{I \times J}$ be a real $I \times J$ matrix. For $0 \leq i_1 \leq i_2 \leq I$ and $0 \leq j_1 \leq j_2 \leq J$ the sum of element of $\mathbf{x}$ between these indices is defined as:

$$\mathfrak{F}(i_1 : i_2, j_1 : j_2; \mathbf{x}) := \sum_{\substack{i_1 \leq i \leq i_2 \\ j_1 \leq j \leq j_2}} x_{ij}. \tag{1.40}$$

When $i_1 = i_2$ or $j_1 = j_2$, write i_1 and j_1 instead of $i_1 : i_1$ and $j_1 : j_1$, respectively. Let $\boldsymbol{\pi}$ be an $I \times J$ multinomial probability. Define the following short notations ($i = 1, ..., I-1, j =$

[1] [Lehmann, 1966, pp. 1150]
[2] [Shaked and Shanthikumar, 2007, p. 17]
[3] [Lehmann, 1966, pp. 1143]
[4] [Lehmann, 1966, pp. 1137]

$1, ..., J-1$):

$$\begin{aligned}
a_{ij}^{L}(\boldsymbol{\pi}) &= \mathfrak{F}(i,j;\boldsymbol{\pi}) & b_{ij}^{L}(\boldsymbol{\pi}) &= \mathfrak{F}(i,j+1;\boldsymbol{\pi}) \\
c_{ij}^{L}(\boldsymbol{\pi}) &= \mathfrak{F}(i+1,j;\boldsymbol{\pi}) & d_{ij}^{L}(\boldsymbol{\pi}) &= \mathfrak{F}(i+1,j+1;\boldsymbol{\pi}) \\
a_{ij}^{CO}(\boldsymbol{\pi}) &= \mathfrak{F}(i,j;\boldsymbol{\pi}) & b_{ij}^{CO}(\boldsymbol{\pi}) &= \mathfrak{F}(i,(j+1):J;\boldsymbol{\pi}) \\
c_{ij}^{CO}(\boldsymbol{\pi}) &= \mathfrak{F}(i+1,j;\boldsymbol{\pi}) & d_{ij}^{CO}(\boldsymbol{\pi}) &= \mathfrak{F}(i+1,(j+1):J;\boldsymbol{\pi}) \\
a_{ij}^{C}(\boldsymbol{\pi}) &= \mathfrak{F}(i,1:j;\boldsymbol{\pi}) & b_{ij}^{C}(\boldsymbol{\pi}) &= \mathfrak{F}(i,(j+1):J;\boldsymbol{\pi}) \\
c_{ij}^{C}(\boldsymbol{\pi}) &= \mathfrak{F}(i+1,1:j;\boldsymbol{\pi}) & d_{ij}^{C}(\boldsymbol{\pi}) &= \mathfrak{F}(i+1,(j+1):J;\boldsymbol{\pi}) \\
a_{ij}^{G}(\boldsymbol{\pi}) &= \mathfrak{F}(1:i,1:j;\boldsymbol{\pi}) & b_{ij}^{G}(\boldsymbol{\pi}) &= \mathfrak{F}(1:i,(j+1):J;\boldsymbol{\pi}) \\
c_{ij}^{G}(\boldsymbol{\pi}) &= \mathfrak{F}((i+1):I,1:j;\boldsymbol{\pi}) & d_{ij}^{G}(\boldsymbol{\pi}) &= \mathfrak{F}((i+1):I,(j+1):J;\boldsymbol{\pi})
\end{aligned} \tag{1.41}$$

Each generalised log-odds ratio can be defined by

$$\log \theta_{ij}^{h} = \log \theta_{ij}^{h}(\boldsymbol{\pi}) = \log\left(\frac{a_{ij}^{h}(\boldsymbol{\pi})d_{ij}^{h}(\boldsymbol{\pi})}{b_{ij}^{h}(\boldsymbol{\pi})c_{ij}^{h}(\boldsymbol{\pi})}\right), \qquad h = L, CO, C, G. \tag{1.42}$$

Thus, the sums (1.41) are called *essential sums* in the following.

The generalised odds ratios can also be defined using matrices. Let $\mathbf{A}^h$ be an appropriate $IJ \times (I-1)(J-1)$ matrix and let $\mathbf{C}^h$ be an appropriate $(I-1)(J-1) \times IJ$ matrix, both consisting of elements $\{-1, 0, 1\}$, such that

$$\log \boldsymbol{\theta}^h = \mathbf{C}^h \log(\mathbf{A}^h \boldsymbol{\pi}), \qquad h = L, CO, C, G. \tag{1.43}$$

The matrices $\mathbf{A}^h$ and $\mathbf{C}^h$ is calculated using the R functions `local.odds.DM`, `cont.odds.DM`, `cum.odds.DM` and `global.odds.DM` available in the web-appendix of Kateri [2014]. The MLE of $\log \boldsymbol{\theta}^h$ becomes

$$\log \hat{\boldsymbol{\theta}}^h = \mathbf{C}^h \log(\mathbf{A}^h \hat{\boldsymbol{\pi}}), \qquad h = L, CO, C, G. \tag{1.44}$$

The representation (1.44) simplifies the calculation of the asymptotic distribution:

Theorem 1.9.6. Let $\mathbf{N} \sim \mathcal{M}(n, \boldsymbol{\pi})$ be multinomial distributed with sample size $n \in \mathbb{N}$ and probability vector $\boldsymbol{\pi} = (\pi_{ij}) \in \Delta_{I,J}$. Let $\hat{\boldsymbol{\pi}} = \mathbf{N}/n$ be the MLE of $\boldsymbol{\pi}$. Let $\mathbf{A}^h$ and $\mathbf{C}^h$ be the matrices as in (1.44). Then the MLE $\log \hat{\boldsymbol{\theta}}^h = \mathbf{C}^h \log(\mathbf{A}^h \hat{\boldsymbol{\pi}})$ of $\log \boldsymbol{\theta}^h, h = L, CO, C, G$ is asymptotic normal distributed with variance-covariance matrix

$$\boldsymbol{\Sigma}^h(\boldsymbol{\pi}) := \mathbf{C}^h \text{diag}\left(\frac{1}{\mathbf{A}^h \boldsymbol{\pi}}\right) \mathbf{A}^h \boldsymbol{\Sigma}(\boldsymbol{\pi})(\mathbf{A}^h)^T \text{diag}\left(\frac{1}{\mathbf{A}^h \boldsymbol{\pi}}\right)(\mathbf{C}^h)^T, \tag{1.45}$$

where $\boldsymbol{\Sigma}(\boldsymbol{\pi}) = \text{diag}(\boldsymbol{\pi}) - \boldsymbol{\pi}^T \boldsymbol{\pi}$ and $\frac{1}{\mathbf{A}^h \boldsymbol{\pi}}$ is the vector of component-wise inverse elements of $\mathbf{A}^h \boldsymbol{\pi}$. In other words, it holds

$$\sqrt{n}(\log \hat{\boldsymbol{\theta}}^h - \log \boldsymbol{\theta}^h) \xrightarrow{d} \mathcal{N}(0, \boldsymbol{\Sigma}^h(\boldsymbol{\pi})), \qquad \text{for } n \to \infty.$$

Proof. By Lemma 1.4.2, it holds

$$\sqrt{n}(\hat{\boldsymbol{\pi}} - \boldsymbol{\pi}) \xrightarrow{d} \mathcal{N}(0, \boldsymbol{\Sigma}(\boldsymbol{\pi})), \qquad \text{for } n \to \infty.$$

Applying the chain rule on $\log(\mathbf{A}^h\boldsymbol{\pi})$, it holds for the Jacobian matrix

$$\mathbf{D} := \frac{\partial}{\partial \boldsymbol{\pi}} \log(\mathbf{A}^h\boldsymbol{\pi}) = \text{diag}\left(\frac{1}{\mathbf{A}^h\boldsymbol{\pi}}\right)\mathbf{A}^h.$$

Clearly, $\mathbf{D} \neq \mathbf{0}$ for any multinomial probability vector $\boldsymbol{\pi} \in \Delta_{I,J}$. Applying the delta method (Theorem 1.4.3) gives

$$\sqrt{n}(\log(\mathbf{A}^h\hat{\boldsymbol{\pi}}) - \log(\mathbf{A}^h\boldsymbol{\pi})) \xrightarrow{d} \mathcal{N}\left(0, \text{diag}\left(\frac{1}{\mathbf{A}^h\boldsymbol{\pi}}\right)\mathbf{A}^h\boldsymbol{\Sigma}(\boldsymbol{\pi})(\mathbf{A}^h)^T\text{diag}\left(\frac{1}{\mathbf{A}^h\boldsymbol{\pi}}\right)\right)$$

for $n \to \infty$. Then the result follows as

$$\begin{aligned}\sqrt{n}(\log\hat{\boldsymbol{\theta}}^h - \log\boldsymbol{\theta}^h) &= \sqrt{n}\mathbf{C}^h(\log(\mathbf{A}^h\hat{\boldsymbol{\pi}}) - \log(\mathbf{A}^h\boldsymbol{\pi}))\\ &\xrightarrow{d} \mathcal{N}(0, \mathbf{C}^h\text{diag}\left(\frac{1}{\mathbf{A}^h\boldsymbol{\pi}}\right)\mathbf{A}^h\boldsymbol{\Sigma}(\boldsymbol{\pi})(A^h)^T\text{diag}\left(\frac{1}{\mathbf{A}^h\boldsymbol{\pi}}\right)(\mathbf{C}^h)^T)\end{aligned}$$

for $n \to \infty$. □

The MLEs (1.44) and asymptotic variance-covariance matrices (1.45) for $h = L, CO, C, G$ is calculated using the R function `log.genOR` (see Appendix B.1).

Remark 1.9.7. The asymptotic variance-covariance for the local log-odds ratio $\boldsymbol{\Sigma}^L(\boldsymbol{\pi})$ has already been calculated in Agresti [2013, Section 16.1.6, p. 619]. In this case, it holds asymptotically

$$\begin{aligned}\mathbb{V}\text{ar}\ (\sqrt{n}\log\hat{\theta}^L_{ij}) &\overset{asymp.}{=} \frac{1}{\pi_{ij}} + \frac{1}{\pi_{ij+1}} + \frac{1}{\pi_{i+1j}} + \frac{1}{\pi_{i+1j+1}}\\ \text{Cov}(\sqrt{n}\log\hat{\theta}^L_{ij}, \sqrt{n}\log\hat{\theta}^L_{i+1j}) &\overset{asymp.}{=} -\frac{1}{\pi_{ij}} - \frac{1}{\pi_{i+1j+1}}\\ \text{Cov}(\sqrt{n}\log\hat{\theta}^L_{ij}, \sqrt{n}\log\hat{\theta}^L_{i+1j+1}) &\overset{asymp.}{=} \frac{1}{\pi_{i+1j+1}}\\ \text{Cov}(\sqrt{n}\log\hat{\theta}^L_{i+1j}, \sqrt{n}\log\hat{\theta}^L_{ij+1}) &\overset{asymp.}{=} \frac{1}{\pi_{i+1j+1}}\\ \text{and Cov}(\sqrt{n}\log\hat{\theta}^L_{ij}, \sqrt{n}\log\hat{\theta}^L_{kl}) &\overset{asymp.}{=} 0, \text{if } \hat{\theta}_{ij} \text{ and } \hat{\theta}_{kl} \text{ use mutually exclusive set of cells.}\end{aligned}$$

1.9.4 Connection of Generalized Odds Ratios to Models

Contingency tables can be analysed by loglinear models, which are special members of the generalised linear models (GLMs) introduced by P. McCullagh, J. Nelder and R. Wedderburn (cf. McCullagh and Nelder [1989]). They can be used to detect and test underlying association structures.

Consider an $I \times J$ contingency table $\mathbf{n} = (n_{ij})$ with classification variables X and Y for rows and columns, respectively. Assuming that $\mathbf{n} = (n_{ij})$ is a realisation of $\mathbf{N} = (N_{ij}) \sim$

$\mathcal{M}(n, \boldsymbol{\pi})$, the expected frequency counts $\mu_{ij} = \mathbb{E}N_{ij} = n\,\pi_{ij}, i = 1, ..., I, j = 1, ..., J$ are linked to a linear predictor using the log-link. Then, the saturated log-linear model is defined as

$$\log \mu_{ij} = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_{ij}^{XY}, \qquad i = 1, ..., I, j = 1, ..., J. \tag{1.46}$$

The parameters $\{\lambda_i^X\}$ and $\{\lambda_j^Y\}$ give the row resp. column main effects, while the set $\{\lambda_{ij}^{XY}\}$ encodes the association. To enable identifiability, a reference constrain such as

$$\lambda_1^X = 0, \quad \lambda_1^Y = 0, \quad \lambda_{1j}^{XY} = 0, j = 1, ..., J, \quad \lambda_{i1}^{XY} = 0,\ i = 1, ..., I, \tag{1.47}$$

or

$$\lambda_I^X = 0, \quad \lambda_J^Y = 0, \quad \lambda_{Ij}^{XY} = 0, j = 1, ..., J, \quad \lambda_{iJ}^{XY} = 0,\ i = 1, ..., I, \tag{1.48}$$

or the sum-to-zero constraint

$$\begin{aligned} &\sum_{i=1}^{I} \lambda_i^X = 0, && \sum_{j=1}^{J} \lambda_j^Y = 0, \\ &\sum_{i=1}^{I} \lambda_{ij}^{XY} = 0, \quad j = 1, ..., J, && \sum_{j=1}^{J} \lambda_{ij}^{XY} = 0, \quad i = 1, ..., I, \end{aligned} \tag{1.49}$$

is required. The saturated model (1.46) just give a reparameterization of the observed table counts, providing a perfect fit (the corresponding goodness-of-fit statistics (Section 1.5) equal zero and have $df = 0$).

The fact, that the interaction parameters $\{\lambda_{ij}^{XY}\}$ express the underlying associations in a contingency tables becomes clear through their connection to the local odds ratios. In particular, the local odds ratios (1.34) can equivalently be defined in terms of expected cell frequencies which can be expressed by (1.46). Doing so, it is easy to verify that

$$\log \theta_{ij}^L = \log \left(\frac{\mu_{ij}\mu_{i+1j+1}}{\mu_{i+1j}\mu_{ij+1}} \right) = \lambda_{ij}^{XY} + \lambda_{i+1j+1}^{XY} - \lambda_{i+1j}^{XY} - \lambda_{ij+1}^{XY}, \tag{1.50}$$

for $i = 1, ..., I-1, j = 1, ..., J-1$. Using (1.50) in 2×2 tables with reference constraint (1.47), it holds $\log \theta = \lambda_{22}^{XY}$ for the log-odds ratio.

When setting all $\lambda_{ij}^{XY} = 0, i = 1, ..., I, j = 1, ..., J$ in (1.46), all local log-odds ratio in (1.50) become zero, which by (1.39) is equivalent to independence. The associated loglinear model of independence is then

$$\log \mu_{ij} = \lambda + \lambda_i^X + \lambda_j^Y, \qquad i = 1, ..., I, j = 1, ..., J. \tag{1.51}$$

The goodness-of-fit statistics (Section 1.5) for the independence model (1.51) have $df = IJ - 1 - (I - 1 + J - 1) = (I-1)(J-1)$ degrees of freedom.

So far, the very restrictive independence model (1.51) and the saturated model (1.46) for contingency tables were introduced, giving $(I-1)(J-1)$ or 0 degrees of freedom, respectively. To cover more complex association structures than independence, but more parsimonious models than the saturated model, association models are introduced. As seen in Section 1.9 association can be described using the generalised odds ratios, where $\theta_{ij}^h = 1, i = 1, ..., I-1, j = 1, ..., J-1$ is equivalent to independence. The classical association models can be derived by assuming patterns of the underlying association in terms of the local odds ratios. Setting

$$\log \theta_{ij}^L = \varphi, \qquad i = 1, ..., I-1, j = 1, ..., J-1, \tag{1.52}$$

for a constant $\varphi \in (-\infty, \infty)$, gives a model with the same association in every 2×2 subtable formed by the adjacent row and column categories. This model was studied by Goodman [1979] and has already been mentioned in Haberman [1974b]. Since association is uniform over the rows and column categories for the cross-classified variables X and Y, model (1.52) is called the *uniform association model*. It has one parameter more than the independence model and independence is a special case when $\varphi = 0$. Thus the goodness-of-fit statistics (Section 1.5) have $df = (I-1)(J-1)-1$ degrees of freedom. Using (1.35), it holds

$$\log \theta_{i_1 j_1}^{i_2 j_2} = \varphi(i_2 - i_1)(j_2 - j_1), \qquad i_1 < i_2, j_1 < j_2. \tag{1.53}$$

Thus, the strength of association increases for row and column categories further apart.

As seen in (1.53), distances between categories are an important part for association models. For ordinal variables, it is often natural to assume different distances between categories. This can be captured when (known) scores $(\mu_1, ..., \mu_I)$ and $(\nu_1, ..., \nu_J)$ are assigned to the row and column categories, respectively. Using an intrinsic association parameter φ, model (1.52) can be generalised to

$$\log \theta_{ij}^L = \varphi(\mu_{i+1} - \mu_i)(\nu_{j+1} - \nu_j), \qquad i = 1, ..., I-1, j = 1, ..., J-1. \tag{1.54}$$

This model is called *linear-by-linear association model*, since the local log-odds ratios depend linearly on the differences of the scores, while φ determines the association strength. In the case $\mu_i = i, i = 1, ..., I$ and $\nu_j = j, j = 1, ..., J$ or any other choice of equidistant scores, model (1.54) becomes the uniform association model (1.52).

To enable inference, model (1.54) requires a representation as loglinear model. Setting the interaction parameter $\lambda_{ij}^{XY} = \varphi\mu_i\nu_j, i = 1, ..., I, j = 1, ..., J$, the model can be expressed as

$$\log \mu_{ij} = \lambda + \lambda_i^X + \lambda_j^Y + \varphi\mu_i\nu_j, \qquad i = 1, ..., I, j = 1, ..., J. \tag{1.55}$$

To ensure identifiability, either the sum constraints

$$\sum_{i=1}^{I} \omega_{1i}\mu_i = 0 = \sum_{j=1}^{J} \omega_{2j}\nu_j, \qquad \text{and} \qquad \sum_{i=1}^{I} \omega_{1i}\mu_i^2 = 1 = \sum_{j=1}^{J} \omega_{2j}\nu_j^2, \tag{1.56}$$

or the reference constraints $\mu_1 = 0 = \nu_1$ (or $\mu_I = 0 = \nu_J$) are assigned to the scores (cf. (1.47), (1.48) and (1.49)), where $\{\omega_{1i}\}$ and $\{\omega_{2j}\}$ are weights. Popular choices for the weights are uniform weights ($\omega_{1i} = 1, i = 1, ..., I$ and $\omega_{2j} = 1, j = 1, ..., J$) or marginal weights ($\omega_{1i} = \pi_{i+}, i = 1, ..., I$ and $\omega_{2j} = \pi_{+j}, j = 1, ..., J$).

Like the uniform association model (1.52), model (1.55) has one additional parameter compared to the independence model (1.51) and has $df = (I-1)(J-1)-1$ degrees of freedom. Therefore, the model is very parsimonious. When the linear-by-linear association model is insufficient, the number of parameters can be increased by assuming that either $\{\mu_i\}$ or $\{\nu_j\}$ are parametric. In this case, the *row-effect* (R) model and *column-effect* (C) model are obtained, respectively. Since φ can be encoded in the parametric scores, it becomes redundant. The R model has $df = (I-1)(J-2)$ degrees of freedom, while the C model has $df = (I-2)(J-1)$. For the R (C) model, the local odds ratio becomes equal in every row (column), i.e.

$$\log \theta_{ij}^L = c_{i1}, \qquad i = 1, ..., I-1, j = 1, ..., J-1 \qquad \text{(R model)}, \tag{1.57}$$
$$\log \theta_{ij}^L = c_{j2}, \qquad i = 1, ..., I-1, j = 1, ..., J-1 \qquad \text{(C model)}, \tag{1.58}$$

for parameters $c_{i1} \in (-\infty, \infty)$ or $c_{j2} \in (-\infty, \infty)$, respectively.

When both scores $\{\mu_i\}$ and $\{\nu_j\}$ are unknown, model (1.55) becomes the *RC model*,

$$\log \theta_{ij}^L = c_{i1} c_{2j}, \qquad i = 1, ..., I-1, j = 1, ..., J-1 \qquad \text{(RC model)}.$$

Notice that the RC model is no longer a member of the loglinear model class, as the predictor is nonlinear. This class of models can be captured by the **gnm** (generalised nonlinear models) package in R.

Remark 1.9.8. The loglinear models are presented here for multinomial sampling schemes. Birch [1963] showed that MLEs in loglinear models are the same for the three different sampling schemes introduced in Section 1.2 as long, as the likelihood equations of the model fix the corresponding marginal conditions, i.e. for the Poisson sampling scheme the total sample size is fixed, while for the product multinomial sampling scheme the row (or column) marginals are fixed. For more information see Agresti [2002, p. 339, Section 8.6.7] or Birch [1963].

1.9.5 HLP Models

Dale [1984, 1986] generalised model (1.55) using the global odds ratios (1.38) instead of the local ones,

$$\log \theta_{ij}^G = \varphi(\mu_{i+1} - \mu_i)(\nu_{j+1} - \nu_j), \qquad i = 1, ..., I-1, j = 1, ..., J-1.$$

Extension of Dale's approach to the other generalised odds ratio and for other association models requires introduction of new models.

Let $\boldsymbol{\pi}$ be an $I \times J$ multinomial probability. Lang and Agresti [1994] introduced *generalised log-linear models* (GLLM) as a broad class of multivariate categorical response data models, which have the form

$$\mathbf{C} \log(\mathbf{A}\boldsymbol{\pi}) = \mathbf{X}\boldsymbol{\beta}. \tag{1.59}$$

The matrices $\mathbf{C}, \mathbf{A}, \mathbf{X}$ are used to specify the model with model parameters $\boldsymbol{\beta}$. GLLMs include many models for categorical data, for example logit- and cumulative models and they can be applied on a wide range of data classes (longitudinal, rater agreement, crossover data, ...). Maximum likelihood methods for model (1.59) have been studied in Lang

[1996]. This model is fitted in R through the `mph.fit` package by Lang. GLLMs open the opportunity allow to model the association directly on level of the generalised log-odds ratio and do not require a defined on $\boldsymbol{\pi}$, showing their flexibility. For example, the global uniform association model ($\log \theta_{ij}^G = \varphi, i = 1, ..., I-1, j = 1, ..., J-1$) is obtained by using $\mathbf{A}^G$ and $\mathbf{C}^G$ from (1.43), and the appropriate choice for the model matrix $\mathbf{X} = \mathbf{1}_{(I-1)(J-1)}$ as vector of ones with size $(I-1)(J-1)$.

Lang [2004, 2005] generalised the GLLMs further to the *homogeneous linear predictor* (HLP) models, which have the form

$$\mathbf{L}(\boldsymbol{\pi}) = \mathbf{X}\boldsymbol{\beta}, \tag{1.60}$$

for a link function $\mathbf{L}$ fulfilling special requirements mentioned in Appendix A.3 (Definition A.3.3). Clearly, for $\mathbf{L}(\boldsymbol{\pi}) = \mathbf{C}\log(\mathbf{A}\boldsymbol{\pi})$ model (1.60) becomes model (1.59).

Therefore, the generalised log-odds ratios are used to define the *generalised association models*

$$\log \theta_{ij}^h = \varphi(\mu_{i+1} - \mu_i)(\nu_{j+1} - \nu_j), \qquad i = 1, ..., I-1, j = 1, ..., J-1, \tag{1.61}$$

for $h = L, CO, C, G$ using model (1.60) and the choice $\mathbf{L}(\boldsymbol{\pi}) = \mathbf{C}^h \log(\mathbf{A}^h\boldsymbol{\pi})$ for the link function, where $\mathbf{C}^h$ and $\mathbf{A}^h$ are the matrices from (1.43).

1.9.6 Symmetry Models

In many study designs, one has to deal with matched classification variables. Such data often occurs in surveys at two different time points for the same group of subjects, surveys of married couples or otherwise related persons as for example in social mobility tables. Cross-classification of two such commensurable variables require special models. One very basic model is the symmetry model. For a multinomial probability vector $\boldsymbol{\pi} = (\pi_{ij}) \in \Delta_{I,I}$, symmetry holds when

$$\pi_{ij} = \pi_{ji}, \qquad i, j = 1, ..., I. \tag{1.62}$$

Symmetry implies marginal homogeneity, i.e.

$$\pi_{i+} = \pi_{+i}, \qquad i = 1, ..., I. \tag{1.63}$$

For $I = 2$ both (1.62) and (1.63) are equivalent. The symmetry model (1.62) can also be represented as loglinear model. For $\mathbf{N} = (N_{ij}) \sim \mathcal{M}(n, \boldsymbol{\pi})$ and $\boldsymbol{\mu} = \mathbb{E}\mathbf{N} = n\boldsymbol{\pi}$, the model of symmetry becomes

$$\log \mu_{ij} = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_{ij}^{XY}, \qquad i, j = 1, ..., I, \tag{1.64}$$

with the additional constraints (beyond the identifiable constraints)

$$\lambda_{ij}^{XY} = \lambda_{ji}^{XY} \qquad i, j = 1, ..., I, \tag{1.65}$$

$$\lambda_i^X = \lambda_i^Y \qquad i = 1, ..., I. \tag{1.66}$$

Relaxing the constraint on the main effects (1.66), model (1.64) with constraint (1.65) defines the *quasi-symmetry model* by Caussinus [1965]. The model of quasi-symmetry can be expressed in terms of the local log-odds ratios $\boldsymbol{\theta}^L$ (1.34). It is equivalent to

$$\theta_{ij}^L = \theta_{ji}^L, \qquad i,j = 1,...,I-1. \tag{1.67}$$

Kateri [2014, Section 9.2] and Agresti [2013, Section 11.4] can be used as reference for more information on symmetry model.

1.10 Information Theory: The ϕ-Divergence

A very basic concept in statistical information theory is the measuring of divergence between probability distributions. The goodness-of-fit test statistics, such as Pearson's X^2 (1.10), likelihood ratio statistic G^2 (1.11) and Cressie-Read's test statistic $CR_\lambda, \lambda \neq -1, 0$, (1.28) are based on such divergences between the observed frequency distributions and the estimated frequencies under the assumed hypothesis (or model).

Measures of divergence between probability distributions are in the focus of statistical research since the 1925 (Lévy [1925], Mahalanobis [1936], Bhattacharya [1943], Bhattacharya [1946],Rényi [1961]). As one of the most fundamental works, Shannon [1948] introduced the information theory as a mathematical theory of communication, which is based on divergence measures. Its applications within statistical science have been studied during the last centuries. A traditional reference for the application of Information Theory within statistics is Kullback [1959].

However, in the 1970s a generalised theory of divergence measures was independently introduced by Csiszár [1963] and Ali and Silvey [1966] that covers many of the already studied distance concepts and gave a good basis for further development:

Definition 1.10.1. The ϕ-divergence between two discrete finite bivariate distributions $\mathbf{p} = (p_{ij})$ and $\mathbf{q} = (q_{ij})$ in $\Delta_{I,J}$ is defined as

$$\mathcal{D}^\phi(\mathbf{p}, \mathbf{q}) = \sum_{i,j=1}^{I,J} q_{ij}\phi\left(\frac{p_{ij}}{q_{ij}}\right), \tag{1.68}$$

with $\phi \in \Phi$, where Φ is the set of convex functions on $[0, \infty)$, satisfying $\phi(1) = \phi'(1) = 0$, $0 \cdot \phi(0/0) = 0$, and $0 \cdot \phi(x/0) = x \cdot \lim_{u\to\infty} \phi(u)/u$.

The ϕ-divergence reflects the ability to distinguish between $\mathbf{p}$ and $\mathbf{q}$. Notice that the additional requirements on $\phi \in \Phi$ in Definition 1.10.1 define the unique convex and lower semi-continuous extension of the function $(0,\infty)^2 \to (-\infty, \infty], (u, v) \mapsto v\phi\left(\frac{u}{v}\right)$ on $[0,\infty)^2$ (cf. Vajda [1989]). In addition, Ali and Silvey [1966, Theorem 1] showed that the convexity of ϕ is a necessary condition for $\mathcal{D}^\phi$ to fulfil the so called information loss property: Let $\mathbf{p}$ and $\mathbf{q}$ be some probability measures and G be a measurable transformation, such that $\mathbf{p}^G$ and $\mathbf{q}^G$ are the probability measures on the image of G, then $D^\phi(\mathbf{p}, \mathbf{q}) \geq D^\phi(\mathbf{p}^G, \mathbf{q}^G)$. Thus, the information loss property states, that the ability to distinguish between $\mathbf{p}$ and $\mathbf{q}$ does not increase under transformations. While focus of this work lies in the context of categorical data and therefore formula (1.68) is only given for discrete probability functions $\mathbf{p} = (p_{ij})$ and $\mathbf{q} = (q_{ij})$, simple extension for general measure theoretic settings is possible. There are some nice other properties for the ϕ-divergence, which shall not be presented here. For more information about the mathematical background and properties of ϕ-divergences see Liese and Vajda [1987], Vajda [1989] and Pardo [2006].

The most important members of the ϕ-divergences family are presented.

Example 1.10.2. (Kullback-Leibler Divergence)

Based on Shannon [1948] the *Kullback-Leibler* (KL) divergence between two distributions has been introduced by Kullback and Leibler [1951]. The KL divergence in the case of discrete finite bivariate distributions $\mathbf{p} = (p_{ij}) \in \Delta_{I,J}$ and $\mathbf{q} = (q_{ij}) \in \Delta_{I,J}$ is defined by

$$I(\mathbf{p}, \mathbf{q}) = \sum_{i,j=1}^{I,J} p_{ij} \log \left(\frac{p_{ij}}{q_{ij}} \right). \tag{1.69}$$

In terms of ϕ-divergence it is the special case when

$$\phi_{KL}(x) = x \log x - x + 1. \tag{1.70}$$

It is also called *Information Divergence*. Kullback [1959] studied and summarized properties of the KL divergence useful for Information Theory and statistics. Next to some other, he also summarized its adjustment to multivariate data (including contingency tables).

Let $\mathbf{n} = (n_{ij})$ be a realisation of $\mathbf{N} \sim \mathcal{M}(n, \boldsymbol{\pi})$ a multinomial distribution with sample size $n \in \mathbb{N}$ and multinomial probability vector $\boldsymbol{\pi} \in \Delta_{I,J}$. Let $\hat{\boldsymbol{\pi}}_0$ be the MLE of $\boldsymbol{\pi}$ under a specified model $\boldsymbol{\pi} \in \Theta_0 \subseteq \Delta_{I,J}$ and let $\hat{\boldsymbol{\pi}} = \mathbf{N}/n$ be the MLE of $\boldsymbol{\pi}$ under the saturated model. Then the LR statistic (1.11) is $G^2 = 2nI(\hat{\boldsymbol{\pi}}, \hat{\boldsymbol{\pi}}_0)$.

Example 1.10.3. (Pearson Divergence)

The very first test statistic by Pearson [1900] given in (1.10) can also be represented as ϕ-divergence via

$$\phi_P(x) = \frac{1}{2}(x-1)^2, \tag{1.71}$$

called *Pearson Divergence*. In the case of $\mathbf{p} = (p_{ij}) \in \Delta_{I,J}$ and $\mathbf{q} = (q_{ij}) \in \Delta_{I,J}$ it is defined by

$$\mathcal{D}^{\phi_P}(\mathbf{p}, \mathbf{q}) = \frac{1}{2} \sum_{i,j=1}^{I,J} \frac{(p_{ij} - q_{ij})^2}{q_{ij}}. \tag{1.72}$$

The Pearson statistic (1.10) is $X^2 = 2n\mathcal{D}^{\phi_P}(\hat{\boldsymbol{\pi}}, \hat{\boldsymbol{\pi}}_0)$.

Example 1.10.4. (Power Divergence)

The *power divergence* is one of the most important members of the ϕ-divergence family, that forms a parametric family of divergences, controlled by a single parameter $\lambda \in \mathbb{R}$. It was introduced by Cressie and Read [1984] and is also called *Cressie-Read Divergence*:

$$\phi_\lambda(x) = \begin{cases} -\log x + x - 1 & \text{for } \lambda = -1 \\ \frac{x^{\lambda+1} - x - \lambda(x-1)}{\lambda(\lambda+1)} & \text{for } \lambda \neq -1, 0 \\ x \log x - x + 1 & \text{for } \lambda = 0, \end{cases} \tag{1.73}$$

such that in the case of $\mathbf{p} = (p_{ij}) \in \Delta_{I,J}$ and $\mathbf{q} = (q_{ij}) \in \Delta_{I,J}$ it holds

$$\mathcal{D}^{\phi_\lambda}(\mathbf{p},\mathbf{q}) = \frac{1}{\lambda(\lambda+1)} \sum_{i,j=1}^{I,J} p_{ij} \left(\left(\frac{p_{ij}}{q_{ij}} \right)^\lambda - 1 \right), \qquad \lambda \neq -1, 0. \tag{1.74}$$

Clearly, $\lambda = 1$ gives the Pearson divergence. The power divergence also includes the Kullback-Leibler divergence for $\lambda \to 0$, since

$$\lim_{\lambda \to 0} \mathcal{D}^{\phi_\lambda}(\mathbf{p},\mathbf{q}) = I(\mathbf{p},\mathbf{q}). \tag{1.75}$$

As seen, well-known statistics are smoothly connected by λ in the power divergence family, producing new test statistics with new properties. The Cressie-Read test statistic (1.28) is $CR_\lambda = 2n\mathcal{D}^{\phi_\lambda}(\hat{\boldsymbol{\pi}}, \hat{\boldsymbol{\pi}}_0)$. Cressie and Read [1984] pointed out, that the goodness-of-fit test statistic in the multinomial set-up, CR_λ for $\lambda = 2/3$ is a good compromise within the power divergence goodness-of-fit tests on the basis of test power and small sample approximation.

Divergence	ϕ-function
Kullback and Leibler [1951]	$\phi_{KL}(x) = x \log x - x + 1$
Pearson [1900]	$\phi_P(x) = \frac{1}{2}(x-1)^2$
Cressie and Read [1984]	$\phi_\lambda(x) = \frac{x^{\lambda+1} - x - \lambda(x-1)}{\lambda(\lambda+1)}$, $\lambda \neq -1, 0$
J-Divergence	$\phi(x) = (x-1)\log x$
Balakrishnan and Sanghvi [1968]	$\phi(x) = (x-1)^2(x+1)^{-2}$
Minimum Discrimination Information	$\phi(x) = -\log x + x - 1$

Table 1.4: Selection of ϕ-divergence functions, $\phi \in \Phi$, Definition 1.10.1.

Table 1.4 shows a selection of important ϕ-functions. This listing of ϕ-divergences is not exhaustive. An extension of the power divergence family is found in Kus et al. [2008]. Another family of divergence by Kateri et al. [2015] is introduced in the Appendix A.2.

As seen in Section 1.5, Pearson, KL and Cressie-Read divergences can be used to construct goodness-of-fit tests for multinomial models $\boldsymbol{\pi} = \boldsymbol{\pi}(\boldsymbol{\varphi}), \boldsymbol{\varphi} = (\varphi_1, ..., \varphi_K)$. This results holds for general $\phi \in \Phi$. Let $\mathbf{N} \sim \mathcal{M}(n, \boldsymbol{\pi})$ and let $\hat{\boldsymbol{\pi}} = \mathbf{N}/n$ be the MLE of $\boldsymbol{\pi}$ in the saturated model. Let $\hat{\boldsymbol{\pi}}_0 = \boldsymbol{\pi}_0(\hat{\boldsymbol{\varphi}})$ be the MLE of $\boldsymbol{\pi}$ under $H_0 : \boldsymbol{\pi} = \boldsymbol{\pi}_0(\boldsymbol{\varphi})$. Then for any strictly convex $\phi \in \Phi$, the statistic

$$T^\phi := \frac{2n}{\phi''(1)} \mathcal{D}^\phi(\hat{\boldsymbol{\pi}}, \hat{\boldsymbol{\pi}}_0)$$

has an asymptotic χ^2_{df} distribution with $df = IJ - 1 - K$ degrees of freedom (cf. [Pardo, 2006, Theorem 6.1, p. 259]). The hypothesis $H_0 : \boldsymbol{\pi} = \boldsymbol{\pi}(\varphi)$ is than rejected in favor for $H_1 : \boldsymbol{\pi} \neq \boldsymbol{\pi}(\varphi)$, if $T^\phi > \chi^2_{df,\alpha}$, where $\chi^2_{df,\alpha}$ is the $(1-\alpha)$-quantile of a χ^2_{df} distribution. Inverting the test statistic leads to a confidence set for φ,

$$\{\varphi \mid T^\phi \leq \chi^2_{df,\alpha}\}.$$

Inversion of the test statistic for the odds ratio θ in 2×2 tables has already been presented in Section 1.7.1 and will be applied in Chapter 2 to study the effect of different correction techniques on the odds ratio θ in 2×2 tables with small sample size.

1.11 ϕ-Divergence based Models

For the analysis of contingency tables, association model (1.55), which can be written as

$$\pi_{ij} = \pi_{i+}\pi_{+j} \exp\left(\alpha_i + \beta_j + \varphi\mu_i\nu_j\right), \qquad i = 1, ..., I, j = 1, ..., J, \tag{1.76}$$

and the correlation model (Gilula and Haberman [1986]; Goodman [1985]) given by

$$\pi_{ij} = \pi_{i+}\pi_{+j}(1 + \alpha_i + \beta_j + \varphi\mu_i\nu_j), \qquad i = 1, ..., I, j = 1, ..., J \tag{1.77}$$

are well-known. They were developed in the 80s simultaneously and were in a competition with each other. Later, the pioneering paper of Gilula et al. [1988] pointed out, that both models are - under certain conditions - the closest to independence in terms of Kullback-Leibler divergence (Example 1.10.2) and Pearson divergence (Example 1.10.3), respectively. In other words, let $\mathcal{A}$ be the set of multinomial probabilities vector $\boldsymbol{\pi} = (\pi_{ij}) \in \Delta_{I,J}$ with fixed marginals $\boldsymbol{\pi}^X = (\pi_{i+}) \in \Delta_I, \boldsymbol{\pi}^Y = (\pi_{+j}) \in \Delta_J$, fixed scores $\{\mu_i\}, \{\nu_j\}$ and fixed correlation $\rho = \sum_{ij} \pi_{ij}\mu_i\nu_j$. Then model (1.76) and (1.77) are solutions to

$$\min_{\boldsymbol{\pi}\in\mathcal{A}} I(\boldsymbol{\pi}, \boldsymbol{\pi}^I) \qquad \text{and} \qquad \min_{\boldsymbol{\pi}\in\mathcal{A}} \mathcal{D}^{\phi_P}(\boldsymbol{\pi}, \boldsymbol{\pi}^I), \tag{1.78}$$

respectively, where $\boldsymbol{\pi}^I = (\pi_{i+}\pi_{+j})$.

Rom and Sarkar [1992] introduced generalised association models using the power divergence family instead of Kullback-Leibler and Pearson divergence in (1.78), which was then generalised by Kateri and Papaioannou [1995] through the ϕ-divergence (Definition 1.10.1). Solving

$$\min_{\boldsymbol{\pi}\in\mathcal{A}} \mathcal{D}^{\phi}(\boldsymbol{\pi}, \boldsymbol{\pi}^I), \tag{1.79}$$

leads to *ϕ-association models*, $\phi \in \Phi$. For invertible ϕ', the it takes the form

$$\pi_{ij} = \pi_{i+}\pi_{+j}(\phi')^{-1}(\alpha_i + \beta_j + \varphi\mu_i\nu_j), \qquad i = 1, ..., I, j = 1, ..., J, \tag{1.80}$$

with sum-to-zero constraint for the scores as given in (1.56). It is the closest model to independence in $\mathcal{A}$ in terms of ϕ-divergence.

Some calculations show, that (1.80) leads to

$$\begin{aligned} \phi'\left(\frac{\pi_{ij}}{\pi_{i+}\pi_{+j}}\right) - \phi'\left(\frac{\pi_{ij+1}}{\pi_{i+}\pi_{+j+1}}\right) - \phi'\left(\frac{\pi_{i+1j}}{\pi_{i+1+}\pi_{+j}}\right) + \phi'\left(\frac{\pi_{i+1j+1}}{\pi_{i+1+}\pi_{+j+1}}\right) \\ = \varphi(\mu_{i+1} - \mu_i)(\nu_{j+1} - \nu_j), \qquad i = 1, ..., I-1, j = 1, ..., J-1, \end{aligned} \tag{1.81}$$

Clearly, (1.81) coincides with the local log-odds ratio (1.54) in the case of the Kullback-Leibler divergence, since $\phi'_{KL}(x) = \log(x)$ and comparison between both formulas shows that the left part of (1.81) can be regarded as a ϕ-scaled odds ratio. Thus, there is a connection between the ϕ-association model and objects of the form (1.81), the so called ϕ-scaled odds ratio, which will extensively be studied in Chapter 3. When the column classification variable is a binary response ($J = 2$) and the sampling scheme is the product binomial, Kateri and Agresti [2010] studied a ϕ-divergence based logit model for binary responses.

Other examples for modelling based on ϕ-divergence are the ϕ-quasi-symmetry models by Kateri and Papaioannou [1997]. Let $\mathcal{S}$ be the set of $I \times I$ multinomial probability vectors $\boldsymbol{\pi} = (\pi_{ij}) \in \Delta_{I \times I}$ with fixed sums $2\pi_{ij}^S = \pi_{ij} + \pi_{ji}$ and fixed marginals $\{\pi_{i+}\}$ (and thus $\{\pi_{+i}\}$ is also fixed.) Then $\boldsymbol{\pi}^S = (\pi_{ij}^S)$ is the probability of symmetry induced by $\boldsymbol{\pi}$. The *ϕ-quasi-symmetry model* is defined as

$$\pi_{ij} = \pi_{ij}^S (\phi')^{-1}(\alpha_i + \gamma_{ij}), \qquad i, j = 1, ..., I, \tag{1.82}$$

for parameter sets $\{\alpha_i\}$ and $\{\gamma_{ij}\}$ fulfilling $\gamma_{ij} = \gamma_{ji}$. This model is the closest to symmetry in terms of ϕ-divergence in $\mathcal{S}$, i.e. it is solution to

$$\min_{\boldsymbol{\pi} \in \mathcal{S}} \mathcal{D}^{\phi}(\boldsymbol{\pi}, \boldsymbol{\pi}^S).$$

For the Kullback-Leibler divergence (Example 1.10.2), the model (1.82) reduces to the classical model of quasi-symmetry (Section 1.9.6). Taking differences in (1.82) between cell probabilities π_{ij} and π_{ji}, and using $\gamma_{ij} = \gamma_{ji}$, gives the relation

$$\alpha_i - \alpha_j = \phi'\left(\frac{\pi_{ij}}{\pi_{ij}^S}\right) - \phi'\left(\frac{\pi_{ji}}{\pi_{ij}^S},\right), \qquad i, j = 1, ..., I, \tag{1.83}$$

which can be used to measure asymmetries in $I \times I$ contingency tables. The term on the right side of (1.83) will be studied more detailed in Chapter 6 and leads to a directed ϕ-scaled asymmetry measure. As further reference Kateri and Agresti [2006] introduced a ϕ-quasi-symmetry model for ordinal classification variables.

Chapter 2

Correction Study for the log-Odds Ratio

The odds ratio θ, as already pointed out in Section 1.7, is the most fundamental measure of association in 2×2 contingency tables, cross-classifying two binary variables. Construction of confidence intervals (CIs) for θ is thus an important task for categorical data analysis. But 2×2 contingency table are often of small sample size or have sampling zeros and thus the odds ratio may be confronted with the problem of infinite estimates and undefined or infinite CIs as described in Section 1.7.2. In addition, asymptotic methods like the Wald or score CIs presented in Section 1.7.1 are not always able to hold the coverage level. Agresti [1999] showed that the Wald CIs often behave degenerate. In the case of small sample sizes for $|\log\theta| > 4$, these CIs give coverage probability below the nominal.

Another approach to construct CIs for θ is the use of exact methods (cf. Agresti [2013, Section 16.6.4, p. 606]). Although often conservative, the exact methods ensure compliance of a given significance level (Agresti and Min [2002]). For small samples, exact methods are not always superior to asymptotic ones as discovered by Agresti and Coull [1998a].

Modern statistical science still discusses intensively on the old classical problem, which type of CIs for the odds ration is the right choice in categorical data analysis. Next to other, Brown et al. [2001] induced an intense discussion on interval estimation for binomial proportions with participation of Alan Agresti, T. Tony Cai, George Casella, Brent A. Coull, Chris Corcoran, Anirban DasGupta, Malay Ghosh, Cyrus Mehta and Thomas J. Santner.

The number of possible $I \times J$ contingency tables with sample size n is equal to

$$\binom{n + IJ - 1}{IJ - 1}$$

in the multinomial sampling scheme (cf. Diaconis and Efron [1985]). Thus the number of possible 2×2 tables with fixed n is $\frac{1}{6}(n + 3)(n + 2)(n + 1)$. In case of product binomial sampling with fixed row marginal counts, n_1 and n_2, the number of possible 2×2 tables is equal to $(n_1 + 1)(n_2 + 1)$. Although the number of possible tables is large, sampling zeros often occur in contingency tables with small sample sizes: In the product binomial sampling scheme, the relative frequency of tables with sampling zeros is $2(n_1+n_2)\cdot((n_1+1)(n_2+1))^{-1}$. For $n_1 = 10, n_2 = 10$, about 33.058% of the tables have sampling zeros when allowing zero marginal counts and about 31.933% when not allowing zero marginal counts. A classical approach to bypass the problem with sampling zeros is the use of continuity correction. Different correction techniques have been developed. The most classical is the constant correction (Anscombe [1956]; Gart [1962]; Haldane [1956]), which adds a constant to each cell. Alternatively, the single cell correction (Plackett [1962]) exclusively adds a constant to the zero cell, while the independence smoothed corrections (cf. Agresti [1999]; Bishop et al. [1975]) push the table into the direction of independence. Agresti [1999] already demonstrated the use of independence smoothed corrections to improve odds ratio CIs in terms of coverage probability. Bayesian methods (Fienberg and Holland [1970, 1972]) estimate a 'risk minimal' constant, which is added to each cell.

All mentioned correction methods are presented in the context of Bayesian statistics in a unified way. Each correction corresponds to a Dirichlet prior for the multinomial sampling scheme in 2×2 tables, which gives a conjugate family (Bishop et al. [1975, Chapter 12]). The same holds for two independent beta distributions as priors in the product binomial sampling scheme. Bayesian techniques are a common approach for inference in 2×2 table and can also be used to get posterior distribution of the odds ratio (Altham [1969]; Nurminen and Mutanen [1987]).

The correction type may heavily impact the outcome in sampling zero data sets. An extensive comparison study is essential for the different correction and CI methods for the odds ratio, because correction is often applied and improvements seem to be possible by changing the correction technique.

In this chapter, the correction techniques will be presented as Bayesian estimators for categorical data (Section 2.1). An extensive evaluation study is carried out in Section 2.2, analysing the corrected CIs in terms of coverage probability and mean length for odds ratios in 2×2 tables to give suggestions for an appropriate choice of correction technique and CI type. This is done for all CIs presented in Section 1.7.1. CIs based on the Cressie-Read statistic, CR_λ (1.28), have not been considered so far in the context of odds ratios. Results and suggestions are summarized in Section 2.3, where additional remarks on small corrections and combined corrections are also provided.

2.1 Treating Sampling Zeros: A Unified Presentation

In the following, known correction methods are described in terms of (empirical) Bayesian estimators in $I \times J$ tables. The theory is given in the more general framework although application is later done in 2×2 tables only.

Let $\boldsymbol{\pi} = (\pi_{ij}) \in \Delta_{I,J}$ be a multinomial probability vector and let $\mathbf{N} = (N_{ij}) \sim \mathcal{M}(n, \boldsymbol{\pi})$. Fienberg and Holland [1970, 1972] present (empirical) Bayesian estimators for multinomial probabilities using a Dirichlet distribution (see also Bishop et al. [1975, Chapter 12]). The conjugate prior for a multinomial distribution is the Dirichlet distribution:

$$P(\boldsymbol{\pi}|\boldsymbol{\beta}) = \frac{1}{B(\boldsymbol{\beta})} \prod_{i,j=1}^{I,J} \pi_{ij}^{\beta_{ij}-1}, \qquad \boldsymbol{\beta} = (\beta_{ij}), \beta_{ij} > 0,$$

where $B(\boldsymbol{\beta}) = \frac{\prod_{i,j=1}^{I,J} \Gamma(\beta_{ij})}{\Gamma(\sum_{i,j=1}^{I,J} \beta_{ij})}$ and Γ is the gamma function $\Gamma(y) = \int_0^\infty e^{-z} z^{y-1} dz$. The Dirichlet distribution with parameter $\boldsymbol{\beta}$ is denoted as $\mathrm{Dir}(\boldsymbol{\beta})$.

The Dirichlet distribution can be regarded as a generalisation of the Beta distribution: If $\boldsymbol{\pi} \sim \mathrm{Dir}(\boldsymbol{\beta})$ is a prior for the 2×2 table probability $\boldsymbol{\pi}$ in the multinomial sampling scheme $\mathbf{N} \sim \mathcal{M}(n, \boldsymbol{\pi})$, then conditioning on the row marginals $N_{i+} = n_i$ leads to independent $(\pi_i, 1-\pi_i) = (\frac{\pi_{i1}}{\pi_{i+}}, \frac{\pi_{i2}}{\pi_{i+}}) \sim \mathrm{Beta}(\beta_{i1}, \beta_{i2})$ as priors for the independent binomials $(N_{i1}|N_{i+} = n_i) \sim \mathcal{B}(n_i, \pi_i), i = 1, 2$.

The prior mean of the Dirichlet distribution is $\mathbb{E}(\pi_{ij}|\boldsymbol{\beta}) = \left(\frac{\beta_{ij}}{\sum_{i,j=1}^{I,J} \beta_{ij}}\right)$. The posterior distribution is the Dirichlet distribution with parameter $\mathbf{n} + \boldsymbol{\beta}$. The posterior mean with parametrization $K = K(\boldsymbol{\beta}) = \sum_{i,j=1}^{I,J} \beta_{ij}$ and $\lambda_{ij} = \lambda_{ij}(\boldsymbol{\beta}) = \frac{\beta_{ij}}{K}$ is thus for $i = 1, ..., I, j = 1, ..., J$

$$\mathbb{E}(\pi_{ij}|\mathbf{n}, \boldsymbol{\beta}) = \left(\frac{n_{ij} + \beta_{ij}}{\sum_{i,j=1}^{I,J}(n_{ij} + \beta_{ij})}\right) = \left(\frac{n}{n+K}\left(\frac{n_{ij}}{n}\right) + \frac{K}{n+K}\lambda_{ij}\right).$$

The posterior mean is a natural estimate for $\boldsymbol{\pi}$ as it minimizes the mean square error (cf. Robert [2001, Proposition 2.5.1, p. 78]). Thus, $\boldsymbol{\pi}^{\boldsymbol{\beta}}(\mathbf{n}) = \mathbb{E}(\boldsymbol{\pi}|\mathbf{n}, \boldsymbol{\beta})$ gives the Bayesian estimator of $\boldsymbol{\pi}$ under the squared error loss. With vector notation, $\mathbf{n} = (n_{ij})$ and $\boldsymbol{\lambda} = (\lambda_{ij})$, $\boldsymbol{\pi}^{\boldsymbol{\beta}}(\mathbf{n})$ obtains the form

$$\boldsymbol{\pi}^{\boldsymbol{\beta}}(\mathbf{n}) = \frac{n}{n+K}\hat{\boldsymbol{\pi}} + \frac{K}{n+K}\boldsymbol{\lambda}, \qquad K = K(\boldsymbol{\beta}), \qquad \boldsymbol{\lambda} = \boldsymbol{\lambda}(\boldsymbol{\beta}), \tag{2.1}$$

where $\hat{\boldsymbol{\pi}} = \mathbf{N}/n$ is the sample proportion. Selecting the prior parameters in (2.1) $(K, \boldsymbol{\lambda})$ gives a (Bayesian) estimator for the probability $\boldsymbol{\pi}$, which can be used to prevent zero estimates in sampling zero situations. The parameter K is the total sum of elements added to the table $\mathbf{n}$ while $\boldsymbol{\lambda}$ indicates the distribution of the elements added to the single cells. Setting $\omega = \frac{n}{n+K}$, the estimator (2.1) has the structure

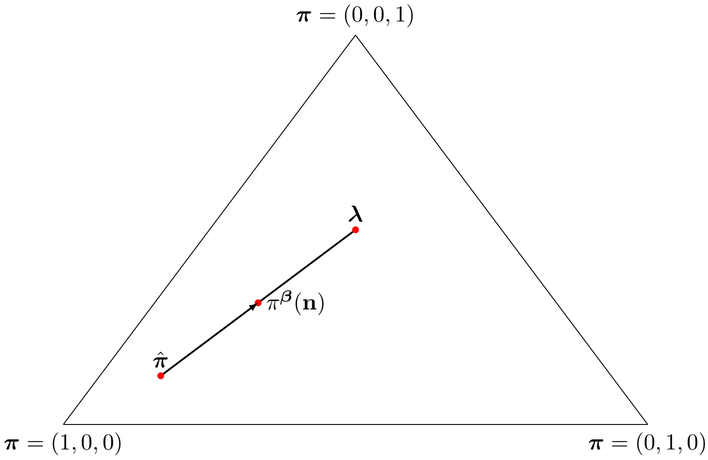
$$\boldsymbol{\pi}^{\boldsymbol{\beta}}(\mathbf{n}) = \omega\hat{\boldsymbol{\pi}} + (1-\omega)\boldsymbol{\lambda}.$$

The Bayesian estimator can be regarded geometrically as a point $\boldsymbol{\pi}^{\boldsymbol{\beta}}(\mathbf{n})$ on the line $\hat{\boldsymbol{\pi}}$ and $\boldsymbol{\lambda}$. Its position is determined by the weight ω, so the ordinary MLE $\hat{\boldsymbol{\pi}}$ is pushed forward to the probability $\boldsymbol{\lambda}$ with a weight induced by K (Figure 2.1).

The use of the Bayesian estimator $\boldsymbol{\pi}^{\boldsymbol{\beta}}(\mathbf{n})$ in (2.1) requires knowledge of the Dirichlet parameters $(K, \boldsymbol{\lambda})$ (resp. $\boldsymbol{\beta}$). Next to direct parameter choice, new estimators are achieved by letting K or $\boldsymbol{\lambda}$ depend on the data, leading to empirical Bayesian estimators. For fixed prior mean $\boldsymbol{\lambda}$, Fienberg and Holland [1972] suggested the use of

$$\hat{K} = \hat{K}(\hat{\boldsymbol{\pi}}, \boldsymbol{\lambda}) = \frac{1 - ||\hat{\boldsymbol{\pi}}||^2}{||\hat{\boldsymbol{\pi}} - \boldsymbol{\lambda}||^2} = \frac{n^2 - \sum_{i,j=1}^{I,J} n_{ij}^2}{\sum_{i,j=1}^{I,J} n_{ij}^2 - 2n\sum_{i,j=1}^{I,J} n_{ij}\lambda_{ij} + n^2\sum_{i,j=1}^{I,J} \lambda_{ij}^2}. \tag{2.2}$$

Figure 2.1: Geometric interpretation of the empirical Bayesian estimator $\boldsymbol{\pi}^{\boldsymbol{\beta}}(\mathbf{n})$ based on a Dirichlet prior with parameter $\boldsymbol{\beta}$ for a multinomial distribution with three categories.



as a quadratic risk minimal estimator of K. Notice that one can argue against empirical Bayesian techniques, since the prior information is not given via external knowledge but rather estimated using the internal information of the analysed data set. However, such estimators can turn out to be useful in certain settings.

The probability vector $\boldsymbol{\pi}^{\boldsymbol{\beta}}(\mathbf{n})$ is always a non-zero probability table, if $\boldsymbol{\lambda}$ lies in the interior of the space $\Delta_{I,J}$ and $K > 0$. In other words, the corrected vector has no zero probabilities and is thus suitable for further sampling zero sensitive analysis like association analysis based on generalised odds ratios. For further details on Bayesian estimators for multinomials and geometric interpretations, see Bishop et al. [1975, pp. 420, 12.5].

2.1.1 Cc: Constant Correction

Let $c > 0$ be a correction factor. Set $K = cIJ$ and $\boldsymbol{\lambda} = \mathbf{c}_0 = c_0 \mathbf{1}_{I\times J} > 0$ a matrix of constants. Since $\sum_{i,j=1}^{I,J} \lambda_{ij} = 1$, it holds $c_0 = 1/(IJ)$. The estimator (2.1) becomes

$$\boldsymbol{\pi}^{\boldsymbol{\beta}(\mathbf{C}c)}(\mathbf{n}) = \frac{n}{n + cIJ}\hat{\boldsymbol{\pi}} + \frac{cIJ}{n + cIJ}\mathbf{c}_0 = \frac{\mathbf{n} + c}{n + cIJ},$$

with

$$\boldsymbol{\beta}(\mathbf{C}c) := c\mathbf{1}_{I\times J}. \tag{2.3}$$

$\boldsymbol{\pi}^{\boldsymbol{\beta}(\mathbf{C}c)}(\mathbf{n})$ is thus the Bayesian estimator, which adds a constant c to each cell. The constant continuity correction introduced by Anscombe [1956], Haldane [1956] and Gart [1962] can be regarded as a Bayesian estimator with Dirichlet prior parameter $\boldsymbol{\beta}(\mathbf{C}c) > 0$. Geometrically, the Bayesian estimator $\boldsymbol{\pi}^{\boldsymbol{\beta}(\mathbf{C}c)}(\mathbf{n})$ is the MLE $\hat{\boldsymbol{\pi}}$ that is pushed towards the equiprobable model, $\pi_{ij} = 1/(IJ), i = 1, ..., I, j = 1, ..., J$, by a weight induced by c.

Remark 2.1.1. The Dirichlet distribution with constant parameter set $\boldsymbol{\beta} = \frac{1}{2}\mathbf{1}_{I\times J}$ is the Jeffrey's Prior, $\mathbf{J}(\boldsymbol{\pi}) \propto \prod_{i,j=1}^{I,J} \pi_{ij}^{1/2}$. Thus the constant continuity correction, $c = 1/2$, is the least non-informative prior. The constant correction for contingency tables has also been captured through the penalized likelihood (see Firth [1993]). Firth introduced penalized likelihoods to remove the first-order term from the bias of maximum likelihood estimates, where he suggests a penalization by the Jeffrey's prior in exponential families. Hence, the constant correction for 2×2 tables occurs in different mathematical contexts. Under this approach, the penalized MLE (pMLE) is asymptotically equivalent to the MLE, as long as the derivative of the penalty term in the likelihood is $O_P(1)$ as $n \to \infty$. Note that the pMLE is equivalent to the posterior mode estimator for an appropriate chosen prior distribution (see Tutz [2012, p. 178]).

2.1.2 DC: Data dependent weight correction

The weight K can be data dependent. Choosing $\boldsymbol{\lambda} = \mathbf{c}_0 = 1/(IJ) \cdot \mathbf{1}_{I \times J}$, the estimator (2.2) becomes

$$\hat{K} = \frac{n^2 - \sum_{i,j=1}^{I,J} n_{ij}^2}{\sum_{i,j=1}^{I,J} n_{ij}^2 - \frac{n^2}{IJ}}.$$

The empirical Bayesian estimator becomes the MLE of the data by adding $\hat{K}/(IJ)$ to each cell,

$$\boldsymbol{\pi}^{\boldsymbol{\beta}(\mathbf{DC})}(\mathbf{n}) = \frac{n}{n+\hat{K}}\hat{\boldsymbol{\pi}} + \frac{\hat{K}}{n+\hat{K}}\left(\frac{1}{IJ}\right) = \frac{\mathbf{n} + \hat{K}/(IJ)}{n+\hat{K}},$$

where

$$\boldsymbol{\beta}(\mathbf{DC}) := \hat{K}/(IJ) \cdot \mathbf{c}_0 = \hat{K} \cdot \mathbf{1}_{I \times J}. \tag{2.4}$$

This estimator occured first in Fienberg and Holland [1970, 1972] for $I \times J$ tables. It turns out, that $\hat{K}$ may adobt very large values in small tables. For example

$$\mathbf{n} = \begin{pmatrix} 6 & 4 \\ 1 & 9 \end{pmatrix}$$

gives an estimate of $\hat{K} = 7.82$. Thus, about 8 additional "fake" counts are added to the data, which is a large amount in a table with a sample size of 20.

2.1.3 Ic and DI: Independence Smoothed Estimators

Until now, the parameter $\boldsymbol{\lambda}$ within the Dirichlet distribution has been assumed to be constant. In the Bayesian context, $\boldsymbol{\lambda}$ can be interpreted as a null model, reflecting the prior believe of the analyst. One famous null model is the independence model, which is estimated using the marginal counts of $\mathbf{n} = (n_{ij})$:

$$\hat{\boldsymbol{\lambda}}^I = (\hat{\lambda}_{ij}^I), \qquad \hat{\lambda}_{ij}^I = \frac{n_{i+}n_{+j}}{n^2}, \qquad i = 1, ..., I, j = 1, ..., J.$$

Let the weight be constant, $K = cIJ, c > 0$. The empirical Bayesian estimator is the MLE of the tables with corrected cell counts, $n_{ij} + cIJ\frac{n_{i+}n_{+j}}{n^2}$,

$$\boldsymbol{\pi}^{\boldsymbol{\beta}(\mathbf{Ic})}(\mathbf{n}) = \frac{n}{n+cIJ}\hat{\boldsymbol{\pi}} + \frac{cIJ}{n+cIJ}\hat{\boldsymbol{\lambda}}^I,$$

where

$$\boldsymbol{\beta}(\mathbf{Ic}) := K \cdot \hat{\boldsymbol{\lambda}}^I = cIJ \cdot \hat{\boldsymbol{\lambda}}^I. \tag{2.5}$$

Geometrically, the table counts are corrected into the direction of independence with a weight induced by c, reasoning the nomenclature of estimators based on $\hat{\boldsymbol{\lambda}}^I$ as *independence smoothed estimators*. Agresti [1999] already studied the coverage probability of log-odds ratio asymptotic Wald CIs based on the estimator $\boldsymbol{\pi}^{\boldsymbol{\beta}(\mathbf{I}0.5)}(\mathbf{n})$ and $\boldsymbol{\pi}^{\boldsymbol{\beta}(\mathbf{C}0.5)}(\mathbf{n})$ in 2×2 contingency tables, where the independence smoothed estimator was able to improve the asymptotic Wald CI for the log-odds ratio $\log\theta$ in terms of coverage probability.

Letting K in (2.5) be data dependent using the risk optimal estimate $\hat{K} = \hat{K}(\boldsymbol{\pi}, \boldsymbol{\lambda})$ from (2.2), the empirical Bayesian estimator $\boldsymbol{\pi}^{\boldsymbol{\beta}}(\mathbf{n})$ becomes

$$\boldsymbol{\pi}^{\boldsymbol{\beta}(\mathbf{DI})}(\mathbf{n}) = \frac{n}{n+\hat{K}}\hat{\boldsymbol{\pi}} + \frac{\hat{K}}{n+\hat{K}}\boldsymbol{\lambda}^I,$$

where

$$\boldsymbol{\beta}(\mathbf{DI}) := \hat{K} \cdot \hat{\boldsymbol{\lambda}}^I, \tag{2.6}$$

which is an independence smoothed estimator with estimated weight. The correction in each cell count is $n_{ij} + \hat{K}\frac{n_{i+}n_{+j}}{n^2}, i = 1, ..., I, j = 1, ..., J$. The $\mathbf{D}$ in the nomenclature indicating data dependence of $\boldsymbol{\lambda}$. Another possibility to define $\hat{K}$ is to use $\hat{\boldsymbol{\lambda}}^I$ in (2.2). This estimator is denoted as $\hat{K}(\hat{\boldsymbol{\lambda}}^I)$. This estimator gives very large estimates, reflecting a high amount of fake data added to the table. For example,

$$\mathbf{n} = \begin{pmatrix} 4 & 6 \\ 1 & 9 \end{pmatrix}$$

gives an estimate $\hat{K}(\hat{\boldsymbol{\lambda}}^I) = 29.56$.

2.1.4 Sc: Single cell correction

In the presented Bayesian setting, single cell correction is intuitively done by setting the parameters β_{ij} (respectively λ_{ij}) equal to zero if $n_{ij} > 0$, so that the (i, j)'s cell count remains uncorrected in the posterior distribution. Single cell correction requires the introduction of improper priors (see Robert [2001, pp. 26]) since in the Dirichlet distribution the parameter assumption $\beta_{ij} > 0$ is crucial to define a probability distribution:

If the prior distribution is defined based on a σ-infinite measure on a parameter space $\mathfrak{B}$ instead of a probability measure, e.g.

$$\int_{\mathfrak{B}} \boldsymbol{\pi}(\boldsymbol{\beta})d\boldsymbol{\beta} = +\infty,$$

then $\boldsymbol{\pi}$ is called improper prior. Although $\boldsymbol{\beta}$ does not induce a probability distribution on $\boldsymbol{\pi}$, the posterior distribution $(\boldsymbol{\pi}|\mathbf{n}, \boldsymbol{\beta})$ can be well-defined, justifying thus the use of improper priors.

Robert [2001] argued from a more practical perspective because several Bayesian statisticians have not been convinced from the application of improper priors (see Lindley [1965]). He stresses that the focus of of Bayesian statistics and inference about it (p. 29):

> *"According to the Bayesian version of the Likelihood Principle, only posterior distributions are of importance. Therefore, the generalisation from proper to improper prior distributions should not cause problems, in the sense that the posterior distribution corresponding to an improper prior can be used similarly to a regular posterior distribution, when it is defined."*

Agresti [2013, 1.6.5, p.26] describes the improper prios for multinomial likelihoods: The improper prior has the same structure as die Dirichlet distribution:

$$\prod_{i,j=1}^{I,J} \pi_{ij}^{\widetilde{\beta}_{ij}-1}, \qquad \widetilde{\boldsymbol{\beta}} = (\widetilde{\beta}_{ij}), \qquad \widetilde{\beta}_{ij} \geq 0 \tag{2.7}$$

where some $\widetilde{\beta}_{ij}$ can be equal to zero. For a multinomial likelihood, the posterior distribution, if defined, has the kernel of a Dirichlet distribution:

$$\mathbf{P}(\boldsymbol{\pi}|\mathbf{n}) \propto \mathbf{P}(\mathbf{n}|\boldsymbol{\pi})\mathbf{P}(\boldsymbol{\pi}) \propto \prod_{i,j=1}^{I,J} \pi_{ij}^{n_{ij}} \prod_{i,j=1}^{I,J} \pi_{ij}^{\widetilde{\beta}_{ij}-1} = \prod_{i,j=1}^{I,J} \pi_{ij}^{n_{ij}+\widetilde{\beta}-1}.$$

Assume that $n_{ij} + \widetilde{\beta}_{ij} > 0$ for all $i = 1, .., I, j = 1, ..., J$, then the posterior distribution is well-defined:

$$\boldsymbol{\pi}|\mathbf{n} \sim \text{Dir}(\mathbf{n} + \widetilde{\boldsymbol{\beta}}).$$

The single cell correction is integrated into the Bayesian set-up by setting $\widetilde{\boldsymbol{\beta}} = c\mathbf{1}_{\{\mathbf{n}=0\}}, c > 0$ in the impoper prior (2.7), where $\mathbf{1}_{\{\mathbf{n}=0\}}$ is the matrix $(\mathbf{1}_{\{n_{ij}=0\}})$. Then clearly $\mathbf{n}+\widetilde{\boldsymbol{\beta}} > 0$, which leads to the Dirichlet posterior with the parameter $\mathbf{n} + \widetilde{\boldsymbol{\beta}}$. Setting $\hat{K}^S = c|\mathbf{1}_{\{\mathbf{n}=0\}}|$ and $\boldsymbol{\lambda} = \mathbf{1}_{\{\mathbf{n}=0\}}/|\mathbf{1}_{\{\mathbf{n}=0\}}|$, where $|\mathbf{1}_{\{\mathbf{n}=0\}}|$ stands for the number of sampling zeros in the table. The empirical Bayesian estimator is derived as

$$\boldsymbol{\pi}^{\boldsymbol{\beta}(\mathbf{Sc})}(\mathbf{n}) = \frac{n}{n + c|\mathbf{1}_{\{\mathbf{n}=0\}}|}\hat{\boldsymbol{\pi}} + \frac{c\mathbf{1}_{\{\mathbf{n}=0\}}}{n + c|\mathbf{1}_{\{\mathbf{n}=0\}}|} = \frac{\mathbf{n} + c\mathbf{1}_{\{\mathbf{n}=0\}}}{n + c|\mathbf{1}_{\{\mathbf{n}=0\}}|},$$

which is a correction of c in each zero sampled cell and

$$\boldsymbol{\beta}(\mathbf{Sc}) := c\mathbf{1}_{\{\mathbf{n}=0\}}. \tag{2.8}$$

In the estimator $\boldsymbol{\pi}^{\boldsymbol{\beta}(\mathbf{Sc})}(\mathbf{n})$, the prior parameter $\boldsymbol{\beta}(\mathbf{Sc})$ is data dependent. Geometrically, the MLE $\hat{\boldsymbol{\pi}}$ is pushed towards the equiprobable model $\mathbf{1}_{\{\mathbf{n}=0\}}/|\mathbf{1}_{\{\mathbf{n}=0\}}|$ in the orthogonal complement space with a weight induced by c. The orthogonal complement is the set of probabilities, which are non-zero in the sampling zero cells and zero otherwise. The single cell correction was suggested by Plackett [1962] for 2×2 tables. Notice that the single cell correction defined here, applies only on tables with sampling zeros, letting ordinary tables untouched.

An overview of estimators is shown in Table 2.2. There are more possibilities for selecting $(K, \boldsymbol{\lambda})$ but only the famous corrections are highlighted here. For example, the Agresti-Coull interval for binomial proportions (Agresti and Coull [1998b]) adds two successes and two failures to obtain pseudo-score CIs to adjust Wald CIs, which is covered by the presented Bayesian set-up.

Table 2.1: Short notation for the different correction methods.

Notation	Correction method
C*c*	constant correction with $c > 0$
DC	constant correction with estimated weight $\hat{K}$
I*c*	independence smoothed correction with $c > 0$
DI	independence smoothed correction with estimated weight $\hat{K}$
S*c*	single cell correction with $c > 0$

Table 2.2: Overview over different corrections for $I \times J$ tables.

Estimator	Eqn.	Correction	Reference
$\boldsymbol{\pi}^{\boldsymbol{\beta}(\mathbf{Cc})}$	(2.3)	$\mathbf{n} + c$	Anscombe [1956]; Gart [1962]; Haldane [1956]
$\boldsymbol{\pi}^{\boldsymbol{\beta}(\mathbf{DC})}$	(2.4)	$\mathbf{n} + \hat{K}/(IJ)$	Fienberg and Holland [1970, 1972]
$\boldsymbol{\pi}^{\boldsymbol{\beta}(\mathbf{Ic})}$	(2.5)	$n_{ij} + cIJn_{i+}n_{+j}/n^2$	Agresti [1999]
$\boldsymbol{\pi}^{\boldsymbol{\beta}(\mathbf{DI})}$	(2.6)	$n_{ij} + \hat{K}/(IJ) \cdot n_{i+}n_{+j}/n^2$	Bishop et al. [1975]
$\boldsymbol{\pi}^{\boldsymbol{\beta}(\mathbf{Sc})}$	(2.8)	$n_{ij} + c\mathbf{1}(n_{ij} = 0)$	Plackett [1962]

Remark 2.1.2. The correction types presented in Table 2.2 can be calculated using the R-functions given in the digital attachment (see Appendix B.2). In detail, the following R programms are used: `constant.cell.correction` for the **C***c* correction as in (2.3), `DD.const` for the **DC** correction as in (2.4), `ind.smooth` for the **I***c* correction as in (2.5), `DD.ind.smooth` for the **DI** correction as in (2.6) and `single.cell.correction` for the **S***c* correction as in (2.8).

The Wald, score, LR or CR ($\lambda \neq -1, 0$) CIs for the $\mathbf{C}c, \mathbf{DC}, \mathbf{I}c, \mathbf{DI}$ and **S***c* correction is calculated using `generate.CI.cor`.

Posterior Dirichlet parameter $\mathbf{n} + \boldsymbol{\beta}$, estimates $\log \hat{\theta}(\mathbf{n} + \boldsymbol{\beta})$, standard error estimated (SE) and the asymptotic 95 % CIs (Wald, score, CR, LR) for the presented corrections are given for Example 1.2.1 and Example 1.2.2 in Table 2.3. In sampling zero cases, the choice of the correction has a great effect on the estimates, standard errors and CIs. Due to the element $1/c$ in the SE formula, low-invasive corrections ($c = 0.1$) have an high effect on the standard error estimate, giving a value of 10.331 for Example 1.2.2 and **C**0.1, while the corresponding value for **C**0.5 is 2.315. The estimate itself is effected in the same way giving an odds ratio estimate of $\hat{\theta} = \exp(4.021) = 55.751$ in the **I**0.5 case while it is about five times greater for the less-invasive correction **I**0.1 with $\hat{\theta} = \exp(5.637) = 280.620$. The correction impact for sampling zero data sets is thus vast, leading to a rejection of the independence hypothesis for the $\mathbf{C}0.5, \mathbf{S}0.5, \mathbf{DC}$ and **DI** correction, while the other corrections cannot reject independence.

The correction has an inferior role for non-critical data sets, e.g. for data sets without sampling zeros. As analysis for Example 1.2.1 shows, the estimates and CIs are quite stable for non-zero tables under correction change. Notice that **C**0.5, **DD** and **DI** corrections give the greatest changes as they effect the posterior parameters the most, adding many fake data counts ($\approx 2, 4, 2$) to the table.

Remark 2.1.3. It is not immediately clear, that the presented correction leave the asymp-

Correction	$\mathbf{n}+\boldsymbol{\beta}$		$\log\hat{\theta}$	SE	Wald	Score	CR	LR
	Example 1.2.1 (Larynx data)							
none	21.000 15.000	2.000 3.000	0.742	0.948	[0.312, 14.152]	[0.366, 11.859]	[0.346, 13.211]	[0.311, 17.477]
C05	21.500 15.500	2.500 3.500	0.664	0.797	[0.338, 11.169]	[0.381, 9.775]	[0.365, 10.610]	[0.336, 12.919]
S05	21.000 15.000	2.000 3.000	0.742	0.948	[0.312, 14.152]	[0.366, 11.859]	[0.346, 13.211]	[0.311, 17.477]
I05	21.985 15.771	2.137 3.107	0.707	0.899	[0.316, 12.996]	[0.367, 11.031]	[0.349, 12.191]	[0.314, 15.698]
C01	21.100 15.100	2.100 3.100	0.724	0.912	[0.317, 13.412]	[0.369, 11.350]	[0.350, 12.568]	[0.317, 16.297]
S01	21.000 15.000	2.000 3.000	0.742	0.948	[0.312, 14.152]	[0.366, 11.859]	[0.346, 13.211]	[0.311, 17.477]
I01	21.197 15.154	2.027 3.021	0.735	0.937	[0.313, 13.904]	[0.366, 11.678]	[0.347, 12.983]	[0.312, 17.090]
DD	21.968 15.968	2.968 3.968	0.609	0.697	[0.358, 9.447]	[0.395, 8.496]	[0.381, 9.075]	[0.355, 10.543]
DI	22.907 16.493	2.265 3.207	0.676	0.858	[0.320, 12.079]	[0.369, 10.367]	[0.350, 11.377]	[0.317, 14.331]
	Example 1.2.2 (Prednisolone data)							
none	7.000 0.000	8.000 15.000	∞	∞	-	[2.923, ∞]	[3.293, ∞]	[5.114, ∞]
C05	7.500 0.500	8.500 15.500	3.309	2.315	[1.386, 540]	[2.242, 298]	[2.369, 444]	[2.753, 3716]
S05	7.000 0.500	8.000 15.000	3.268	2.335	[1.314, 524]	[2.124, 288]	[2.238, 429]	[2.587, 3584]
I05	7.233 0.233	8.767 15.767	4.021	4.601	[0.832, 3734]	[2.493, 1097]	[2.692, 2114]	[3.375, 590837]
C01	7.100 0.100	8.100 15.100	4.886	10.331	[0.243, 72046]	[2.742, 5517]	[3.021, 15080]	[4.087, $>8\cdot10^{10}$]
S01	7.000 0.100	8.000 15.000	4.877	10.335	[0.241, 71530]	[2.713, 5477]	[2.987, 14954]	[4.036, $>7.9\cdot10^{10}$]
I01	7.047 0.047	8.153 15.153	5.637	21.759	[0.030, $>2.6\cdot10^{6}$]	[2.820, 24025]	[3.139, 92983]	[4.477, $>3.5\cdot10^{16}$]
DD	8.243 1.243	9.243 16.243	2.455	1.095	[1.498, 91]	[1.757, 72]	[1.805, 87]	[1.940, 160]
DI	7.580 0.580	9.906 16.906	3.104	2.015	[1.380, 360]	[2.082, 216]	[2.188, 310]	[2.505, 1780]

Table 2.3: Posterior Dirichlet parameters $\mathbf{n}+\boldsymbol{\beta}$, estimate $\log\hat{\theta} = \log\hat{\theta}(\mathbf{n}+\boldsymbol{\beta})$, estimated standard error SE and 95% CI on θ-scale for different correction types for the Prednisolone and Larynx data. CIs are Wald, Score, Cressie-Read (CR, $\lambda = 2/3$) and Likelihood-Ratio (LR). Large numbers (> 70) are rounded on zero digits.

totic results of $\hat{\boldsymbol{\pi}}$ and the goodness-of-fit test statistics unaffected. All presented continuity correction can be captured through penalized likelihoods (cf. Remark 2.1.1). As noted, the pMLEs are asymptotic equivalent to the MLEs as long as the derivations of the penalty term in the likelihood are $O_P(1)$ for $n \to \infty$. Following the idea of Serfling [1980, p. 145, Theorem 4.2.2] the pMLE obtains the same asymptotic normality as the (unpenalized) MLE and the test statistics presented in Section 1.5 have the same asymptotic distribution when using the pMLEs instead of MLEs.

The $\mathbf{C}c, \mathbf{S}c$ and $\mathbf{I}c$ lead to such pMLEs. For any $\boldsymbol{\pi} = (\pi_{ij}) \in \Delta_{2\times 2}$ and any realisation $\mathbf{n} = (n_{ij})$ let $l(\boldsymbol{\pi}) = \sum_{i,j=1}^{2} n_{ij} \log \pi_{ij}$ be the log-likelihood function. The penalized log-likelihood function is defined as $l^*(\boldsymbol{\pi}) = l(\boldsymbol{\pi}) + \log \mathcal{A}(\boldsymbol{\pi})$ for a penalization term $\mathcal{A}(\boldsymbol{\pi})$. Let $\boldsymbol{\beta} = (\beta_{ij})$ be a parameter in the Dirichlet distribution for a 2×2 contingency table. The penalization term corresponding to the Dirichlet prior in the log-likelihood function is

$$\log \mathcal{A}(\boldsymbol{\pi}) = \sum_{i,j=1}^{2} \beta_{ij} \log \pi_{ij},$$

which has derivatives

$$\frac{\partial \log \mathcal{A}(\boldsymbol{\pi})}{\partial \pi_{ij}} = \frac{\beta_{ij}}{\pi_{ij}}, \qquad i, j = 1, 2. \tag{2.9}$$

By (2.3) it holds $\beta_{ij}(\mathbf{C}c) = c < 4c$ for the constant cell correction, $\beta_{ij}(\mathbf{S}c) = c1_{\{n_{ij}=0\}} < 4c$ by (2.8) for the single cell correction and $\beta_{ij}(\mathbf{I}c) = 4cn_{i+}n_{+j}/n^2 \leq 4cn^2/n^2 = 4c$ by (2.5) for the independence smoothed correction. Thus, using (2.9) it holds

$$\left| \frac{\partial \log \mathcal{A}(\boldsymbol{\pi})}{\partial \pi_{ij}} \right| \leq \frac{4c}{\pi_{ij}}$$

and the derivative of the penalization term is $O_P(1)$ for $n \to \infty$ as it does not depend on n.

The data dependent estimators using $\hat{K}$ are not bounded and hence are not $O_P(1)$ for $n \to \infty$. A simple counterexample to show unboundedness can be formulated, using the subsequence $\mathbf{n}_m = (n, n, n+1, n-1)$, $|\mathbf{n}_m| = m = 4n$ for which $\hat{K} = \hat{K}(\mathbf{n}_m) \to \infty$ if $m \to \infty$ and $\mathbf{P}(\mathbf{N}_m = \mathbf{n}_m) > 0$, $\mathbf{N}_m \sim \mathcal{M}(m, \boldsymbol{\pi})$ for all $m \in \mathbb{N}$. Using $\beta_{ij}(\mathbf{DC}) = \hat{K}/4$ from (2.4), it holds by (2.9) that $\frac{\partial \log \mathcal{A}(\boldsymbol{\pi})}{\partial \pi_{ij}} \to \infty$ with non-zero probability.

2.2 Simulation and Evaluation Study

Let $\boldsymbol{\beta}$ be the (im)proper prior parameter for a selected correction method. The asymptotic 95% CIs, $CI(\mathbf{N} + \boldsymbol{\beta}) = [CI^L, CI^U](\mathbf{N} + \boldsymbol{\beta})$, for $\hat{\theta}(\mathbf{n} + \boldsymbol{\beta})$ are evaluated for different $\boldsymbol{\beta}$ in terms of their coverage probability (CP)

$$\begin{aligned} \mathbf{P}(\theta_0 \in CI(\mathbf{N} + \boldsymbol{\beta})) &= \mathbb{E}\left(\mathbf{1}_{\{\theta_0 \in CI(\mathbf{N}+\boldsymbol{\beta})\}}\right) \\ &= \sum_{\mathbf{n} \in \mathcal{S}_0} \mathbf{1}_{\{\theta_0 \in CI(\mathbf{n}+\boldsymbol{\beta})\}} \mathbf{P}(\mathbf{N} = \mathbf{n} \,|\, n_{i+}, n_{+j} \neq 0, i, j = 1, 2), \end{aligned} \tag{2.10}$$

and mean length

$$\begin{aligned} &\mathbb{E}\left((\log CI^U - \log CI^L)(\mathbf{N} + \boldsymbol{\beta})\right) \\ &\quad = \sum_{\mathbf{n} \in \mathcal{S}_0} \left(\log CI^U - \log CI^L\right)(\mathbf{n} + \boldsymbol{\beta}) \mathbf{P}(\mathbf{N} = \mathbf{n} \,|\, n_{i+}, n_{+j} \neq 0, i, j = 1, 2), \end{aligned} \tag{2.11}$$

where θ_0 is the true odds ratio within the probability setting and $\mathcal{S}_0$ is the set of all possible 2×2 tables $\mathbf{n}$ of fixed sample size n or fixed row marginals n_1, n_2, for the multinomial or product binomial sampling scheme, respectively, with non-zero marginals. Conditioning on non-zero marginal counts enables the use and comparison of independence smoothed empirical Bayesian estimators. The mean length is calculated on log-scale to enable better comparison.

The adopted sampling scheme in this study is the product binomial. The CP and mean length are computed analytically by generating each time all possible tables $\mathbf{n}$ with non-zero row and column marginals and applying (2.10) and (2.11) directly, since the sample sizes considered are relatively small. The size of $|\mathcal{S}_0|$ gives reasoning for the choice of the binomial instead of the multinomial sampling scheme. In the following study for a correction technique with Dirichlet prior parameter $\boldsymbol{\beta}$, the corresponding continuity correction is applied to all tables $\mathbf{n} \in \mathcal{S}_0$ as given in (2.10) and (2.11). This even applies to tables without sampling zeros. Thus, the estimator $\boldsymbol{\pi}^{\boldsymbol{\beta}}$ is applied on the whole space $\mathcal{S}_0$ and the evaluation results do not get distorted by the unproblematic cases. This simulation design has already been applied by Agresti [1999] and Lui and Lin [2003]. Coverage probability and mean length can be calculated using the algorithm and R functions presented in Remark B.2.1.

Following the set-up of a study by Lui and Lin [2003], the CP and mean length for selected combinations of binomial parameters ($n_1 = n_2 = 10, 20, 30, 50$, $\theta = 1, 2, 4, 8$, $\pi_2 = 0.1, 0.2, 0.5, 0.9$) and corrections $\boldsymbol{\beta} \in \{\boldsymbol{\beta}(\mathbf{C}0.5), \boldsymbol{\beta}(\mathbf{I}0.5), \boldsymbol{\beta}(\mathbf{S}0.5)\}$ are given in Table 2.4 and Table 2.5, respectively. Wald CIs tend to give higher CP than score, LR and CR CIs. LR CI often gives CP below the nominal for small (large) success probabilities, when n_1 and θ are small. All CIs and corrections give CP closer to the 95% nominal, when the expected outcome in each cell increases, i.e. for large n_1 or $\pi_1 \approx 0.5$ and small θ. For $\theta = 1$, the CP behaves symmetric around $\pi_1 = 0.5$ with large CP, when π_1 is small (or large) and n_1 is small. Score and CR seem to perform best. **I**0.5 correction always gives larger ML compared to single cell and constant cell correction. The mean length on log-scale decreases with increasing n_1. In nearly every case, score and CR CI give smaller ML than Wald and LR CI, where the LR CI is nearly always larger than Wald CI. ML is increasing when the expected cell count is decreasing, i.e. when θ and π_1 are large.

The left (lower tail) and right (upper tail) noncoverage probabilities are given in Table 2.6. In each case considered, they sum up to $1 - CP$, when taking rounding errors into account. The noncoverage probabilities become skew for increasing θ in all cases, giving lower noncoverage probability for the lower tail compared to the upper tail. The LR seems to have a more balanced noncoverage probability, while Wald and score CIs tend to have a very low left tail noncoverage probability in the unsymmetrical cases ($\theta > 1$).

A random study for the CP and mean length of 95% CIs (Wald, Score, CR, LR) has been conducted under the product binomial sampling scheme as in Agresti [1999]. In particular, $K = 10.000$ random points $(\pi_1, \pi_2) \in (0, 1)^2$ were uniformly generated, and for each of them the associated CP and ML were computed. The plots with sample sizes $(n_1, n_2) = (10, 10)$ are presented in Figure 2.3 and Figure 2.4, for CP and ML, respectively. For increasing sample sizes, the CP of all correction types approach the nominal, as expected due to their asymptotic roots. The ML behaves similarly for **C**0.5 and **S**0.5 correction. In consistence with the previous observation (Table 2.5), the **I**0.5 gives very large MLs, especially for the Wald and LR CI. Wald and LR tend to give very high ML over all corrections and should be avoided. The CIs are very conservative and thus give high CP, when $\log\theta$ is close to zero. Wald, score and CR CI give degenerate CP ($CP < 0.85$), for constant and single cell correction when $|\log\theta| > 4$, which has already

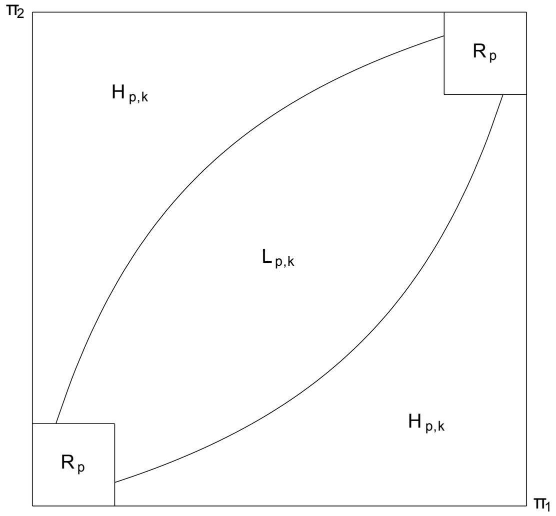


Figure 2.2: Separation of the parameter space $(0,1)^2$ of (π_1, π_2) into regions of high ($\mathbf{H}_k^p$) and low ($\mathbf{L}_k^p$) association. The regions $\mathbf{R}_k^p$ are the regions of rare probabilities.

been pointed out by Agresti [1999] for the constant cell corrected Wald CI. The LR does not produce degenerate CP for any correction at the cost of higher ML. The random study conducted for unbalanced sample ($n_1 \neq n_2$) sizes is not displayed here. The results are very similar to the balanced case ($n_1 = n_2$).

To get an overview over the whole parameter space, contour plots of the 95% Wald CP and ML for $(\pi_1, \pi_2) \in (0,1)^2$ ranging from 0.01 to 0.99 by increments 0.01, for $(n_1, n_2) = (10, 10)$ and all presented corrections (Table 2.2, $c \in \{0.1, 0.5\}$) have been made (Figures 2.5 and 2.6), which show a similar structure as for the ordinary log-odds ratio. The contour plots allow separation of the parameter space of (π_1, π_2) into regions of of rare probabilities (**R**), low to moderate association (**L**) and high association (**H**) (see Figure 2.2 and descriptions below). The rare probabilities subspace is defined by the regions of the parametric space, where both π_1 and π_2 are simultaneously very low (below p) or very high (over $1-p$), for a small probability p. Thus, $\mathbf{R}_p$ and B_k might be defined as $\mathbf{R}_p = \{(\pi_1, \pi_2) \in (0,1)^2 \mid \pi_1, \pi_2 \leq p, \quad \text{or} \quad \pi_1, \pi_2 \geq 1-p\}$ and $\mathrm{B}_k = \{(\pi_1, \pi_2) \in (0,1)^2 \mid \ |\log\theta| \leq k\}$ for a given $k \geq 0$. The subspaces of low to moderate and high association are defined as $\mathbf{L}_{p,k} = \mathbf{R}_p^c \cap \mathrm{B}_k$ and $\mathbf{H}_{p,k} = \mathbf{R}_p^c \cap \mathrm{B}_k^c$, respectively. $\mathbf{R}_p, \mathbf{L}_{p,k}$ and $\mathbf{H}_{p,k}$ separate the parameter space of (π_1, π_2) in an arbitrary way depending on the choice for p and the level of association strength k chosen for distinguishing between high and low association. For $p = 0$, $\mathbf{R}_p = \emptyset$ and the separation reduces to the high-low association subspaces.

Corrections with small correction size ($c = 0.1$) and the data dependent corrections using $\hat{K}$ (e.g. **DC** and **DI**) are given in the contour plots. They are not followed further in the study as $c = 0.1$ corrections give very large mean length, and **DC** and **DI** show a highly degenerate behaviour in terms of CP. They are far below the nominal for $|\log\theta| > 3$ and are extremely conservative for $|\log\theta| \leq 3$, which shows a high instability over the whole parameter space.

The $\mathbf{C}c, \mathbf{S}c, \mathbf{I}c, c = 0.1$ corrections give very large ML, especially at the edges when π_1 or π_2 are small (or large). Including $c = 0.5$, they also give high CP in the **R** region, i.e.

when the expected cell count is small.

The summary statistics based on the **HRL** separation for these contour plot data are given in Table 2.7. **S**0.5 gives higher CP and ML than **I**0.5 for all CIs (Wald, Score, CR, LR). The later one has always a larger ML which coincides with the previous observations. The LR CI does not have any degenerate CP ($CP < 0.85$) while this is the case for Wald and score CI for all three correction techniques ($\mathbf{C}0.5, \mathbf{S}0.5, \mathbf{I}0.5$).

Table 2.4: CP (multiplied by 1000) for the 95% CIs under the independent binomial sampling scheme ($n_1 = n_2$) conditional on $n_{+j} \neq 0, j = 1, 2$. The asymptotic CIs are (a) Wald, (b) Score, (c) CR ($\lambda = 2/3$) and (d) LR.

		$\pi_1 = 0.1$			$\pi_1 = 0.2$			$\pi_1 = 0.5$			$\pi_1 = 0.7$			$\pi_1 = 0.9$		
n_1	θ	C05	S05	I05	C05	S05	I05	C05	S05	I05	C05	S05	I05	C05	S05	I05
10	1	1000[a]	1000	1000	995	995	995	960	960	960	980	980	980	1000	1000	1000
		990[b]	999	990	970	989	970	958	959	958	963	974	963	990	999	990
		990[c]	997	990	970	975	970	958	949	958	963	949	963	990	997	990
		990[d]	988	990	970	955	970	958	948	958	963	938	963	990	988	990
20		998	998	998	978	975	980	961	961	961	968	958	968	998	998	998
		983	983	961	956	955	951	961	957	961	955	947	955	983	983	961
		983	983	961	956	955	951	961	957	961	955	947	955	983	983	961
		961	983	914	951	955	945	961	957	957	955	947	947	961	983	914
30		989	991	992	965	966	966	948	948	948	959	959	959	989	991	992
		974	961	946	957	951	951	948	948	948	958	951	951	974	961	946
		974	961	946	955	951	951	948	948	948	951	951	951	974	961	946
		959	960	925	954	941	951	948	948	948	951	949	951	959	960	925
50		979	974	980	959	957	959	943	943	943	955	955	955	979	974	980
		964	952	952	957	949	951	943	943	943	955	950	950	964	952	952
		964	942	952	952	948	951	943	943	943	955	950	950	964	942	952
		952	942	939	951	948	948	943	943	943	950	950	950	952	942	939
10	2	993	993	995	976	976	979	962	962	964	980	980	983	999	999	999
		984	984	957	966	962	961	958	945	958	970	967	961	993	993	958
		984	984	984	966	961	966	958	937	958	970	967	970	993	993	993
		984	984	950	966	961	946	958	937	953	970	967	947	993	993	957
20		983	985	986	961	961	962	950	950	950	965	965	966	996	998	998
		970	969	961	957	948	952	950	947	947	959	951	954	983	983	957
		966	969	959	952	948	946	947	947	947	954	951	948	982	983	957
		965	965	931	946	944	942	947	941	947	948	947	940	982	982	942
30		972	977	978	959	956	960	957	953	957	960	958	961	989	994	995
		961	960	951	956	946	954	953	950	953	956	946	953	974	975	959
		961	957	951	956	946	954	953	950	953	956	946	953	974	974	959
		952	945	935	950	945	947	953	950	953	950	943	946	970	970	929
50		967	966	967	955	953	956	958	954	958	957	955	959	981	982	982
		959	955	954	952	948	950	956	953	956	953	949	950	966	964	959
		957	951	953	951	948	950	956	951	956	953	949	950	961	964	952
		953	944	945	952	946	950	956	951	954	953	948	952	953	957	926
10	4	979	986	985	967	975	967	967	975	967	979	987	986	985	998	999
		965	965	965	953	950	953	953	950	953	966	966	966	985	985	985
		965	978	978	950	959	959	950	959	959	966	978	978	985	985	985
		978	981	974	959	970	946	959	970	946	978	981	975	985	985	985
20		966	976	977	965	967	968	965	967	968	966	977	977	988	992	994
		962	963	961	960	956	956	960	956	956	962	964	961	952	977	910
		963	967	959	956	946	954	956	946	954	964	968	960	977	977	977
		964	968	960	956	944	954	956	944	954	965	969	961	977	981	981
30		967	973	973	961	957	962	961	957	962	967	973	973	978	987	989
		958	966	958	953	953	953	953	953	953	958	966	958	949	951	935
		955	965	957	954	945	956	954	945	956	955	966	957	949	972	972
		957	957	940	954	947	949	954	947	949	957	959	940	972	981	980
50		966	967	965	954	954	956	954	954	956	966	968	966	966	980	981
		954	951	955	951	949	951	951	949	951	954	951	955	956	957	955
		955	953	952	951	948	954	951	948	954	955	954	952	956	971	970
		953	948	947	953	949	952	953	949	952	953	949	947	970	979	964
10	8	963	978	975	962	981	962	964	978	973	962	980	982	989	992	995
		949	965	962	947	978	960	948	967	963	951	953	953	929	929	929
		962	975	962	960	965	960	963	974	963	953	980	953	929	992	929
		963	975	974	964	965	965	964	974	974	953	980	971	929	992	931
20		965	972	970	957	970	957	966	972	969	956	967	968	977	977	981
		954	969	952	953	945	951	955	968	952	937	967	954	880	948	948
		952	967	966	951	950	952	952	965	963	954	967	967	948	977	948
		961	971	961	950	946	950	958	968	958	967	981	972	948	986	957
30		960	973	973	962	958	962	962	973	972	954	971	965	955	970	972
		952	964	955	953	943	955	951	960	955	943	965	956	903	957	957
		955	965	958	957	945	957	955	962	956	954	971	965	943	970	957
		956	953	951	951	944	951	954	950	950	964	976	968	957	983	971
50		959	968	962	955	954	955	958	966	960	956	969	966	945	966	967
		954	957	953	953	948	953	954	956	952	955	965	962	933	964	947
		951	951	953	950	951	948	951	950	952	954	967	959	943	966	964
		956	944	950	951	948	951	956	944	950	959	968	943	964	980	978

Table 2.5: Mean length on log-scale for the 95% CIs under the independent binomial sampling scheme ($n_1 = n_2$) conditional on $n_{+j} \neq 0, j = 1, 2$. The asymptotic CIs are (a) Wald, (b) Score, (c) CR ($\lambda = 2/3$) and (d) LR.

		$\pi_1 = 0.1$			$\pi_1 = 0.2$			$\pi_1 = 0.5$			$\pi_1 = 0.7$			$\pi_1 = 0.9$		
n_1	θ	C05	S05	I05	C05	S05	I05	C05	S05	I05	C05	S05	I05	C05	S05	I05
10	1	5.53[a]	6.03	10.47	4.54	5	6.1	3.53	3.75	3.56	3.92	4.27	4.32	5.53	6.03	10.47
		4.62[b]	4.91	6.39	4.03	4.34	4.67	3.39	3.58	3.41	3.65	3.9	3.84	4.62	4.91	6.39
		4.97[c]	5.34	7.19	4.24	4.61	5.03	3.46	3.67	3.49	3.77	4.06	4	4.97	5.34	7.19
		6.39[d]	7.09	18.18	4.97	5.57	8.51	3.6	3.85	3.66	4.12	4.54	4.92	6.39	7.09	18.02
20		4.35	4.80	7.31	3.20	3.44	3.51	2.48	2.55	2.49	2.73	2.86	2.79	4.35	4.80	7.31
		3.80	4.10	4.85	3.00	3.18	3.17	2.44	2.51	2.44	2.65	2.75	2.68	3.80	4.10	4.85
		4.02	4.38	5.32	3.09	3.30	3.29	2.46	2.53	2.47	2.69	2.8	2.73	4.02	4.38	5.32
		4.83	5.43	10.5	3.33	3.62	3.90	2.51	2.58	2.51	2.79	2.92	2.85	4.83	5.43	10.43
30		3.57	3.91	4.90	2.58	2.70	2.67	2.03	2.06	2.03	2.22	2.28	2.24	3.57	3.91	4.90
		3.23	3.48	3.76	2.48	2.58	2.54	2.00	2.04	2.01	2.18	2.23	2.19	3.23	3.48	3.76
		3.38	3.67	4.01	2.53	2.64	2.60	2.02	2.05	2.02	2.20	2.26	2.22	3.38	3.67	4.01
		3.83	4.27	6.18	2.64	2.78	2.76	2.04	2.07	2.04	2.25	2.31	2.27	3.83	4.27	6.16
50		2.71	2.89	2.98	1.98	2.03	2.01	1.57	1.58	1.57	1.72	1.74	1.72	2.71	2.89	2.98
		2.55	2.70	2.71	1.94	1.98	1.96	1.56	1.57	1.56	1.70	1.72	1.70	2.55	2.70	2.71
		2.63	2.79	2.81	1.96	2.01	1.99	1.56	1.58	1.56	1.71	1.73	1.72	2.63	2.79	2.81
		2.81	3.02	3.19	2.00	2.06	2.03	1.57	1.59	1.57	1.73	1.75	1.74	2.81	3.02	3.18
10	2	5.18	5.66	8.57	4.19	4.58	4.93	3.66	3.93	3.75	4.37	4.78	5.43	5.86	6.32	12.25
		4.42	4.71	5.70	3.82	4.10	4.15	3.49	3.71	3.54	3.93	4.22	4.39	4.81	5.07	7.05
		4.72	5.08	6.31	3.98	4.30	4.38	3.57	3.81	3.63	4.11	4.45	4.68	5.21	5.55	8.01
		5.89	6.54	13.93	4.49	4.98	6.06	3.77	4.08	3.93	4.74	5.26	7.06	6.89	7.55	22.13
20		3.89	4.26	5.32	2.93	3.11	3.06	2.57	2.65	2.58	3.07	3.28	3.28	4.91	5.38	10.28
		3.49	3.75	4.08	2.80	2.94	2.88	2.51	2.59	2.52	2.91	3.07	3.03	4.16	4.45	5.95
		3.66	3.96	4.37	2.86	3.02	2.96	2.54	2.62	2.55	2.98	3.16	3.12	4.46	4.82	6.66
		4.22	4.69	7.21	3.02	3.22	3.23	2.60	2.69	2.61	3.19	3.43	3.56	5.63	6.26	15.73
30		3.15	3.41	3.71	2.38	2.46	2.42	2.09	2.14	2.10	2.48	2.59	2.55	4.2	4.62	7.64
		2.93	3.12	3.20	2.31	2.38	2.34	2.06	2.11	2.07	2.40	2.49	2.45	3.67	3.95	4.84
		3.03	3.25	3.35	2.34	2.42	2.38	2.08	2.12	2.08	2.44	2.54	2.50	3.88	4.22	5.32
		3.32	3.64	4.42	2.42	2.51	2.47	2.11	2.15	2.11	2.53	2.66	2.63	4.68	5.23	10.77
50		2.40	2.52	2.53	1.83	1.87	1.84	1.62	1.64	1.62	1.91	1.95	1.93	3.27	3.57	4.49
		2.30	2.40	2.39	1.80	1.83	1.81	1.61	1.63	1.61	1.87	1.91	1.89	2.98	3.2	3.47
		2.35	2.46	2.45	1.82	1.85	1.83	1.61	1.63	1.61	1.89	1.93	1.91	3.11	3.36	3.69
		2.47	2.61	2.67	1.85	1.88	1.86	1.63	1.65	1.63	1.93	1.98	1.95	3.5	3.86	5.56
10	4	4.91	5.31	6.92	4.08	4.43	4.43	4.08	4.43	4.43	4.95	5.36	7.08	6.11	6.52	13.61
		4.27	4.53	5.11	3.77	4.04	3.95	3.77	4.04	3.95	4.29	4.56	5.18	4.96	5.18	7.57
		4.51	4.83	5.55	3.90	4.20	4.11	3.90	4.20	4.11	4.55	4.87	5.63	5.39	5.68	8.66
		5.52	6.04	10.12	4.33	4.75	5.02	4.33	4.75	5.02	5.57	6.11	10.44	7.29	7.87	25.39
20		3.63	3.94	4.38	2.86	3.02	2.93	2.86	3.02	2.93	3.67	3.99	4.50	5.42	5.81	13.25
		3.31	3.54	3.68	2.75	2.88	2.79	2.75	2.88	2.79	3.34	3.57	3.74	4.48	4.72	7.02
		3.44	3.70	3.88	2.80	2.94	2.85	2.8	2.94	2.85	3.48	3.74	3.95	4.83	5.13	7.98
		3.91	4.30	5.47	2.93	3.12	3.04	2.93	3.12	3.04	3.97	4.36	5.69	6.37	6.92	21.42
30		2.95	3.18	3.27	2.32	2.40	2.34	2.32	2.40	2.34	2.99	3.23	3.35	4.86	5.25	11.29
		2.77	2.94	2.95	2.26	2.33	2.28	2.26	2.33	2.28	2.80	2.98	3.00	4.10	4.35	6.19
		2.85	3.04	3.06	2.29	2.37	2.31	2.29	2.37	2.31	2.88	3.08	3.11	4.39	4.69	6.95
		3.10	3.37	3.70	2.35	2.44	2.38	2.35	2.44	2.38	3.15	3.43	3.82	5.61	6.13	17.41
50		2.25	2.37	2.34	1.78	1.82	1.79	1.78	1.82	1.79	2.29	2.40	2.38	4.03	4.38	7.76
		2.17	2.27	2.24	1.76	1.79	1.77	1.76	1.79	1.77	2.20	2.30	2.27	3.52	3.77	4.82
		2.21	2.31	2.28	1.77	1.81	1.78	1.77	1.81	1.78	2.24	2.35	2.32	3.71	4.01	5.29
		2.31	2.44	2.44	1.80	1.84	1.81	1.8	1.84	1.81	2.35	2.48	2.49	4.51	4.97	11.34
10	8	4.79	5.15	5.86	4.20	4.59	4.39	4.68	5.05	5.54	5.47	5.80	8.68	6.29	6.64	14.51
		4.22	4.47	4.72	3.90	4.20	4.00	4.16	4.42	4.57	4.61	4.82	5.92	5.06	5.24	7.92
		4.43	4.72	5.04	4.02	4.35	4.14	4.35	4.65	4.85	4.92	5.18	6.53	5.51	5.76	9.09
		5.34	5.79	7.60	4.45	4.91	4.80	5.18	5.64	6.95	6.34	6.78	13.8	7.56	8.07	27.57
20		3.54	3.84	3.97	2.93	3.11	2.97	3.42	3.70	3.74	4.40	4.73	6.33	5.77	6.09	15.47
		3.25	3.47	3.48	2.83	2.98	2.86	3.17	3.38	3.34	3.83	4.07	4.70	4.70	4.88	7.82
		3.36	3.61	3.64	2.88	3.04	2.90	3.27	3.51	3.47	4.04	4.31	5.08	5.09	5.33	8.95
		3.8	4.17	4.59	3.00	3.20	3.06	3.64	3.99	4.21	4.97	5.39	9.23	6.90	7.35	25.85
30		2.89	3.11	3.08	2.38	2.46	2.39	2.78	2.97	2.91	3.74	4.06	4.93	5.37	5.67	14.47
		2.72	2.89	2.84	2.33	2.41	2.33	2.64	2.79	2.72	3.34	3.58	3.92	4.43	4.62	7.33
		2.79	2.98	2.92	2.35	2.43	2.36	2.70	2.87	2.79	3.49	3.76	4.17	4.77	5.00	8.34
		3.02	3.29	3.35	2.41	2.50	2.42	2.89	3.12	3.11	4.13	4.53	6.65	6.35	6.77	23.42
50		2.2	2.31	2.26	1.83	1.87	1.83	2.12	2.21	2.16	2.93	3.19	3.43	4.74	5.04	11.66
		2.13	2.22	2.17	1.81	1.84	1.81	2.06	2.14	2.09	2.71	2.91	2.99	4.00	4.21	6.30
		2.16	2.26	2.21	1.82	1.86	1.82	2.09	2.17	2.12	2.80	3.02	3.12	4.26	4.51	7.06
		2.26	2.38	2.33	1.84	1.88	1.85	2.16	2.27	2.21	3.13	3.44	4.09	5.50	5.90	18.51

Table 2.6: Noncoverage probability for lower (left) and upper (right) tail. Set-up as in Table 2.5. Values are given in relation 1 : 1000, e.g. 1 equals 0.001.

		$\pi_1 = 0.1$			$\pi_1 = 0.2$			$\pi_1 = 0.5$			$\pi_1 = 0.7$			$\pi_1 = 0.9$		
n_1	θ	**C**0.5	**S**0.5	**I**0.5	**C**0.5	**S**0.5	**I**0.5	**C**0.5	**S**0.5	**I**0.5	**C**0.5	**S**0.5	**I**0.5	**C**0.5	**S**0.5	**I**0.5
10	1	0^a 0	0 0	0 0	3 3	3 3	3 3	20 20	20 20	20 20	10 10	10 10	10 10	0 0	0 0	0 0
		5^b 5	1 1	5 5	15 15	6 6	15 15	21 21	21 21	21 21	19 19	13 13	19 19	5 5	1 1	5 5
		5^c 5	1 1	5 5	15 15	13 13	15 15	21 21	26 26	21 21	19 19	25 25	19 19	5 5	1 1	5 5
		5^d 5	6 6	5 5	15 15	22 22	15 15	21 21	26 26	21 21	19 19	31 31	19 19	5 5	6 6	5 5
20		1 1	1 1	1 1	11 11	13 13	10 10	19 19	20 20	19 19	16 16	21 21	16 16	1 1	1 1	1 1
		9 9	9 9	20 20	22 22	23 23	25 25	20 20	21 21	20 20	23 23	27 27	23 23	9 9	9 9	20 20
		9 9	9 9	20 20	22 22	23 23	25 25	20 20	21 21	20 20	23 23	27 27	23 23	9 9	9 9	20 20
		20 20	9 9	43 43	25 25	23 23	27 27	20 20	21 21	21 21	23 23	27 27	27 27	20 20	9 9	43 43
30		5 5	5 5	4 4	17 17	17 17	17 17	26 26	26 26	26 26	21 21	21 21	21 21	5 5	5 5	4 4
		13 13	20 20	27 27	22 22	24 24	24 24	26 26	26 26	26 26	21 21	24 24	24 24	13 13	20 20	27 27
		13 13	20 20	27 27	23 23	24 24	24 24	26 26	26 26	26 26	24 24	24 24	24 24	13 13	20 20	27 27
		21 21	20 20	37 37	23 23	30 30	25 25	26 26	26 26	26 26	24 24	25 25	24 24	21 21	20 20	37 37
50		11 11	13 13	10 10	21 21	21 21	21 21	28 28	28 28	28 28	23 23	23 23	23 23	11 11	13 13	10 10
		18 18	24 23	24 24	21 21	26 24	24 24	28 28	28 28	28 28	23 23	25 25	25 25	18 18	24 23	24 24
		18 18	29 29	24 24	24 24	26 26	24 24	28 28	28 28	28 28	23 23	25 25	25 25	18 18	29 29	24 24
		24 24	29 29	30 30	24 24	26 26	26 26	28 28	28 28	28 28	25 25	25 25	25 25	24 24	29 29	30 30
10	2	0 6	0 6	0 5	3 21	3 21	1 19	10 28	10 28	8 28	2 18	2 18	1 16	0 1	0 1	0 1
		1 14	2 14	1 42	9 24	14 24	9 30	14 28	27 28	14 28	7 23	10 23	7 31	0 6	0 6	0 42
		1 14	2 14	1 14	9 24	15 24	9 24	14 28	35 28	14 28	7 23	11 23	7 23	0 6	0 6	0 6
		1 14	2 14	8 42	9 24	15 24	24 30	14 28	35 28	18 28	7 23	11 23	22 31	0 6	0 6	1 42
20		0 16	0 15	0 14	9 30	9 30	8 30	21 30	21 30	21 30	6 29	6 29	6 29	0 4	0 2	0 2
		3 27	4 27	7 32	12 31	21 31	17 31	21 30	24 30	23 30	10 31	18 31	16 31	0 17	0 17	1 42
		7 27	4 27	8 32	17 31	21 31	23 31	23 30	24 30	24 30	16 31	18 31	22 31	1 17	0 17	1 42
		8 27	8 27	37 32	23 31	25 31	27 31	24 30	29 30	24 30	22 31	22 31	29 31	1 17	1 17	15 42
30		1 27	2 21	1 21	14 27	16 27	13 27	20 24	23 24	20 24	11 28	14 28	11 28	0 11	0 6	0 5
		9 30	9 30	18 31	17 27	26 27	18 27	23 24	26 24	23 24	15 28	25 28	19 28	2 25	1 25	5 36
		9 30	13 30	18 31	17 27	27 27	18 27	23 24	26 24	23 24	15 28	26 28	19 28	2 25	2 25	5 36
		18 30	25 30	34 31	22 27	28 27	26 27	24 24	26 24	24 24	21 28	29 28	26 28	5 25	6 25	35 36
50		7 27	7 26	6 26	18 28	22 25	18 25	20 22	24 22	20 22	16 27	21 24	17 24	0 19	0 18	0 17
		14 27	18 27	19 27	19 29	25 28	23 28	22 22	25 22	22 22	17 30	24 27	23 27	5 30	6 30	10 31
		16 27	22 27	19 27	21 28	25 28	23 28	22 22	27 22	22 22	21 27	24 27	23 27	9 30	6 30	17 31
		19 27	29 27	28 27	23 25	28 25	24 25	22 22	27 22	24 22	23 24	28 24	24 24	17 30	13 30	43 31
10	4	0 21	0 13	0 15	1 32	4 21	1 31	1 32	4 21	1 31	0 21	0 13	0 14	0 15	0 2	0 1
		1 34	1 34	1 34	6 41	9 41	6 41	6 41	9 41	6 41	1 34	1 34	1 34	0 15	0 15	0 15
		1 34	1 21	1 21	9 41	9 32	9 32	9 41	9 32	9 32	1 34	1 21	1 21	0 15	0 15	0 15
		1 21	1 18	4 21	9 32	9 21	22 32	9 32	9 21	22 32	1 21	1 18	4 21	0 15	0 15	0 15
20		0 34	0 24	0 23	7 28	8 25	7 25	7 28	8 25	7 25	0 34	0 23	0 23	0 12	0 8	0 6
		3 35	3 34	3 36	12 28	17 28	17 28	12 28	17 28	17 28	2 36	3 34	3 36	0 48	0 23	0 90
		3 34	4 28	7 34	17 28	29 25	19 28	17 28	29 25	19 28	3 34	4 29	7 34	0 23	0 23	0 23
		7 28	9 24	16 24	19 25	31 25	21 25	19 25	31 25	21 25	7 29	8 24	15 24	0 23	0 19	0 19
30		1 32	1 27	0 27	12 27	16 26	12 26	12 27	16 26	12 26	0 32	1 27	0 27	0 22	0 13	0 11
		6 36	7 27	9 33	17 30	21 26	17 30	17 30	21 26	17 30	5 37	7 27	9 33	0 51	0 49	0 65
		9 36	8 27	16 27	17 30	28 26	18 26	17 30	28 26	18 26	9 37	7 27	15 27	0 51	0 28	0 28
		16 27	20 22	33 27	20 26	30 23	25 26	20 26	30 23	25 26	15 27	19 22	33 27	0 28	0 19	1 19
50		4 31	7 26	5 30	15 30	20 25	18 26	15 30	20 25	18 26	3 31	6 27	4 30	0 34	0 20	0 19
		11 35	20 30	15 31	19 30	25 26	19 30	19 30	25 26	19 30	11 36	19 30	15 31	0 44	0 43	0 44
		14 31	20 26	18 31	19 30	27 25	19 27	19 30	27 25	19 27	14 31	20 27	18 31	0 44	0 29	1 29
		21 26	29 23	28 25	21 26	28 23	22 26	21 26	28 23	22 26	21 27	28 23	29 24	1 29	0 21	15 21
10	8	0 37	0 22	0 25	0 38	0 19	0 38	0 36	0 22	0 27	0 38	0 20	0 18	0 11	0 8	0 5
		0 51	0 35	0 38	2 51	2 20	2 38	0 51	0 32	0 37	0 49	0 47	0 47	0 71	0 71	0 71
		0 38	2 22	0 38	2 38	16 19	2 38	0 37	4 22	0 37	0 47	0 20	0 47	0 71	0 8	0 71
		2 35	2 22	2 23	16 20	16 19	16 19	4 32	4 22	4 23	0 47	0 20	0 29	0 71	0 8	0 69
20		0 35	0 28	0 30	7 36	7 23	7 36	0 34	0 28	0 31	0 44	0 33	0 32	0 23	0 23	0 19
		1 45	2 28	4 44	10 37	25 30	12 37	2 42	4 28	6 42	0 63	0 33	0 46	0 120	0 52	0 52
		4 44	5 28	4 30	12 37	27 23	12 36	6 42	7 28	6 31	0 46	0 33	0 33	0 52	0 23	0 52
		11 28	7 22	11 28	20 30	36 18	20 30	14 28	11 21	14 28	0 33	0 19	0 27	0 52	0 14	0 43
30		1 39	1 26	1 27	11 27	18 24	11 27	1 37	2 26	1 27	0 46	0 29	0 35	0 45	0 30	0 28
		6 42	9 27	6 39	16 31	30 27	16 30	8 41	12 27	8 37	0 57	0 35	0 44	0 97	0 43	0 43
		6 39	9 26	10 32	16 27	30 24	16 27	8 37	12 26	11 32	0 46	0 29	0 35	0 57	0 30	0 43
		17 27	21 25	22 27	22 27	32 24	22 27	19 27	25 25	23 27	1 35	1 22	4 28	0 43	0 17	0 29
50		4 37	5 26	4 35	15 30	22 24	15 30	6 36	9 26	6 34	0 44	0 31	0 34	0 55	0 34	0 33
		9 37	16 27	13 35	16 30	24 27	17 30	10 36	18 26	13 34	1 44	1 34	4 34	0 67	0 36	0 53
		12 37	23 26	17 30	20 30	25 24	22 30	13 36	25 26	18 30	2 44	2 31	6 34	0 57	0 34	0 36
		17 26	34 22	23 26	22 27	32 21	22 27	18 26	33 22	23 26	11 31	9 24	32 25	0 36	0 20	0 22

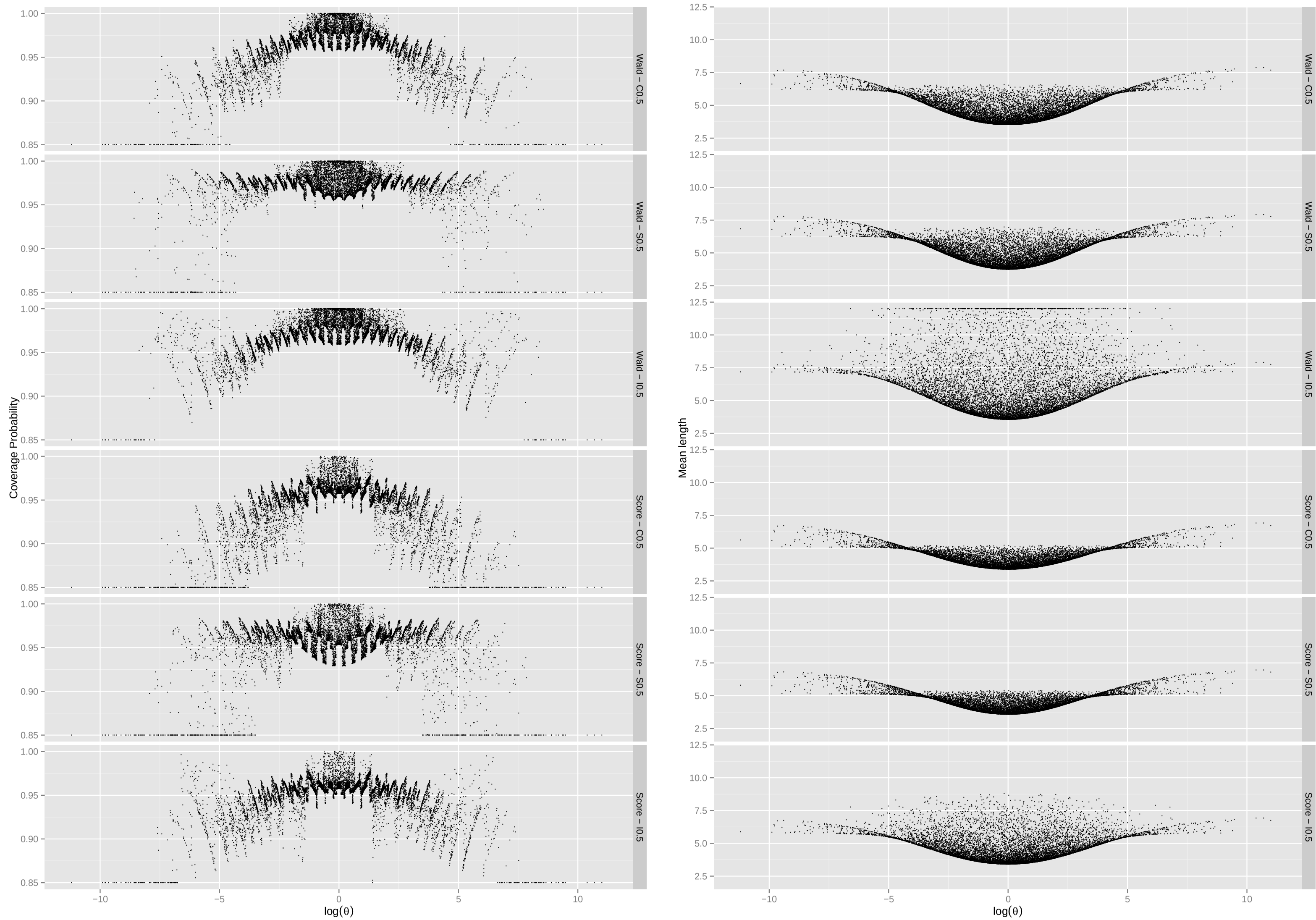


Figure 2.3: Random study with 10.000 uniformly simulated pairs (π_1, π_2) in $(0,1)^2$ giving values for the coverage probability (left) and the mean length on log-scale (right) for the different correction methods and asymptotic CIs (Wald and Score) in the independent binomial setting with $n_1 = 10$ and $n_2 = 10$. CP and mean length are cut at 0.85 and 12.5, respectively. Values on the 85% line indicate lower CP and values on the 12.5 line indicate higher mean length.

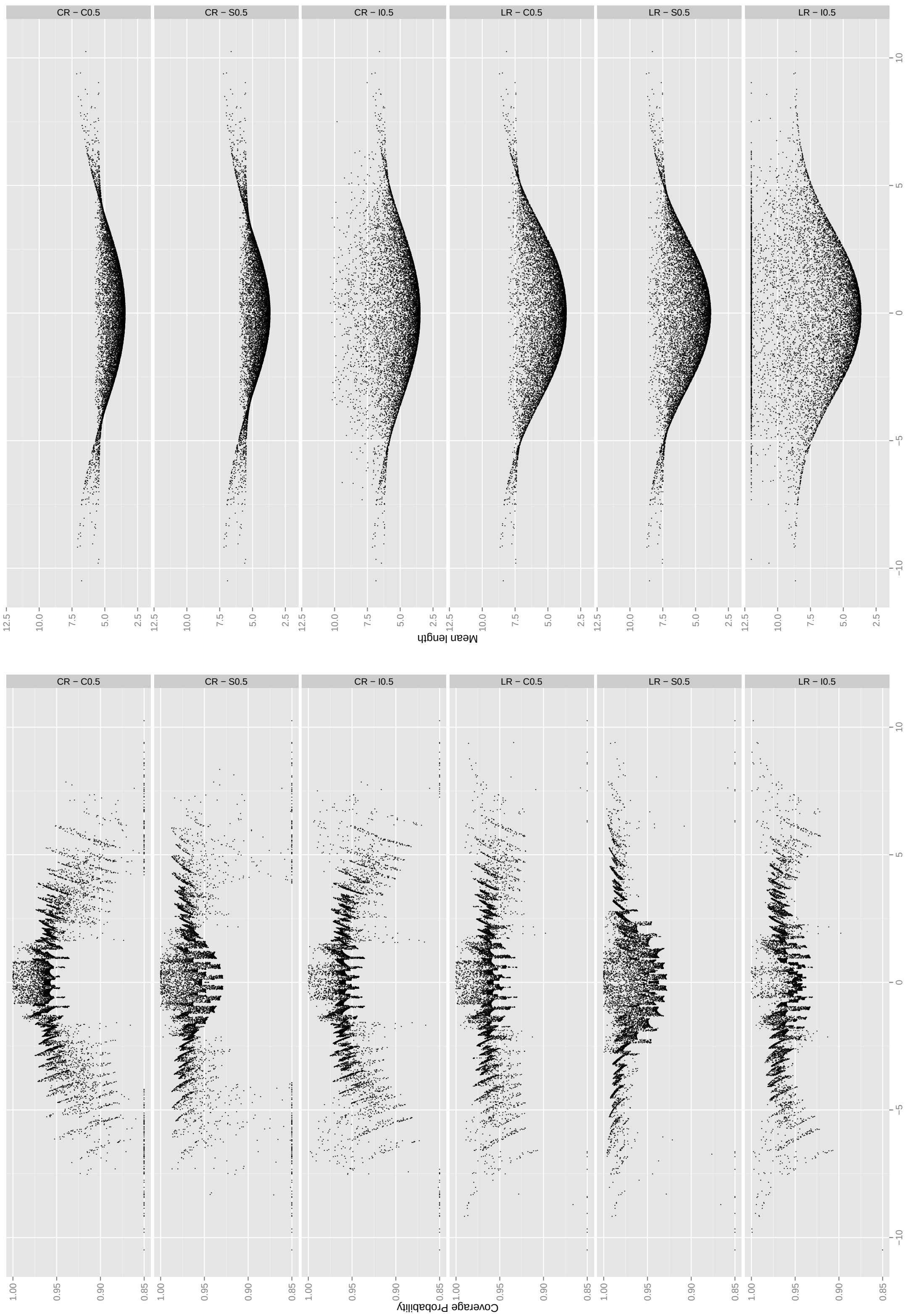


Figure 2.4: Same set-up as in Figure 2.3 for the Cressie-Read ($\lambda = 2/3$) and Likelihood Ratio CI.

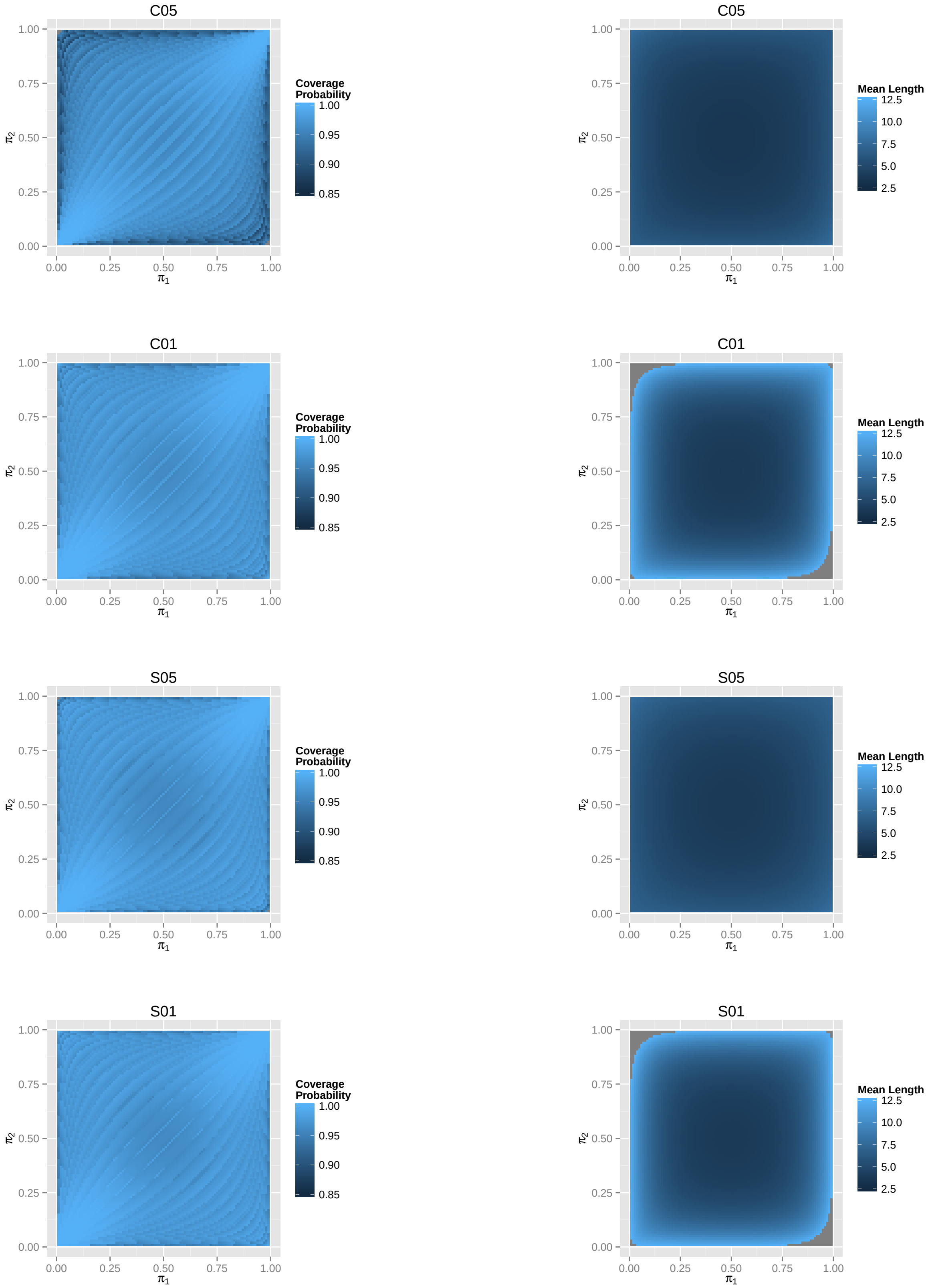


Figure 2.5: Contour plots for the CP (left) and the mean length (right) for the $\log \hat{\theta}(\mathbf{n}+\boldsymbol{\beta})$ 95% Wald CIs for the constant cell ($\mathbf{C}c$) and single cell correction ($\mathbf{S}c$) with $c \in \{0.1, 0.5\}$ (cf. Table 2.2) under the independent binomial setting with $n_1 = 10, n_2 = 10$ for π_1, π_2 ranging from 0.01 to 0.99 by 0.01.

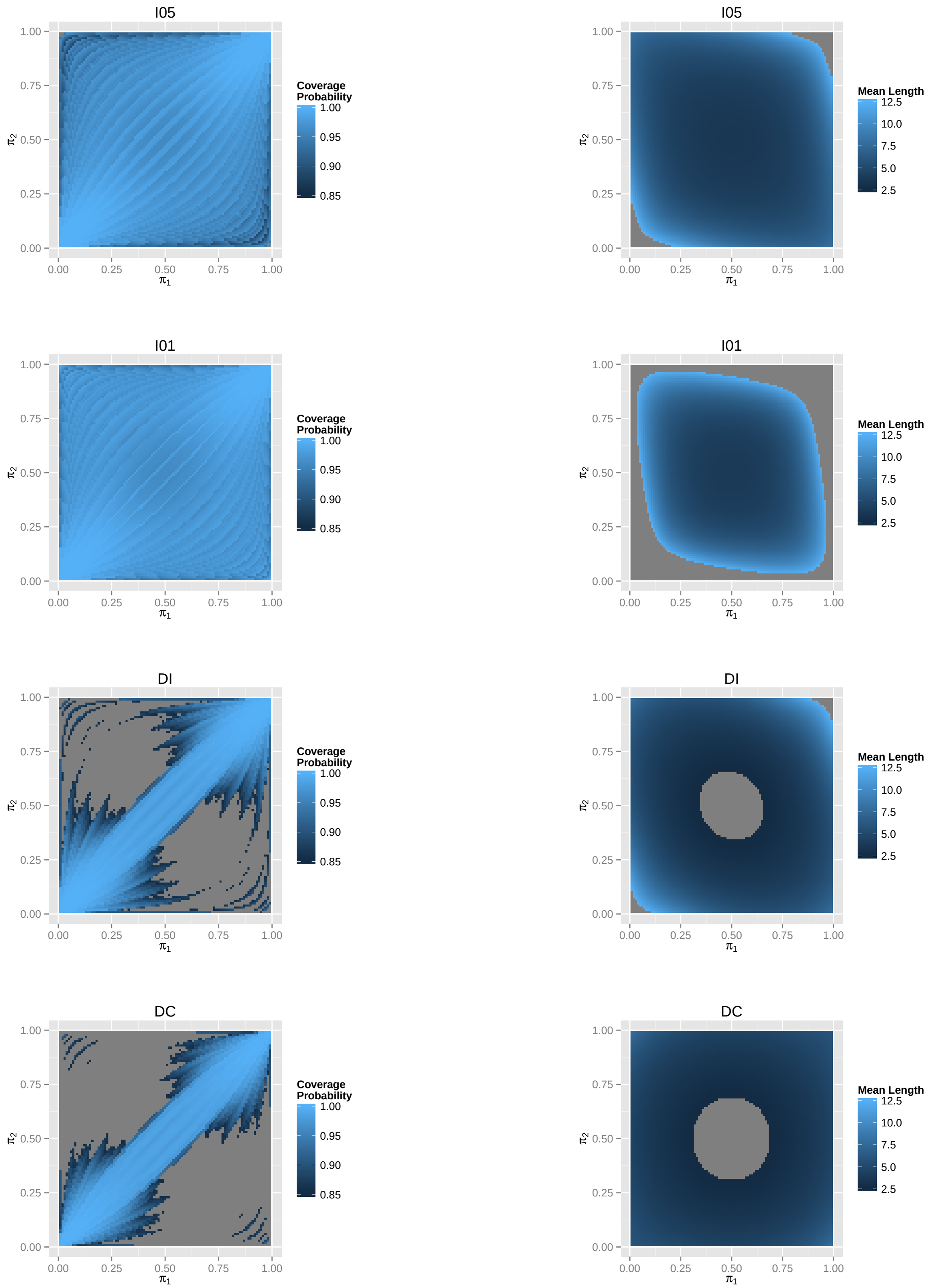


Figure 2.6: Contour plots in the same set-up as in Figure 2.5 for the independence smoothed correction (**I**c, $c \in \{0.1, 0.5\}$), the data dependent independence smoothed (**DI**) and data dependent constant cell correction (**DC**) (cf. Table 2.2).

Table 2.7: CP and mean length for the 95% CIs under the independent binomial sampling scheme ($n_1 = n_2$) conditional on $n_{+j} \neq 0, j = 1, 2$.

		Wald						Score						CR ($\lambda = 2/3$)						LR					
		Min.	1Q	Med	Mean	3Q	Max.	Min.	1Q	Med	Mean	3Q	Max.	Min.	1Q	Med	Mean	3Q	Max.	Min.	1Q	Med	Mean	3Q	Max.
		Coverage Probability																							
C0.5	full	739	957	967	964	979	1000	601	942	956	947	964	1000	667	949	958	953	964	1000	912	955	962	961	967	1000
	$R_4^{0.1}$	928	997	999	994	1000	1000	871	979	992	975	998	1000	872	980	993	978	998	1000	915	984	992	986	996	1000
	$L_4^{0.1}$	889	961	969	968	979	1000	806	948	957	953	964	995	881	953	959	957	964	995	922	957	962	961	967	995
	$H_4^{0.1}$	739	912	923	923	938	961	601	868	893	887	922	951	667	902	917	917	937	959	912	940	953	952	963	991
S0.5	full	818	969	975	975	981	1000	667	953	963	958	970	1000	739	954	966	963	974	1000	929	952	971	967	982	1000
	$R_4^{0.1}$	991	999	1000	999	1000	1000	914	991	996	990	999	1000	928	992	996	991	998	1000	982	994	996	996	998	1000
	$L_4^{0.1}$	940	969	975	976	981	1000	742	953	963	961	970	998	860	954	966	963	973	996	929	950	967	965	979	996
	$H_4^{0.1}$	818	958	972	967	978	990	667	910	955	931	968	984	739	946	969	958	977	987	974	983	987	986	990	996
I0.5	full	739	961	970	969	981	1000	601	946	956	951	962	1000	739	951	958	955	963	1000	912	952	962	961	969	1000
	$R_4^{0.1}$	996	1000	1000	1000	1000	1000	840	969	987	968	994	1000	872	970	988	975	993	1000	915	976	985	980	993	1000
	$L_4^{0.1}$	919	964	972	973	981	1000	881	950	957	955	962	992	904	954	959	958	963	992	926	952	962	960	969	989
	$H_4^{0.1}$	739	920	932	931	946	969	601	902	916	911	930	955	739	917	929	928	944	963	912	951	962	960	970	991
		Mean Length																							
C0.5	full	3.53	3.95	4.54	4.69	5.36	7.75	3.39	3.69	4.08	4.18	4.59	6.79	3.46	3.8	4.26	4.37	4.87	7.12	3.6	4.15	4.96	5.18	6.09	8.66
	$R_4^{0.1}$	5.53	5.91	6.12	6.1	6.32	6.53	4.62	4.84	4.96	4.95	5.07	5.19	4.97	5.25	5.4	5.38	5.54	5.69	6.39	6.98	7.29	7.26	7.59	7.92
	$L_4^{0.1}$	3.53	3.88	4.38	4.5	5.03	6.32	3.39	3.65	3.98	4.04	4.4	5.08	3.46	3.75	4.14	4.22	4.64	5.54	3.6	4.06	4.74	4.92	5.65	7.62
	$H_4^{0.1}$	5.54	5.85	6.03	6.15	6.34	7.75	4.85	4.98	5.13	5.27	5.46	6.79	5.11	5.29	5.41	5.56	5.74	7.12	6.16	6.73	7.07	7.07	7.29	8.66
S0.5	full	3.75	4.29	4.96	5.04	5.72	7.83	3.58	3.95	4.38	4.44	4.85	6.87	3.67	4.09	4.6	4.66	5.17	7.2	3.85	4.55	5.46	5.6	6.56	8.72
	$R_4^{0.1}$	6.03	6.38	6.56	6.54	6.7	6.94	4.91	5.1	5.19	5.18	5.27	5.38	5.34	5.59	5.71	5.7	5.81	5.97	7.09	7.64	7.93	7.9	8.16	8.54
	$L_4^{0.1}$	3.75	4.22	4.79	4.85	5.44	6.65	3.58	3.9	4.27	4.3	4.68	5.28	3.67	4.04	4.48	4.51	4.96	5.77	3.85	4.45	5.22	5.35	6.15	8.09
	$H_4^{0.1}$	5.96	6.14	6.29	6.43	6.66	7.83	5.01	5.19	5.42	5.52	5.75	6.87	5.34	5.49	5.69	5.82	6.04	7.2	6.66	7.08	7.3	7.36	7.56	8.72
I0.5	full	3.56	4.18	5.19	5.84	6.72	17.17	3.42	3.81	4.42	4.65	5.27	8.69	3.49	3.95	4.66	4.96	5.66	10.09	3.65	4.58	6.22	7.88	8.72	39.01
	$R_4^{0.1}$	10.47	12.64	13.97	13.88	15.08	17.17	6.39	7.18	7.65	7.61	8.04	8.69	7.19	8.18	8.77	8.72	9.26	10.09	18.35	24.55	27.79	27.85	30.88	39.01
	$L_4^{0.1}$	3.56	4.09	4.91	5.54	6.22	14.56	3.42	3.75	4.25	4.48	4.93	7.96	3.49	3.88	4.46	4.75	5.28	9.14	3.65	4.43	5.75	7.35	8.33	32.02
	$H_4^{0.1}$	5.76	6.36	6.76	6.87	7.14	9.82	5.01	5.36	5.57	5.61	5.8	6.81	5.25	5.65	5.92	5.97	6.19	7.28	6.57	7.46	8.1	8.52	8.74	16.39

Remark 2.2.1. (Small Corrections, $c = 0.1$)

The correction study was also conducted with smaller correction size ($c = 0.1$). Small improvements in terms of CP could be made (reducing the amount of degenerate CPs). The (mean) length of the corrected CIs is large which is increased up to factor 20 on θ-scale. Thus, the use of $c = 0.1$ corrections is not suggested. This was verified by corresponding studies which are not reported here. Since aim of the correction is to reduce the length of the CIs to an acceptable value, which cannot be achieved with $c = 0.1$, this correction is disregarded.

Remark 2.2.2. (Combining Corrections)

The interpretation of correction methods in 2×2 contingency tables as (empirical) Bayesian estimators opens the wide varying toolbox of Bayesian statistics. Regarding each correction method as a single Bayesian model for the contingency table, questions of optimal selection arise. Following the idea of prior mixture and average modelling, which is a combination on prior or posterior level, respectively, the different correction methods were combined and studied in an extensive evaluation setting as in Section 2.2. The results for combined correction are not given here. It turned out that the combination of correction methods leads to new estimators, which inherit most of the structure - and the structural deficits - of the original correction methods. The use of combined correction methods is therefore not suggested.

2.3 Discussion

The behaviour of the various CIs in terms of CP and mean length was assessed based on random studies (Figures 2.3 and 2.4) simulated over a wide range of association strengths, examples (Tables 2.3, 2.4, 2.5 and 2.6) and the summary statistics in the **HRL**-classification (Table 2.7). Standard error and consequently the length of the CIs are heavily effected by the size of the correction since the SE in sampling zero tables is of order c^{-1}. Due to the asymptotic nature of the methods, the coverage probability is closer to the nominal for increasing sample size. The score CIs give the shortest length over all different correction types, while Wald and Likelihood Ratio CIs give the longest CIs. CR and score are similar, although score outperforms the CR intervals in length.

Wald, score and CR produce degenerate CP for $|\log\theta| > 4$ for all kinds of corrections, which for the Wald CI was already pointed out by Agresti [1999]. The degeneration effect shifts to higher association levels ($|\log\theta| > 8$) with the independence correction. The LR statistic gives only few degenerate CPs at the cost of higher mean length. Wald CIs are very conservative in low association settings ($\mathbf{L}_{0.1,4}$). All CIs and corrections are very conservative for small probabilities ($\pi_1, \pi_2 \leq 0.1$ or $\pi_1, \pi_2 \geq 0.9$), reflecting the information aridity in these regions (small expected cell counts).

All studied corrections suffer high CP in the rare probability setting, $\mathbf{R}_{0.1,4}$, giving CP of at least 99%. In the $\mathbf{L}_{0.1,4}$ and $\mathbf{H}_{0.1,4}$ setting, single cell correction (**S**0.5) produces fewer CIs below the nominal level when compared to the constant cell correction (**C**0.5). This is accompanied by a slightly increased mean length. Wald, score or CR ($\lambda = 2/3$) in combination with **C**0.5 or **I**0.5 correction obtain a parabolic structure in the CP plots, reflecting increasing liberality (shrinking CP) with increasing $|\log\theta|$. In addition they always break the nominal level.

At the cost of higher mean length, the LR CI with **C**0.5 correction gives the best CP in terms of stability and nominal compliance. Choice between score and CR CIs is suggested, if mean length is important. The CR CI gives less degenerate CPs for $\mathbf{L}_{0.1,4}$, while the summary statistics for the **C**0.5 correction imply more stable CPs. The single cell correction **S**0.5 for CR gives a more conservative CI for the high association settings ($\mathbf{H}_{0.1,4}$), while the **I**0.5 for CR is a more liberal CI and can outperform the score CI in these set-ups.

Chapter 3

ϕ-Divergence based Measures of Association for 2×2 Tables

In Section 1.7.2 it has been shown that the odds ratio faces the problem of infinity estimation or infinite variance estimates when sampling zeros are present in the 2×2 table. This leads to undefined CIs or CIs with infinite length. Different continuity corrections have been studied in Chapter 2 to dispose these infinite estimates. In this chapter, a new information-theoretic approach is presented to overcome this problem. A generalisation of the log-odds ratio by replacing the log-scale through a family of scales based on the ϕ-divergence (Section 1.10) is proposed, which leads to the ϕ-scaled odds ratio. This new measure is closely connected to the ϕ-association models (cf. Section 1.11) and occurs as natural interpretation parameter. In the same way the log-odds ratio appears in loglinear models.

The aim of this new family is to construct new measure of association, that is more compatible with sampling zeros and enlarge the toolbox of association measurement. The ϕ-scaled odds ratio already occurred in Rom and Sarkar [1992] for the power divergence family and for general $\phi \in \Phi$ in Kateri and Papaioannou [1995], but its properties have not been studied in detail, yet.

In Section 3.1, the ϕ-scaled odds ratio is introduced and its properties are discussed. For the Kullback-Leibler divergence (Example 1.10.2), the ϕ_{KL}-scaled odds ratio turns out to be the classical log-odds ratio. Special attention is given to the power divergence (Example 1.10.4). The asymptotic normality of the maximum likelihood estimator (MLE) of the ϕ-scaled odds ratio is proved in Section 3.2 and enables asymptotic inference for this new association measure. This is done for the case of multinomial as well as product binomial sampling scheme. The boundedness of the newly introduced association measures and the effect of a sampling zero on their estimates are discussed in Section 3.3. The invertibility is analysed in Section 3.4. Focusing on the ϕ-scaled odds ratios based on the power divergence for characteristic choices of λ, MLEs and Wald asymptotic CIs for Examples 1.2.1 and 1.2.2 together with a ϕ-based relative association parameter for interpretation are provided in Section 3.5. The role of λ for the behaviour of the CI is investigated in an extensive evaluation study for various table structure scenarios in Section 3.6. It is known that in cases of high association settings with low success probabilities asymptotic CIs for the log-odds ratios are of low coverage (Agresti [1999]). In this set-up, the measure corresponding to the power divergence resulting from $\lambda = 1/3$ performs better and is thus a promising alternative. The results are summarized and shortly commented in Section 3.7. The relative risk is closely related to the odds ratio (Section 1.8). Such a connection also holds on ϕ-scale, which leads to the ϕ-scaled relative risk. A brief outlook is presented in Section 3.7.1.

3.1 The ϕ-scaled Odds Ratio

Let X and Y denote the binary row and column classification variables, respectively. Let $\mathbf{n} = (n_{ij})$ be a 2×2 contingency table of counts, which are observed under a multinomial sampling scheme with sample size n and probability matrix $\boldsymbol{\pi} = (\pi_{ij}) \in \Delta_{2\times 2}$. The log-odds ratio (Section 1.7) specifies with (1.19) uniquely the underlying probability table $\boldsymbol{\pi}$, when the marginal distributions π_{1+} and π_{+1} are given. The information actually required for completely specifying a 2×2 probability table additionally to $(\pi_{1+},\ \pi_{+1})$ is a measure of the underlying association of the form

$$\theta^g(\boldsymbol{\pi}) = g\left(\frac{\pi_{11}}{\pi_{1+}\pi_{+1}}\right) + g\left(\frac{\pi_{22}}{\pi_{2+}\pi_{+2}}\right) - g\left(\frac{\pi_{12}}{\pi_{1+}\pi_{+2}}\right) - g\left(\frac{\pi_{21}}{\pi_{2+}\pi_{+1}}\right) .$$

The terms $\pi_{ij}/\pi_{i+}\pi_{+j}$ are called *association factors* and were first used by Good [1956]. Thus, $\theta^g(\boldsymbol{\pi})$ compares the divergence of the cell probabilities from independence by scaling the association factors through a strictly convex function g. Is this departure measured in terms of the ϕ-divergence (Section 1.10) for a differentiable strictly convex ϕ-function, then it can be proved that $g = \phi'$. This is done through the ϕ-association models for $I \times J$ tables (Kateri and Papaioannou [1995]; for $I = J = 2$, the model is saturated).

Hence, the ϕ-scaled odds ratio for a strictly convex ϕ is defined as

$$\theta^\phi(\boldsymbol{\pi}) = \sum_{i=j} \phi'\left(\frac{\pi_{ij}}{\pi_{i+}\pi_{+j}}\right) - \sum_{i\neq j} \phi'\left(\frac{\pi_{ij}}{\pi_{i+}\pi_{+j}}\right), \quad i,j = 1,2. \tag{3.1}$$

(3.1) equals the intrinsic association parameter φ of the ϕ-association model (Kateri and Papaioannou [1995]) for the corresponding 2×2 probability table, when the scores assigned to the rows and columns of the table are one unit apart. The ϕ-scaled odds ratio is estimated by $\hat{\theta}^\phi(\mathbf{n}) = \theta^\phi(\hat{\boldsymbol{\pi}}) = \theta^\phi(\mathbf{n}/n)$ (cp. Casella and Berger [2002, Theorem 7.2.10, p. 320]). Let $\boldsymbol{\pi}^I = (\pi^I_{ij}) = (\pi_{i+}\pi_{+j})$ be the corresponding probability of independence based on $\boldsymbol{\pi}$. A useful representation of (3.1) to interpret the structure is

$$\theta^\phi(\boldsymbol{\pi}) = \sum_{i=j} \phi'\left(\frac{\pi_{ij}}{\pi^I_{ij}}\right) - \sum_{i\neq j} \phi'\left(\frac{\pi_{ij}}{\pi^I_{ij}}\right), \quad i,j = 1,2.$$

The association factors $\frac{\pi_{ij}}{\pi^I_{ij}} = \frac{\pi_{ij}}{\pi_{i+}\pi_{+j}}$ measure the divergence from independence in cell $(i,j), i,j = 1,2$. These are scaled by ϕ'. Since $\phi \in \Phi$ is convex and assumed as differentiable, the derivative ϕ' is monotonically increasing, so that ϕ' can be used as a scale function. An increase of probability in cell $(1,1)$ or $(2,2)$ corresponds to an increase of association since the terms $\phi'\left(\frac{\pi_{11}}{\pi^I_{11}}\right)$ and $\phi'\left(\frac{\pi_{22}}{\pi^I_{22}}\right)$ are added positively. The terms $\phi'\left(\frac{\pi_{12}}{\pi^I_{12}}\right)$

resp. $\phi'\left(\frac{\pi_{21}}{\pi_{21}^{I}}\right)$ are added negatively to reflect the decrease of association when the probability in cell $(1,2)$ or $(2,1)$ increases. An overview for selected ϕ-divergences is given in Appendix A.6.

For the Kullback-Leibler divergence (Example 1.10.2) the ϕ_{KL}-scaled odds ratio becomes the log-odds ratio:

$$\theta^{\phi_{KL}}(\boldsymbol{\pi}) = \log \frac{\pi_{11}}{\pi_{1+}\pi_{+1}} - \log \frac{\pi_{12}}{\pi_{1+}\pi_{+2}} - \log \frac{\pi_{21}}{\pi_{2+}\pi_{+1}} + \log \frac{\pi_{22}}{\pi_{2+}\pi_{+2}} = \log \frac{\pi_{11}\pi_{22}}{\pi_{12}\pi_{21}}.$$

Thus, (3.1) can be regarded as a generalisation of the log-odds ratio and a selection $\phi \in \Phi$ determines the scale in which the divergence from independence is measured.

The ϕ-scaled odds ratio $\theta^{\phi}(\boldsymbol{\pi})$ is estimated by $\hat{\theta}^{\phi} = \theta^{\phi}(\hat{\boldsymbol{\pi}})$ by applying the invariance property of maximum likelihood estimators. For the power divergence (Example 1.10.4), the ϕ_λ-scaled odds ratio ($\lambda \neq 0, -1$) becomes

$$\theta^{\phi_\lambda}(\boldsymbol{\pi}) = \frac{1}{\lambda}\left[\left(\frac{\pi_{11}}{\pi_{1+}\pi_{+1}}\right)^{\lambda} - \left(\frac{\pi_{12}}{\pi_{1+}\pi_{+2}}\right)^{\lambda} - \left(\frac{\pi_{21}}{\pi_{2+}\pi_{+1}}\right)^{\lambda} + \left(\frac{\pi_{22}}{\pi_{2+}\pi_{+2}}\right)^{\lambda}\right]. \tag{3.2}$$

The power divergence odds ratio θ^{ϕ_λ} occurred first in Rom and Sarkar [1992] in the context of $I \times J$ tables for forming generalised local odds ratios. However, their approach was different and their focus was on developing association models for $I \times J$ tables. For $\lambda = 1$, the power divergence reduces to the Pearson divergence (1.71). The associated θ^{ϕ_1} becomes a linear combination of the association factors and thus gives a linear measure of association.

So far, the 2×2 table of counts is observed by a multinomial sampling scheme. Quite often, the underlying sampling scheme is product binomial, i.e. the row marginals are fixed ($n_{1+} = n_1$, $n_{2+} = n_2$) and the observed 'success' cell counts n_{i1}, $i = 1, 2$, are realisations of independent binomial distributions $N_{i1} \sim \mathcal{B}(n_i, \pi_i)$. Obviously, the cell counts of the second column are the corresponding 'failures' $n_{i2} = n_i - n_{i1}$, $i = 1, 2$. In this case, the ϕ-scaled odds ratio is defined as

$$\theta^{\phi}_{\mathcal{B}}(\pi_1, \pi_2, r) = \phi'\left(\frac{\pi_1}{\pi}\right) - \phi'\left(\frac{1-\pi_1}{1-\pi}\right) - \phi'\left(\frac{\pi_2}{\pi}\right) + \phi'\left(\frac{1-\pi_2}{1-\pi}\right), \tag{3.3}$$

where $r = n_2/n_1$ and $\pi = \frac{n_1\pi_1 + n_2\pi_2}{n_1+n_2} = \frac{\pi_1 + r\pi_2}{1+r}$. $\theta^{\phi}_{\mathcal{B}}(\pi_1, \pi_2, r)$ is based on the departure of the product binomial probability table that is under consideration, i.e. the table with rows $(\pi_i,\ 1-\pi_i)$, $i = 1, 2$, from the table of equal success probabilities $\pi = \pi_1 = \pi_2$. In an analogous manner, (3.1) is based on the departure of the multinomial probability table from independence. Analogous to the connection between (3.1) and the ϕ-association model, $\theta^{\phi}_{\mathcal{B}}$ in (3.3) is connected to the ϕ-divergence based binary response model (Kateri and Agresti [2010]). This model is saturated for the present case of a single binary explanatory variable. In particular, $\theta^{\phi}_{\mathcal{B}}$ equals the coefficient of slope of this model, when the two levels of the explanatory variable are one unit apart. If $\mathbf{n}$ is a 2×2 table observed under a product binomial sampling scheme with fixed row marginals $n_{i+} = n_i$ $(= 1, 2)$, then the sample version of $\theta^{\phi}_{\mathcal{B}}$ is $\hat{\theta}^{\phi}_{\mathcal{B}} = \theta^{\phi}_{\mathcal{B}}(\hat{\pi}_1, \hat{\pi}_2, r) = \theta^{\phi}_{\mathcal{B}}(n_{11}/n_1, n_{21}/n_2, r) = \theta^{\phi}_{\mathcal{B}}(\mathbf{n})$.

The ϕ-scaled odds ratio is sampling scheme invariant. In particular, the following Lemma holds:

Lemma 3.1.1. Let $\mathcal{B}(n_i, \pi_i)$, $i = 1, 2$, be two independent binomial distributions underlying a 2×2 contingency table, with $n_2/n_1 = r$. If $\boldsymbol{\pi}$ is the 2×2 probability table of a multinomial distribution with $\pi_{i1}/\pi_{i+} = \pi_i$, $i = 1, 2$, and $\pi_{2+}/\pi_{1+} = r$, then

$$\theta^{\phi}_{\mathcal{B}}(\pi_1, \pi_2, r) = \theta^{\phi}(\boldsymbol{\pi}) \ .$$

Proof. Due to $\pi_{i1}/\pi_{i+} = \pi_i$, $i = 1, 2$, it follows that

$$\frac{\pi_{ij}}{\pi_{i+}\pi_{+j}} = \frac{\pi_i \mathbf{1}_{\{j=1\}} + (1-\pi_i)\mathbf{1}_{\{j=2\}}}{\pi_{+j}} \ , \quad i, j = 1, 2,$$

where $\mathbf{1}_{\{\cdot\}}$ is the indicator function, and hence

$$\begin{aligned}
\theta^{\phi}(\boldsymbol{\pi}) &= \phi'\left(\frac{\pi_{11}}{\pi_{1+}\pi_{+1}}\right) - \phi'\left(\frac{\pi_{12}}{\pi_{1+}\pi_{+2}}\right) - \phi'\left(\frac{\pi_{21}}{\pi_{2+}\pi_{+1}}\right) + \phi'\left(\frac{\pi_{22}}{\pi_{2+}\pi_{+2}}\right) \\
&= \phi'\left(\frac{\pi_1}{\pi_{+1}}\right) - \phi'\left(\frac{1-\pi_1}{\pi_{+2}}\right) - \phi'\left(\frac{\pi_2}{\pi_{+1}}\right) + \phi'\left(\frac{1-\pi_2}{\pi_{+2}}\right) = \theta^{\phi}_{\mathcal{B}}(\pi_1, \pi_2, r).
\end{aligned}$$

The last equation is due to the fact that $\pi = \pi_{+1}$, since

$$\pi = \frac{n_1\pi_1 + n_2\pi_2}{n_1 + n_2} = \pi_{1+}\pi_1 + \pi_{2+}\pi_2 = \pi_{11} + \pi_{21} = \pi_{+1} \ .$$

□

Consequently, if $\mathbf{n}_1$ and $\mathbf{n}_2$ are 2×2 frequencies tables observed under a multinomial and a product binomial sampling scheme, respectively, with $\mathbf{n}_1 = \mathbf{n}_2$, then $\hat{\theta}^{\phi}(\mathbf{n}_1) = \hat{\theta}^{\phi}_{\mathcal{B}}(\mathbf{n}_2)$.

Remark 3.1.2. The estimate $\theta^{\phi}(\hat{\boldsymbol{\pi}})$ can be calculated using the R-function `por` in the digital attachment (see Appendix B.3). By Lemma 3.1.1, it can also be used to calculate the product binomial estimate of $\theta^{\phi}_{\mathcal{B}}(\pi_1, \pi_2, r)$.

It can be proved that the ϕ-scaled odds ratio θ^{ϕ} inherits most properties from the log-odds ratio (Proposition 1.7.1) as stated in the next proposition. Due to Lemma 3.1.1, the results stated for the θ^{ϕ} in the sequel also hold for $\theta^{\phi}_{\mathcal{B}}$.

Proposition 3.1.3. The ϕ-scaled odds ratio θ^{ϕ} defined in (3.1) for a strictly convex ϕ, exhibits the following properties:

(i) $\theta^{\phi} = 0$ if and only if X and Y are independent.

(ii) For fixed row and column marginals, θ^{ϕ} is increasing in π_{11}.

(iii) By altering the entries of a single column (row) and keeping the column (row) marginals fixed, θ^{ϕ} is monotonic in the row (column) marginals difference.

(iv) θ^{ϕ} is invariant under table rotation.

(v) θ^{ϕ} changes sign when rows or columns are interchanged (antisymmetry)

(vi) θ^{ϕ} is inversion invariant, e.g. invariant under changing row and column entries.

Hence, the standard properties of the classical log-odds ratio are fulfilled by the ϕ-scaled odds ratio as well. The property of the log-odds ratio that does not hold for the ϕ-scaled odds ratio is the scaling invariance under row and column multiplication (cf. Proposition 1.7.1 (vii)).

Proof. **(i)** Let $\boldsymbol{\pi} \neq \boldsymbol{\pi}^I$, where $\boldsymbol{\pi}^I = (\pi_{i+}\pi_{+j})$ is the probability matrix under independence, having the same marginals as $\boldsymbol{\pi} = (\pi_{ij})$. Then, it can easily be verified that $\pi_{ij} = \pi_{ij}^I + q\,(\mathrm{I}(i=j) - \mathrm{I}(i \neq j))$, with $q \in \mathbb{R}$, satisfying

$$-\min(\pi_{1+}\pi_{+1}, \pi_{2+}\pi_{+2}) < q < \min(\pi_{1+}\pi_{+2}, \pi_{2+}\pi_{+1})$$

and $q \neq 0$ when $\boldsymbol{\pi} \neq \boldsymbol{\pi}^I$. For $q > 0$ and $i,j = 1,2$, it holds

$$\frac{\pi_{ij}}{\pi_{ij}^I} = \begin{cases} 1 + \frac{q}{\pi_{ij}^I} > 1, & i = j, \\ 1 - \frac{q}{\pi_{ij}^I} < 1, & i \neq j. \end{cases}$$

and by the strict convexity of ϕ it follows

$$\theta^\phi(\boldsymbol{\pi}) = \left[\phi'\left(\frac{\pi_{11}}{\pi_{11}^I}\right) + \phi'\left(\frac{\pi_{22}}{\pi_{22}^I}\right)\right] - \left[\phi'\left(\frac{\pi_{12}}{\pi_{12}^I}\right) + \phi'\left(\frac{\pi_{21}}{\pi_{21}^I}\right)\right] > 2\phi'(1) - 2\phi'(1) = 0.$$

Similarly, $\theta^\phi(\boldsymbol{\pi}) < 0$ for $q < 0$ and overall $\theta^\phi(\boldsymbol{\pi}) \neq 0$ for $\boldsymbol{\pi} \neq \boldsymbol{\pi}^I$. The other direction is obvious.

(ii) Let $\Delta_{2\times 2}(\pi_{1+}, \pi_{+1})$ be the set of 2×2 probability tables $\boldsymbol{\pi} = (\pi_{ij})$ of positive entries with fixed marginals $\pi_{1+}, \pi_{+1} \in (0,1)$. Each element $\boldsymbol{\pi} \in \Delta_{2\times 2}(\pi_{1+}, \pi_{+1})$ can be represented using π_{11}, since $\pi_{12} = \pi_{1+} - \pi_{11}$, $\pi_{21} = \pi_{+1} - \pi_{11}$ and $\pi_{22} = 1 - (\pi_{1+} + \pi_{+1}) + \pi_{11}$. Applying $0 < \pi_{ij} < 1, i,j = 1,2$, the bounds for π_{11} are

$$\underline{\pi}_{11} := \max(0, \pi_{1+} + \pi_{+1} - 1) < \pi_{11} < \min(\pi_{1+}, \pi_{+1}) =: \overline{\pi}_{11}.$$

Furthermore, θ^ϕ is a function of π_{11}, e.g. $\theta^\phi(\boldsymbol{\pi}) = \theta^\phi(\pi_{11}|\pi_{1+}, \pi_{+1})$, and for strictly convex ϕ, it is monotonically increasing in π_{11}, since the derivative

$$\begin{aligned} \frac{\partial \theta^\phi(\pi_{11}|\pi_{1+}, \pi_{+1})}{\partial \pi_{11}} &= \frac{1}{\pi_{11}^I}\phi''\left(\frac{\pi_{11}}{\pi_{11}^I}\right) + \frac{1}{\pi_{12}^I}\phi''\left(\frac{\pi_{1+} - \pi_{11}}{\pi_{12}^I}\right) \\ &+ \frac{1}{\pi_{21}^I}\phi''\left(\frac{\pi_{+1} - \pi_{11}}{\pi_{21}^I}\right) + \frac{1}{\pi_{22}^I}\phi''\left(\frac{1 - (\pi_{1+} + \pi_{+1}) + \pi_{11}}{\pi_{22}^I}\right) \end{aligned}$$

is always positive for strictly convex ϕ ($\phi'' > 0$).

(iii) Consider a table having a column and the column marginals fixed, e.g. of the form

$$\boldsymbol{\pi}^{c,1} = \begin{pmatrix} \pi_{11} & \pi_{12} - c \\ \pi_{21} & \pi_{22} + c \end{pmatrix} \quad \text{or} \quad \boldsymbol{\pi}^{c,2} = \begin{pmatrix} \pi_{11} - c & \pi_{12} \\ \pi_{21} + c & \pi_{22} \end{pmatrix},$$

for some c with $0 < c < \min(\pi_{1j}, 1 - \pi_{2j})$ for $j = 1,2$. Then, $\pi_{2+}^{c,j} - \pi_{1+}^{c,j} = \pi_{2+} - \pi_{1+} + 2c$ for $j = 1,2$. By the strict convexity of ϕ, one can show $\partial\theta^\phi(\boldsymbol{\pi}^{c,1})/\partial c > 0$ ($\partial\theta^\phi(\boldsymbol{\pi}^{c,2})/\partial c < 0$) and $\theta^\phi(\boldsymbol{\pi}^{c,1})$ ($\theta^\phi(\boldsymbol{\pi}^{c,2})$) is increasing (decreasing) in c. The proof for the case of fixed row marginals is analogous.

(iv) It holds

$$\theta^\phi(\boldsymbol{\pi}^T) = \phi'\left(\frac{\pi_{11}}{\pi_{1+}\pi_{+1}}\right) - \phi'\left(\frac{\pi_{21}}{\pi_{2+}\pi_{+1}}\right) - \phi'\left(\frac{\pi_{12}}{\pi_{1+}\pi_{+2}}\right) + \phi'\left(\frac{\pi_{22}}{\pi_{2+}\pi_{+2}}\right) = \theta^\phi(\boldsymbol{\pi}).$$

Table 3.1: Properties of the ϕ-scaled odds ratio.

Property	log-scale	ϕ-scale
Sample scheme independence (estimate)	yes	yes
Sample scheme independence (variance estimate)	yes	**no**
Quality criteria (**measure of association**)	yes	yes
Symmetric under variable permutation	yes	yes
Row or column scaling invariance	yes	**no**
Antisymmetric under row/column permutation	yes	yes
Inversion invariance	yes	yes

(v) Let $\mathbf{S} \in \mathbb{R}^{2\times 2}$ be the matrix with 1's in the off-diagonals and 0's in the diagonal, e.g. $\mathbf{S}$ is the permutation matrix.

$$\theta^{\phi}(\mathbf{S}\boldsymbol{\pi}) = \phi'\left(\frac{\pi_{21}}{\pi_{2+}\pi_{+1}}\right) - \phi'\left(\frac{\pi_{22}}{\pi_{2+}\pi_{+2}}\right) - \phi'\left(\frac{\pi_{11}}{\pi_{1+}\pi_{+1}}\right) + \phi'\left(\frac{\pi_{12}}{\pi_{1+}\pi_{+2}}\right) = -\theta^{\phi}(\boldsymbol{\pi}).$$

In addition, it holds $\theta^{\phi}(\boldsymbol{\pi}\mathbf{S}) = -\theta^{\phi}(\boldsymbol{\pi})$.

(vi) The inversion invariance follows directly from the antisymmetry under row and column permutation by $\theta^{\phi}(\mathbf{S}\boldsymbol{\pi}\mathbf{S}) = -\theta^{\phi}(\mathbf{S}\boldsymbol{\pi}) = \theta^{\phi}(\boldsymbol{\pi})$. □

Proposition 3.1.3 shows that the ϕ-scaled odds ratio inherits most of the structure of the log-odds ratio (cp. Table 3.1 for an overview). The properties of Proposition 3.1.3 (i), (ii) and (iii) justify to call the ϕ-scaled odds ratio a measure of association. One property lost during the generalisation is the scaling invariance, e.g. the invariance when multiplying a row or column with a number. Take for example the following data sets:

$$\mathbf{n}_1 = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \quad \text{and} \quad \mathbf{n}_2 = \begin{pmatrix} 2 & 6 \\ 6 & 12 \end{pmatrix} ,$$

such that $\mathbf{n}_2$ is $\mathbf{n}_1$ where the first column is multiplied by 2 and the second is multiplied by 3. It holds for the Pearson divergence $\phi_P(x) = \frac{1}{2}(x-1)^2$:

$$\hat{\theta}^{\phi_P}(\mathbf{n}_1) = -0.3968 \neq -0.3912 = \hat{\theta}^{\phi_P}(\mathbf{n}_2).$$

Note, that this property holds for the log-odds ratio, e.g. $\hat{\theta}^{\phi_{KL}}(\mathbf{n}_1) = -0.4055 = \hat{\theta}^{\phi_{KL}}(\mathbf{n}_2)$, whenever there exists a diagonal matrix $\mathbf{R} = \operatorname{diag}(k_1, k_2)$ with entries $k_1, k_2 > 0$ such that $\mathbf{n}_2 = \mathbf{R}\mathbf{n}_1$, which follows directly using the log-odds ratio:

$$\hat{\theta}^{\phi_{KL}}(\mathbf{n}_2) = \hat{\theta}^{\phi_{KL}}(\mathbf{R}\mathbf{n}_1) = \log\frac{k_1 n_{11} k_2 n_{22}}{k_1 n_{12} k_2 n_{21}} = \log\frac{n_{11}n_{22}}{n_{12}n_{21}} = \hat{\theta}^{\phi_{KL}}(\mathbf{n}_1).$$

The properties of Proposition 3.1.3 will only hold for strictly convex $\phi \in \Phi$. A counterexample would be $\phi(x) = 0, x \in (0, \infty)$, where Proposition 3.1.3 part (i) is clearly not fulfilled. The strict convexity of ϕ does not hold in general. For example, the ϕ-divergence used by Balakrishnan and Sanghvi [1968] is defined as $\phi(x) = \frac{(x-1)^2}{(x+1)^2}$, which has a second derivative $\phi''(x) = -\frac{8(x-2)}{(x+1)^4}$. Thus, ϕ is strictly convex on $[0, 2)$ and strictly concave on $(2, \infty)$ with a turning point at $x = 2$.

The ϕ-divergence is invariant for affine equivalent functions. This relation extends to ϕ-scaled odds ratios:

Lemma 3.1.4. Let $\phi \in \Phi$ be differentiable and define $\widetilde{\phi}(t) = \phi(t) + c(t-1)$ for some $c \in \mathbb{R}$ ($\widetilde{\phi}$ is an affine equivalent function of ϕ). It holds

$$\theta^{\widetilde{\phi}} = \theta^{\phi}.$$

Proof. For $x \in (0,\infty)$ it holds $\widetilde{\phi}'(x) = \phi'(x) + c$, such that for any 2×2 probability vector $\boldsymbol{\pi} = (\pi_{ij})$ it holds

$$\theta^{\widetilde{\phi}}(\boldsymbol{\pi}) = \phi'\left(\frac{\pi_{11}}{\pi_{11}^I}\right) + c - \phi'\left(\frac{\pi_{12}}{\pi_{12}^I}\right) - c - \phi'\left(\frac{\pi_{21}}{\pi_{21}^I}\right) - c + \phi'\left(\frac{\pi_{22}}{\pi_{22}^I}\right) + c = \theta^{\phi}(\boldsymbol{\pi}).$$

□

Remark 3.1.5. (Reordering Effect)

Changing the scale of the ϕ-scaled odds ratio will in general change the order of the set of possible tables when ordered in terms of θ^{ϕ}: Let $\mathbf{n}_1, \mathbf{n}_2$ be two tables with the same row marginal counts. For $\phi \in \Phi$ define an order $\leq^{\phi}$ by

$$\mathbf{n}_1 \leq^{\phi} \mathbf{n}_2 \qquad \Leftrightarrow \qquad \theta^{\phi}(\mathbf{n}_1) \leq \theta^{\phi}(\mathbf{n}_2).$$

For example let ϕ_{KL} be the Kullback-Leibler divergence (Example 1.10.2) and let ϕ_P be the Pearson divergence (Example 1.10.3). Observe the elements

$$\mathbf{n}_1 = \begin{pmatrix} 2 & 3 \\ 4 & 1 \end{pmatrix} \quad \text{and} \quad \mathbf{n}_2 = \begin{pmatrix} 6 & 4 \\ 9 & 1 \end{pmatrix} \;,$$

Due to $\frac{2 \cdot 1}{3 \cdot 4} = \frac{1}{6} = \frac{6 \cdot 1}{4 \cdot 9}$ it holds $\mathbf{n}_1 =^{\phi_{KL}} \mathbf{n}_2$ since $\theta^{\phi_{KL}}(\mathbf{n}_1) = \theta^{\phi_{KL}}(\mathbf{n}_2)$, but

$$\theta^{\phi_P}(\mathbf{n}_1) = -1.667 \qquad < \qquad -1.6 = \theta^{\phi_P}(\mathbf{n}_2)$$

and so $\mathbf{n}_1 <^{\phi_P} \mathbf{n}_2$. The ϕ-scale gives different weights to the divergence of independence in the single cells and thus values the association of tables in a differently.

3.2 Asymptotic Inference

For the classical $\log(\theta)$, a large sample Wald confidence interval (CI) is obtained straightforward, based on the asymptotic normality of $\log(\hat{\theta})$, derived by the delta method (Theorem 1.4.3). Regarding the ϕ-scaled odds ratio, its MLE is also asymptotically normal distributed but with an asymptotic variance that depends on the underlying sampling scheme as it is stated in the next theorems.

Theorem 3.2.1. Let $\phi \in \Phi$ be twice differentiable and strictly convex on $(0,\infty)$ and $\mathbf{N}$ be a random 2×2 contingency table of counts, multinomial distributed $\mathcal{M}(n, \boldsymbol{\pi})$, with $\nabla\theta^{\phi}(\boldsymbol{\pi}) \neq 0$. Let further $\hat{\theta}^{\phi} = \theta^{\phi}(\hat{\boldsymbol{\pi}})$ be the MLE of $\theta^{\phi}(\boldsymbol{\pi})$, where $\hat{\boldsymbol{\pi}} = \mathbf{N}/n$. Then, $\hat{\theta}^{\phi}$ is asymptotically normal distributed

$$\sqrt{n}(\theta^{\phi}(\hat{\boldsymbol{\pi}}) - \theta^{\phi}(\boldsymbol{\pi})) \xrightarrow{d} \mathcal{N}(0, \sigma_{\phi}^2(\boldsymbol{\pi})) \qquad \text{for } n \to \infty,$$

with

$$\sigma_{\phi}^2(\boldsymbol{\pi}) = \sum_{i,j=1}^{2} \pi_{ij} \left(\frac{\partial\theta^{\phi}(\boldsymbol{\pi})}{\partial\pi_{ij}}\right)^2 - \left(\sum_{i,j=1}^{2} \pi_{ij} \left(\frac{\partial\theta^{\phi}(\boldsymbol{\pi})}{\partial\pi_{ij}}\right)\right)^2. \tag{3.4}$$

Proof. Let $\boldsymbol{\pi} = (\pi_{ij})$ be the 2×2 probability. By Lemma 1.4.2 the MLE $\hat{\boldsymbol{\pi}}$ of $\boldsymbol{\pi}$ obtains asymptotic normality

$$\sqrt{n}(\hat{\boldsymbol{\pi}} - \boldsymbol{\pi}) \xrightarrow{d} \mathcal{N}(0, \boldsymbol{\Sigma}(\boldsymbol{\pi})) \qquad \text{for } n \to \infty,$$

with $\boldsymbol{\Sigma}(\boldsymbol{\pi}) = \text{diag}(\boldsymbol{\pi}) - \boldsymbol{\pi}^T\boldsymbol{\pi}$. The partial derivatives $\frac{\partial \theta^\phi(\boldsymbol{\pi})}{\partial \pi_{ij}}, i, j = 1, 2$ exist since $\phi \in \Phi$ is two times differentiable. Using the short notation $\psi_{ij} = \frac{\pi_{ij}}{\pi_{i+}\pi_{+j}}$, they become:

$$\begin{aligned}
\frac{\partial \theta^\phi(\boldsymbol{\pi})}{\partial \pi_{11}} &= \frac{\pi_{1+}\pi_{+1} - \pi_{11}(\pi_{1+} + \pi_{+1})}{(\pi_{1+}\pi_{+1})^2}\phi''(\psi_{11}) + \frac{\pi_{12}\pi_{+2}}{(\pi_{1+}\pi_{+2})^2}\phi''(\psi_{12}) + \frac{\pi_{21}\pi_{2+}}{(\pi_{2+}\pi_{+1})^2}\phi''(\psi_{21}) \\
&= \left(\frac{\psi_{11}}{\pi_{11}} - \frac{\psi_{11}}{\pi_{1+}} - \frac{\psi_{11}}{\pi_{+1}}\right)\phi''(\psi_{11}) + \frac{\psi_{12}}{\pi_{1+}}\phi''(\psi_{12}) + \frac{\psi_{21}}{\pi_{+1}}\phi''(\psi_{21}) \\
&= \left(\frac{1}{\pi_{11}} - \frac{1}{\pi_{1+}} - \frac{1}{\pi_{+1}}\right)\psi_{11}\phi''(\psi_{11}) + \frac{\psi_{12}}{\pi_{1+}}\phi''(\psi_{12}) + \frac{\psi_{21}}{\pi_{+1}}\phi''(\psi_{21}) \qquad (3.5)\\
\frac{\partial \theta^\phi(\boldsymbol{\pi})}{\partial \pi_{12}} &= -\frac{\pi_{11}\pi_{+1}}{(\pi_{1+}\pi_{+1})^2}\phi''(\psi_{11}) - \frac{\pi_{1+}\pi_{+2} - \pi_{12}(\pi_{1+} + \pi_{+2})}{(\pi_{1+}\pi_{+2})^2}\phi''(\psi_{12}) - \frac{\pi_{22}\pi_{2+}}{(\pi_{2+}\pi_{+2})^2}\phi''(\psi_{22}) \\
&= -\frac{\psi_{11}}{\pi_{1+}}\phi''(\psi_{11}) - \left(\frac{\psi_{12}}{\pi_{12}} - \frac{\psi_{12}}{\pi_{+2}} - \frac{\psi_{12}}{\pi_{1+}}\right)\phi''(\psi_{12}) - \frac{\psi_{22}}{\pi_{+2}}\phi''(\psi_{22}) \\
&= -\frac{\psi_{11}}{\pi_{1+}}\phi''(\psi_{11}) - \left(\frac{1}{\pi_{12}} - \frac{1}{\pi_{+2}} - \frac{1}{\pi_{1+}}\right)\psi_{12}\phi''(\psi_{12}) - \frac{\psi_{22}}{\pi_{+2}}\phi''(\psi_{22}) \qquad (3.6)\\
\frac{\partial \theta^\phi(\boldsymbol{\pi})}{\partial \pi_{21}} &= -\frac{\pi_{11}\pi_{1+}}{(\pi_{1+}\pi_{+1})^2}\phi''(\psi_{11}) - \frac{\pi_{2+}\pi_{+1} - \pi_{21}(\pi_{2+} + \pi_{+1})}{(\pi_{2+}\pi_{+1})^2}\phi''(\psi_{21}) - \frac{\pi_{22}\pi_{+2}}{(\pi_{2+}\pi_{+2})^2}\phi''(\psi_{22}) \\
&= -\frac{\psi_{11}}{\pi_{+1}}\phi''(\psi_{11}) - \left(\frac{\psi_{21}}{\pi_{21}} - \frac{\psi_{21}}{\pi_{+1}} - \frac{\psi_{21}}{\pi_{2+}}\right)\phi''(\psi_{21}) - \frac{\psi_{22}}{\pi_{2+}}\phi''(\psi_{22}) \\
&= -\frac{\psi_{11}}{\pi_{+1}}\phi''(\psi_{11}) - \left(\frac{1}{\pi_{21}} - \frac{1}{\pi_{+1}} - \frac{1}{\pi_{2+}}\right)\psi_{21}\phi''(\psi_{21}) - \frac{\psi_{22}}{\pi_{2+}}\phi''(\psi_{22}) \qquad (3.7)\\
\frac{\partial \theta^\phi(\boldsymbol{\pi})}{\partial \pi_{22}} &= \frac{\pi_{12}\pi_{1+}}{(\pi_{1+}\pi_{+2})^2}\phi''(\psi_{12}) + \frac{\pi_{21}\pi_{+1}}{(\pi_{2+}\pi_{+1})^2}\phi''(\psi_{21}) + \frac{\pi_{2+}\pi_{+2} - \pi_{22}(\pi_{2+} + \pi_{+2})}{(\pi_{2+}\pi_{+2})^2}\phi''(\psi_{22}) \\
&= \frac{\psi_{12}}{\pi_{+2}}\phi''(\psi_{12}) + \frac{\psi_{21}}{\pi_{2+}}\phi''(\psi_{21}) + \left(\frac{\psi_{22}}{\pi_{22}} - \frac{\psi_{22}}{\pi_{+2}} - \frac{\psi_{22}}{\pi_{2+}}\right)\phi''(\psi_{22}) \\
&= \frac{\psi_{12}}{\pi_{+2}}\phi''(\psi_{12}) + \frac{\psi_{21}}{\pi_{2+}}\phi''(\psi_{21}) + \left(\frac{1}{\pi_{22}} - \frac{1}{\pi_{+2}} - \frac{1}{\pi_{2+}}\right)\psi_{22}\phi''(\psi_{22}) \qquad (3.8)
\end{aligned}$$

Since it was assumed that $\nabla\theta^\phi(\boldsymbol{\pi}) \neq 0$, one can apply delta method (Theorem 1.4.3) to get asymptotic normality

$$\sqrt{n}(\theta^\phi(\hat{\boldsymbol{\pi}}) - \theta^\phi(\boldsymbol{\pi})) \xrightarrow{d} \mathcal{N}(0, \sigma_\phi^2(\boldsymbol{\pi})) \qquad \text{for } n \to \infty,$$

with $\sigma_\phi^2(\boldsymbol{\pi}) = \nabla\theta^\phi(\boldsymbol{\pi})^T\Sigma(\boldsymbol{\pi})\nabla\theta^\phi(\boldsymbol{\pi})$, which is equal to (3.4). □

The asymptotic normality of the MLE of θ^ϕ under the product binomial sampling is obtained analogously:

Theorem 3.2.2. Let $\phi \in \Phi$ be twice differentiable and strictly convex on $(0, \infty)$ and N_{i1} be independent binomial distributed random variable $\mathcal{B}(n_i, \pi_i)$, $i = 1, 2$, with $n_2/n_1 = r$.

Let further $\hat{\theta}_{\mathcal{B}}^{\phi} = \theta_{\mathcal{B}}^{\phi}(\hat{\pi}_1, \hat{\pi}_2, r)$ be the MLE of $\theta_{\mathcal{B}}^{\phi}$, where $\hat{\pi}_i = N_{i1}/n_i$, $i = 1, 2$. Then, for $n_1, n_2 \to \infty$ with $\lim_{n_1, n_2 \to \infty} \frac{n_2}{n_1} = r$, $\hat{\theta}_{\mathcal{B}}^{\phi}$ obtains asymptotic normality

$$\sqrt{n_1}\left(\theta_{\mathcal{B}}^{\phi}(\hat{\pi}_1, \hat{\pi}_2, r) - \theta_{\mathcal{B}}^{\phi}(\pi_1, \pi_2, r)\right) \xrightarrow{d} \mathcal{N}(0, \sigma_{\mathcal{B},\phi}^2(\pi_1, \pi_2, r)) \ ,$$

where

$$\sigma_{\mathcal{B},\phi}^2(\pi_1, \pi_2, r) = \pi_1(1-\pi_1)\left(\frac{\partial \theta_{\mathcal{B}}^{\phi}(\pi_1, \pi_2, r)}{\partial \pi_1}\right)^2 + \frac{\pi_2(1-\pi_2)}{r}\left(\frac{\partial \theta_{\mathcal{B}}^{\phi}(\pi_1, \pi_2, r)}{\partial \pi_2}\right)^2 . \quad (3.9)$$

Proof. For the independent random sample proportions $\hat{\pi}_i = N_{i1}/n_i, i = 1, 2$, it holds $\sqrt{n_i}\,(\hat{\pi}_i - \pi_i) \xrightarrow{d} \mathcal{N}(0, \pi_i(1-\pi_i)), n_i \to \infty$. Multiplying $\sqrt{n_2}\,(\hat{\pi}_2 - \pi_2)$ by the fixed $r^{-1/2}$, it follows

$$\sqrt{\frac{n_2}{r}}\,(\hat{\pi}_2 - \pi_2) \ \xrightarrow{d}\ \mathcal{N}\left(0, \frac{\pi_2(1-\pi_2)}{r}\right) \quad \text{for} \quad n \to \infty.$$

Using the independence of $\hat{\pi}_1, \hat{\pi}_2$, and the fact that $\sqrt{\frac{n_2}{r}} = \sqrt{n_1}$, the vector $(\hat{\pi}_1, \hat{\pi}_2)$ obtains for $n_1, n_2 \to \infty$ with $\lim_{n_1, n_2 \to \infty} \frac{n_2}{n_1} = r$, a bivariate asymptotic normal distribution

$$\sqrt{n_1}\left(\begin{pmatrix}\hat{\pi}_1\\ \hat{\pi}_2\end{pmatrix} - \begin{pmatrix}\pi_1\\ \pi_2\end{pmatrix}\right) \ \xrightarrow{d}\ \mathcal{N}\left(0, \begin{pmatrix}\pi_1(1-\pi_1) & 0\\ 0 & \frac{\pi_2(1-\pi_2)}{r}\end{pmatrix}\right) \quad \text{for} \quad n \to \infty. \quad (3.10)$$

Applying the notation $\pi_{i+} = \frac{n_i}{n_1+n_2}, i = 1, 2$, such that $\pi = \pi_1\pi_{1+} + \pi_2\pi_{2+}$, it holds for the derivatives

$$\begin{aligned}
\frac{\partial \theta_{\mathcal{B}}^{\phi}(\pi_1, \pi_2, r)}{\partial \pi_1} &= \frac{\pi - \pi_1\pi_{1+}}{\pi^2}\phi''\left(\frac{\pi_1}{\pi}\right) + \frac{(1-\pi) - (1-\pi_1)\pi_{1+}}{(1-\pi)^2}\phi''\left(\frac{1-\pi_1}{1-\pi}\right)\\
&\quad + \frac{\pi_2\pi_{1+}}{\pi^2}\phi''\left(\frac{\pi_2}{\pi}\right) + \frac{(1-\pi_2)\pi_{1+}}{(1-\pi)^2}\phi''\left(\frac{1-\pi_2}{1-\pi}\right)\\
&= \frac{\pi_2\pi_{2+}}{\pi^2}\phi''\left(\frac{\pi_1}{\pi}\right) + \frac{(1-\pi_2)\pi_{2+}}{(1-\pi)^2}\phi''\left(\frac{1-\pi_1}{1-\pi}\right) \qquad (3.11)\\
&\quad + \frac{\pi_2\pi_{1+}}{\pi^2}\phi''\left(\frac{\pi_2}{\pi}\right) + \frac{(1-\pi_2)\pi_{1+}}{(1-\pi)^2}\phi''\left(\frac{1-\pi_2}{1-\pi}\right)\\
\frac{\partial \theta_{\mathcal{B}}^{\phi}(\pi_1, \pi_2, r)}{\partial \pi_2} &= -\frac{\pi_1\pi_{2+}}{\pi^2}\phi''\left(\frac{\pi_1}{\pi}\right) - \frac{(1-\pi_1)\pi_{2+}}{(1-\pi)^2}\phi''\left(\frac{1-\pi_1}{1-\pi}\right)\\
&\quad - \frac{\pi - \pi_2\pi_{2+}}{\pi^2}\phi''\left(\frac{\pi_2}{\pi}\right) - \frac{(1-\pi) - (1-\pi_2)\pi_{2+}}{(1-\pi)^2}\phi''\left(\frac{1-\pi_2}{1-\pi}\right)\\
&= -\frac{\pi_1\pi_{2+}}{\pi^2}\phi''\left(\frac{\pi_1}{\pi}\right) - \frac{(1-\pi_1)\pi_{2+}}{(1-\pi)^2}\phi''\left(\frac{1-\pi_1}{1-\pi}\right) \qquad (3.12)\\
&\quad - \frac{\pi_1\pi_{1+}}{\pi^2}\phi''\left(\frac{\pi_2}{\pi}\right) - \frac{(1-\pi_1)\pi_{1+}}{(1-\pi)^2}\phi''\left(\frac{1-\pi_2}{1-\pi}\right).
\end{aligned}$$

Furthermore, since $\phi'' > 0$ (ϕ is strictly convex), $\frac{\partial \theta_{\mathcal{B}}^{\phi}(\pi_1, \pi_2, r)}{\partial \pi_1} > 0$ and $\frac{\partial \theta_{\mathcal{B}}^{\phi}(\pi_1, \pi_2, r)}{\partial \pi_2} < 0$. Thus, $\nabla\theta_{\mathcal{B}}^{\phi} \neq 0$, and by applying the delta method (Theorem 1.4.3) the asymptotic variance of

$\theta^{\phi}_{\mathcal{B}}(\hat{\pi}_1, \hat{\pi}_2, r)$ is derived as

$$\sigma^2_{\mathcal{B},\phi}(\pi_1, \pi_2, r) = \nabla\theta^{\phi}_{\mathcal{B}}(\pi_1, \pi_2, r)^T \begin{pmatrix} \pi_1(1-\pi_1) & 0 \\ 0 & \frac{\pi_2(1-\pi_2)}{r} \end{pmatrix} \nabla\theta^{\phi}_{\mathcal{B}}(\pi_1, \pi_2, r) \ ,$$

which leads to (3.9). □

Lemma 3.2.3. Consider the set-up of Theorem 3.2.1. Under independence, i.e. for $\boldsymbol{\pi} = \boldsymbol{\pi}^I = (\pi_{i+}\pi_{+j})$, it holds

$$\sigma^2_{\phi}(\boldsymbol{\pi}^I) = \phi''(1)^2 \sum_{i,j=1}^{2} \frac{1}{\pi^I_{ij}} \tag{3.13}$$

Proof. (3.5) to (3.8) lead to

$$\sum_{i,j=1}^{2} \pi_{ij} \left(\frac{\partial\theta^{\phi}(\boldsymbol{\pi})}{\partial\pi_{ij}}\right) = -\psi_{11}\phi''(\psi_{11}) + \psi_{12}\phi''(\psi_{12}) + \psi_{21}\phi''(\psi_{21}) - \psi_{22}\phi''(\psi_{22}), \tag{3.14}$$

which for $\pi_{ij} = \pi^I_{ij} = \pi_{i+}\pi_{+j}$ is equal to 0, since $\psi_{ij} = \frac{\pi_{ij}}{\pi_{i+}\pi_{+j}} = 1, i, j = 1, 2$, and $\phi''(1) = 0$. For the derivatives (3.5) to (3.8), one can easily verify that under independence

$$\frac{\partial\theta^{\phi}(\boldsymbol{\pi}^I)}{\partial\pi_{ij}} = \phi''(1)\frac{1}{\pi^I_{ij}} \qquad i, j = 1, 2,$$

which leads to

$$\sum_{i,j=1}^{2} \pi^I_{ij} \left(\frac{\partial\theta^{\phi}(\boldsymbol{\pi}^I)}{\partial\pi_{ij}}\right)^2 = \phi''(1)^2 \sum_{i,j=1}^{2} \frac{1}{\pi^I_{ij}}.$$

Finally, as $\frac{\partial\theta^{\phi}(\boldsymbol{\pi}^I)}{\partial\pi_{ij}} \neq 0$, it holds $\nabla\theta^{\phi}(\boldsymbol{\pi}^I) \neq 0$ and by Theorem 3.2.1 the asymptotic variance is

$$\sigma^2_{\phi}(\boldsymbol{\pi}^I) = \sum_{i,j=1}^{2} \pi^I_{ij} \left(\frac{\partial\theta^{\phi}(\boldsymbol{\pi}^I)}{\partial\pi_{ij}}\right)^2 - \underbrace{\left(\sum_{i,j=1}^{2} \pi^I_{ij} \frac{\partial\theta^{\phi}(\boldsymbol{\pi}^I)}{\partial\pi_{ij}}\right)^2}_{=0} = \phi''(1)^2 \sum_{i,j=1}^{2} \frac{1}{\pi^I_{ij}},$$

which is (3.13). □

Example 3.2.4. (Kullback-Leibler Divergence)

The Kullback-Leibler divergence (ϕ_{KL}) is the underlying divergence for the log-odds ratio. It holds $\nabla\theta^{\phi_{KL}}(\boldsymbol{\pi}) = \left(\frac{1}{\pi_{11}}, -\frac{1}{\pi_{12}}, -\frac{1}{\pi_{21}}, \frac{1}{\pi_{22}}\right)$ and $\nabla\theta^{\phi_{KL}}(\pi_1, \pi_2, r) = \left(\frac{1}{\pi_1(1-\pi_1)}, -\frac{1}{\pi_2(1-\pi_2)}\right)$. Therefore

$$\sum_{i,j=1}^{2} \pi_{ij} \frac{\partial\theta^{\phi_{KL}}(\boldsymbol{\pi})}{\partial\pi_{ij}} = 0 = \sum_{i=1}^{2} \pi_i(1-\pi_i)\frac{\partial\theta^{\phi_{KL}}(\pi_1, \pi_2, r)}{\partial\pi_i},$$

and for the log-odds ratio it holds

$$\sigma^2_{\phi_{KL}}(\boldsymbol{\pi}) = \sum_{i,j=1}^{2} \frac{1}{\pi_{ij}} \qquad \text{and} \qquad \sigma^2_{\mathcal{B},\phi_{KL}}(\pi_1, \pi_2, r) = \frac{1}{\pi_1(1-\pi_1)} + \frac{1}{r\pi_2(1-\pi_2)}, \tag{3.15}$$

by (3.4) and (3.9), respectively. Considering a multinomial probability table $\boldsymbol{\pi} = (\pi_{ij}) \in \Delta_{2\times 2}$ and a corresponding table having as rows independent binomial probability vectors $(\pi_i, 1-\pi_i), i = 1, 2$, with $\pi_i = \frac{\pi_{i1}}{\pi_{i+}}, i = 1, 2$, it can easily be verified that the two variances in (3.15) are equal.

Example 3.2.5. (Power Divergence)

For the more general power divergence (1.73) and the associated ϕ_λ-scaled odds ratio (3.2), for $\lambda \neq 0, -1$, it holds

$$\begin{aligned}\sum_{i,j=1}^{2} \pi_{ij} \frac{\partial \theta^{\phi_\lambda}(\boldsymbol{\pi})}{\partial \pi_{ij}} &= -\psi_{11}\phi''_\lambda(\psi_{11}) + \psi_{12}\phi''_\lambda(\psi_{12}) + \psi_{21}\phi''_\lambda(\psi_{21}) - \psi_{22}\phi''_\lambda(\psi_{22}) \\ &= -\psi_{11}^\lambda + \psi_{12}^\lambda + \psi_{21}^\lambda - \psi_{22}^\lambda = -\lambda\theta^{\phi_\lambda}(\boldsymbol{\pi}).\end{aligned}$$

The formula for $\sigma^2_{\phi^\lambda}(\boldsymbol{\pi})$ is not given here due to its complexity.

Remark 3.2.6. The derivatives (3.5) to (3.8) in the multinomial sampling scheme can be calculated using the function `por.nabla` in the digital attachment (see Appendix B.3). The derivatives (3.11) and (3.12) in the product binomial sampling scheme can be calculated using the function `por.nabla.binomial`.

The standard error estimated $SE(\hat{\theta}^\phi) = \sqrt{\sigma^2_\phi(\hat{\boldsymbol{\pi}})/n}$ and $SE(\hat{\theta}^\phi_{\mathcal{B}}) = \sqrt{\sigma^2_{\mathcal{B},\phi}(\hat{\pi}_1, \hat{\pi}_2, r)/n}$ can be calculated using the functions `por.SE` and `por.SE.binomial` for the multinomial and product binomial sampling scheme, respectively.

3.2.1 Difference in binomial and multinomial variances

Based on Lemma 3.1.1 and relations (3.4) and (3.9) for the power divergence ϕ_λ, Table 3.5 shows that, although θ^ϕ and $\hat{\theta}^\phi$ are independent of the sampling scheme, their variances are in general not, i.e. for $\hat{\pi}_i = \frac{\hat{\pi}_{ij}}{\hat{\pi}_{i+}}, i = 1, 2,$

$$\mathbb{V}\text{ar }\hat{\theta}^\phi = \mathbb{V}\text{ar }\theta^\phi(\hat{\boldsymbol{\pi}}) \neq \mathbb{V}\text{ar }\theta^\phi_{\mathcal{B}}(\hat{\pi}_1, \hat{\pi}_2, r).$$

The ordinary log-odds ratio ($\lambda \to 0$) is an exception.

Varying variances for the two sampling schemes, multinomial and product binomial are well-known. For example in the context of the Cochran-Mantel-Haenszel test in $2 \times 2 \times K, K \in \mathbb{N}$, contingency tables cross-classifying the presence of a prognostic factor and the binary response of patients across K independent levels. Let $\mathbf{n}_k = (n_{ijk})$ be the cell counts of the k-th partial 2×2 table. Assume

$$N_{i1k} \sim \mathcal{B}(n_{i+k}, \pi_{i+k}), \qquad i = 1, 2, \qquad k = 1, ..., K$$

as independent random variables for the random cell counts. Cochran [1954] proposed a test for conditional independence between X and Y given the level $k = 1, ..., K$, assuming the binomial sampling scheme in every 2×2 table. Thus, expected value estimate becomes $\hat{\mathbb{E}} N_{11k} = n_{1+k} n_{+1k} / n_{++k}$ and variance estimate becomes

$$\widehat{\mathbb{V}\text{ar}}\ N_{11k} = n_{1+k} n_{2+k} n_{+1k} n_{+2k} / n_{++k}^3.$$

The proposed test statistic is

$$CMH = \frac{\left[\sum_k (n_{11k} - \hat{\mathbb{E}}(n_{11k}))\right]^2}{\sum_k \widehat{\mathbb{V}\text{ar}}\ N_{11k}}.$$

Mantel and Haenszel [1959] proposed the same conditional independence test by fixing $\{n_{1+k}, n_{2+k}\}$ and $\{n_{+1k}, n_{+2k}\}$. Thus, N_{11k} is conditional a hypergeometric distribution with the same expected value as in the binomial case. It holds:

$$\widehat{\mathbb{V}\text{ar}}\ N_{11k} = n_{1+k} n_{2+k} n_{+1k} n_{+2k} / n_{++k}^2 (1 - n_{++k}).$$

Therefore, the variances differ with changing sampling schemes.

3.2.2 Hypotheses Tests and Confidence Intervals

Applying Slutsky's Theorem, the following Lemma is useful in defining tests and CIs:

Lemma 3.2.7. Consider the set-up of Theorem 3.2.1 and let $\phi \in \Phi$ be twice continuous differentiable on $(0, \infty)$. Then it holds

$$\sigma_\phi^2(\hat{\boldsymbol{\pi}}) \xrightarrow{P} \sigma_\phi^2(\boldsymbol{\pi}) \qquad \text{for } n \to \infty \tag{3.16}$$

$$\frac{\sqrt{n}(\theta^\phi(\hat{\boldsymbol{\pi}}) - \theta^\phi(\boldsymbol{\pi})}{\sigma_\phi(\hat{\boldsymbol{\pi}})} \xrightarrow{d} \mathcal{N}(0,1) \qquad \text{for } n \to \infty. \tag{3.17}$$

The analogous statements hold for $\theta_{\mathcal{B}}^\phi(\pi_1, \pi_2, r)$ and the associated $\sigma_{\mathcal{B},\phi}^2(\pi_1, \pi_2, r)$ in the set-up of Theorem 3.2.2.

Proof. Since ϕ'' is continuous, the function

$$\boldsymbol{\pi} \mapsto \sigma_\phi^2(\boldsymbol{\pi}) = \sum_{i,j=1}^{2} \pi_{ij} \left(\frac{\partial \theta^\phi(\boldsymbol{\pi})}{\partial \pi_{ij}}\right)^2 - \left(\sum_{i,j=1}^{2} \pi_{ij} \frac{\partial \theta^\phi(\boldsymbol{\pi})}{\partial \pi_{ij}}\right)^2$$

is continuous. The consistency of the MLEs under continuous functions (cf. Casella and Berger [2002, Theorem 10.1.6, p. 470]) implies $\sigma_\phi^2(\hat{\boldsymbol{\pi}}) \xrightarrow{P} \sigma_\phi^2(\boldsymbol{\pi})$ for $n \to \infty$. Thus, (3.16) holds. Using Theorem 3.2.1 it holds

$$Z_n := \frac{\sqrt{n}(\theta^\phi(\hat{\boldsymbol{\pi}}) - \theta^\phi(\boldsymbol{\pi}))}{\sigma_\phi(\boldsymbol{\pi})} \xrightarrow{d} \mathcal{N}(0,1) \qquad \text{for } n \to \infty.$$

Define $Y_n := \frac{\sigma_\phi(\boldsymbol{\pi})}{\sigma_\phi(\hat{\boldsymbol{\pi}})}$. Since $\sigma_\phi(\hat{\boldsymbol{\pi}}) \xrightarrow{P} \sigma_\phi(\boldsymbol{\pi})$ for $n \to \infty$, it holds $Y_n \xrightarrow{P} 1$ for $n \to \infty$. Applying Slutsky's Theorem (cf. Casella and Berger [2002, Theorem 5.5.17, pp. 239]) on Z_n and Y_n gives formula (3.17):

$$\frac{\sqrt{n}(\theta^\phi(\hat{\boldsymbol{\pi}}) - \theta^\phi(\boldsymbol{\pi}))}{\sigma_\phi(\hat{\boldsymbol{\pi}})} = Z_n Y_n \xrightarrow{d} \mathcal{N}(0,1) \qquad \text{for } n \to \infty.$$

The results for $\theta^\phi_{\mathcal{B}}$ and $\sigma^2_{\mathcal{B},\phi}$ are proved analogously. □

Null hypotheses of the form $H_0 : \theta^\phi(\boldsymbol{\pi}) = D_0$ for a fixed $D_0 \in \mathbb{R}$ can be tested asymptotically against two- or one-sided alternatives by the classical Z-test, using the test statistic

$$Z^\phi = \frac{\sqrt{n}(\theta^\phi(\hat{\boldsymbol{\pi}}) - \theta^\phi(\boldsymbol{\pi}))}{\sigma_\phi(\hat{\boldsymbol{\pi}})} \quad \text{resp.} \quad Z^\phi_{\mathcal{B}} = \frac{\sqrt{n}(\theta^\phi_{\mathcal{B}}(\hat{\pi}_1, \hat{\pi}_2, r) - \theta^\phi_{\mathcal{B}}(\pi_1, \pi_2, r))}{\sigma_{\phi,\mathcal{B}}(\hat{\pi}_1, \hat{\pi}_2, r)} \tag{3.18}$$

due to Lemma 3.2.7. The construction of Wald CIs is then straightforward. For example, the asymptotic $(1-\alpha)100\%$ Wald CI for θ^ϕ is

$$CI^\phi_\alpha(\mathbf{n}) = \left[\theta^\phi(\hat{\boldsymbol{\pi}}) - \frac{z_{\alpha/2}}{\sqrt{n}}\sigma_\phi(\hat{\boldsymbol{\pi}}),\ \ \theta^\phi(\hat{\boldsymbol{\pi}}) + \frac{z_{\alpha/2}}{\sqrt{n}}\sigma_\phi(\hat{\boldsymbol{\pi}})\right] , \tag{3.19}$$

where $z_{\alpha/2}$ denotes the $(1-\alpha/2)$-quantile of a standard normal distribution. The values of the test statistic (3.18) can be calculated with the attached R functions `Z.phi.multinomial` and `Z.phi` for the multinomial and product binomial sampling scheme, respectively (see Appendix B.3). The two-sided $(1-\alpha)$ Wald CI (3.19) can be obtained using `calc.CI.phi.multi` and `calc.CI.phi` for the multinomial and product binomial sampling scheme.

3.3 Boundedness

One of the drawbacks of the classical $\log\theta$ is that it is unbounded and, in case of a sampling zero, it is estimated as $\pm\infty$. The newly introduced ϕ-scaled odds ratio can be bounded and estimated finitely in the presence of sampling zeros. The conditions that the ϕ-function has to fulfil in order to achieve boundedness and finite estimation of θ^ϕ are provided in the next Lemma.

Lemma 3.3.1. Let $\phi \in \Phi$ be twice differentiable and strictly convex. Then θ^ϕ is bounded in the set $\mathcal{S}$ of all 2×2 probabilities $\boldsymbol{\pi}$ that has fixed marginals $\pi_{1+} \in (0,1)$ and $\pi_{+1} \in (0,1)$ if

$$\phi'(0) := \lim_{t \searrow 0} \phi'(t) > -\infty \ . \tag{3.20}$$

In this case, in the presence of a sampling zero, the MLE $\hat{\theta}^\phi$ remains finite while its estimated standard error exists finitely if $\phi''(0) := \lim_{t\searrow 0} \phi''(t) < \infty$.

Proof. Recall (Proposition 3.1.3 (ii)) that for $\boldsymbol{\pi} \in \Delta_{2\times2}(\pi_{1+}, \pi_{+1})$, $\theta^\phi(\pi_{11}|\pi_{1+}, \pi_{+1})$ is increasing in π_{11}, for $\pi_{11} \in (\underline{\pi}_{11}, \overline{\pi}_{11})$. Thus, the lower (upper) bound of θ^ϕ is reached for $\pi_{11} \to \underline{\pi}_{11}$ $(\overline{\pi}_{11})$.

If $\underline{\pi}_{11} = 0$, then the lower bound of $\theta^\phi(\pi_{11}|\pi_{1+}, \pi_{+1})$ in $\Delta_{2\times 2}(\pi_{1+}, \pi_{+1})$ becomes

$$\lim_{\pi_{11}\searrow 0}\left(\phi'\left(\frac{\pi_{11}}{\pi_{1+}\pi_{+1}}\right) - \phi'\left(\frac{\pi_{1+}-\pi_{11}}{\pi_{1+}\pi_{+2}}\right) - \phi'\left(\frac{\pi_{+1}-\pi_{11}}{\pi_{2+}\pi_{+1}}\right) + \phi'\left(\frac{1-(\pi_{1+}+\pi_{+1})+\pi_{11}}{\pi_{2+}\pi_{+2}}\right)\right)$$
$$= \lim_{t\searrow 0}\phi'(t) - \underbrace{\left[\phi'\left(\frac{\pi_{1+}}{\pi_{1+}\pi_{+2}}\right) + \phi'\left(\frac{\pi_{+1}}{\pi_{2+}\pi_{+1}}\right) - \phi'\left(\frac{1-(\pi_{1+}+\pi_{+1})}{\pi_{2+}\pi_{+2}}\right)\right]}_{=:A\in\mathbb{R}},$$

which is finite if and only if $\lim_{t\searrow 0}\phi'(t) > -\infty$. Using the same arguments as above, one can verify the same condition for the finiteness of the lower (upper) bound of π_{11} for $\underline{\pi}_{11} = \pi_{1+} + \pi_{+1} - 1$ ($\overline{\pi}_{11} \in \{\pi_{1+}, \pi_{+1}\}$). Overall,

$$\lim_{\pi_{11}\searrow\underline{\pi}_{11}} \theta^\phi \quad \begin{cases} > -\infty, & \phi'(0) > -\infty \\ = -\infty, & \phi'(0) = -\infty \end{cases} \quad \text{and} \quad \lim_{\pi_{11}\nearrow\overline{\pi}_{11}} \theta^\phi \quad \begin{cases} < \infty, & \phi'(0) > -\infty \\ = \infty, & \phi'(0) = -\infty. \end{cases}$$

For the second part of Lemma 3.3.1, since $\operatorname{Var}\hat{\theta}^\phi$ is a weighted sum of $\phi''\left(\frac{\pi_{ij}}{\pi_{i+}\pi_{+j}}\right)$, for $i, j = 1, 2$, it is well-defined on the closure $\overline{\Delta(\pi_{1+}, \pi_{+1})}$ if $\phi''(0)$ exists. □

The derivative criterion of Lemma 3.3.1 is easy to apply. One can verify that for the power divergence θ^{ϕ_λ} is always bounded for $\lambda > 0$. In case of a sampling zero, $\hat{\theta}^{\phi_\lambda}$ is finite for $\lambda > 0$ but its estimated standard error exists finitely for $\lambda \geq 1$. An overview of the boundedness and variance existence for selected ϕ-divergences is given in Table 3.2. The continuity correction will be applied by adding a small positive constant c to all cells of the table since the case $\lambda < 1$ is considered as well. The continuity correction has been intensively studied in Chapter 2.

Table 3.2: Some examples for the boundedness and variance existence of the ϕ-scaled odds ratio.

Divergence			Boundedness	Variance
Power-Divergence ($\lambda \leq 0$)	$\phi'(0) = -\infty$	$\phi''(0) = \infty$	unbounded	undefined
Power-Divergence ($0 < \lambda < 1$)	$\phi'(0) = -\frac{1}{\lambda}$	$\phi'(0) = \infty$	bounded	undefined
Pearson Divergence ($\lambda = 1$)	$\phi'(0) = -1$	$\phi'(0) = 0^0 = 1$	bounded	finite
Power-Divergence ($\lambda > 1$)	$\phi'(0) = -\frac{1}{\lambda}$	$\phi'(0) = 0$	bounded	finite
J-Divergence	$\phi'(0) = -\infty$	$\phi'(0) = \infty$	unbounded	undefined
Rukin ($a > 0$)	$\phi'(0) = -\frac{a+1}{2a^2}$	$\phi'(0) = \frac{1}{a^3}$	bounded	finite
Balakrishnan and Sanghri	$\phi'(0) = -4$	$\phi''(0) = 16$	bounded	finite

In Figure 3.1 and Table 3.3 numerical examples are given based on the set of tables $\mathcal{S}$ with marginal counts $n_{1+} = 10, n_{+1} = 8$ and $n = 20$. The ϕ-scaled odds ratio behave linear near independence, while difference occurs at the edges. The Pearson divergence increases linear on the whole space, which is natural as the formula is just a weighted sum of n_{11}. In general, linearity does not hold for positive or negative association, where a change of n_{11} effects the ϕ-scaled odds ratio in a nonlinear manner.

However, depending on $\phi \in \Phi$, the ϕ-scaled odds ratio can loose its unboundedness in comparison to the log-odds ratio. This is not a disadvantage, since boundedness is a desired property for measures of association (cf. Section 1.6). Before constructing such measures based on the ϕ-scaled odds ratios, the bound is analysed in more detail.

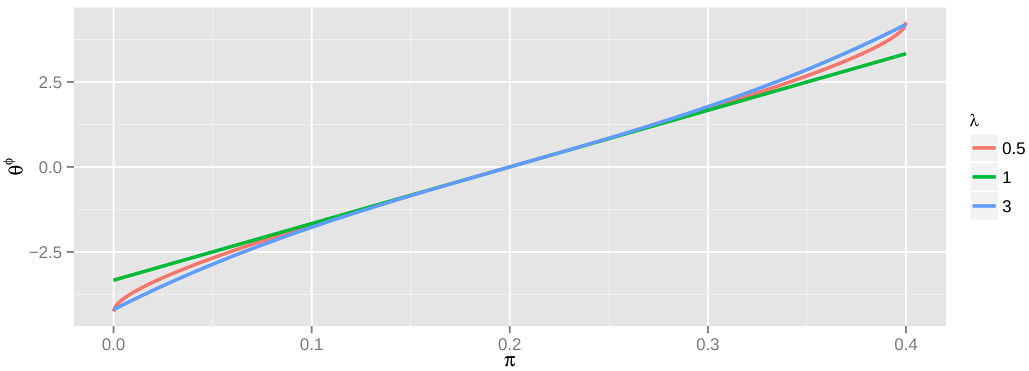


Figure 3.1: Behavior of θ^ϕ for $\pi_{1+} = \frac{1}{2}$ and $\pi_{+1} = \frac{8}{20}$, as a function of $\pi = \pi_{11}$ for $n = 20$.

Table 3.3: Values of $\hat{\theta}^\phi(\mathbf{n})$ for all possible tables $\mathbf{n}$ with marginal counts $n_{1+} = 10$, $n_{+1} = 8$ and sample size $n = 20$.

n_{11}	$\lambda = 3$	$\lambda = -1$	$\lambda = 0.5$	$\lambda = 0$	$\lambda = 0.5$	$\lambda = 1$	$\lambda = 3$
0	$-\infty$	$-\infty$	$-\infty$	$-\infty$	-4.26	-3.33	-4.20
1	-23.84	-4.76	-3.68	-3.04	-2.68	-2.50	-2.86
2	-3.55	-2.08	-1.91	-1.79	-1.71	-1.67	-1.77
3	-0.99	-0.88	-0.86	-0.85	-0.84	-0.83	-0.85
4	0.00	0.00	0.00	0.00	0.00	0.00	0.00
5	0.99	0.88	0.86	0.85	0.84	0.83	0.85
6	3.55	2.08	1.91	1.79	1.71	1.67	1.77
7	23.84	4.76	3.68	3.04	2.68	2.50	2.86
8	∞	∞	∞	∞	4.26	3.33	4.20

3.3.1 Calculation of θ^ϕ's Value Range

Since θ^ϕ can be bounded, its maximum and value range are further studied.

Definition 3.3.2. Let $\phi \in \Phi$ be differentiable. Let $n \in \mathbb{N}$ be the sample size in a multinomial and $(n_1, n_2) \in \mathbb{N}^2$ be the sample sizes in the binomial sampling scheme and define

$$\theta^\phi_{\max,c} = \sup_{\substack{\mathbf{n} \in \mathbb{N}_0^4 \\ |\mathbf{n}| = n}} \hat{\theta}^\phi(\mathbf{n} + c\mathbf{1}_{2\times 2}) \qquad \text{(multinomial)}$$

$$\theta^\phi_{\max,c,\mathcal{B}} = \sup_{\substack{\mathbf{n} \in \mathbb{N}_0^4 \\ n_{1+} = n_1 \\ n_{2+} = n_2}} \hat{\theta}^\phi(\mathbf{n} + c\mathbf{1}_{2\times 2}) \qquad \text{(product binomial)}$$

for the c-corrected maximum value range, $c \geq 0$. Set $\theta^\phi_{\max} := \theta^\phi_{\max,c=0}$ and $\theta^\phi_{\max,\mathcal{B}} := \theta^\phi_{\max,c=0,\mathcal{B}}$ for the maximum value without correction in the multinomial and binomial sampling scheme.

By Lemma 3.3.1, both maxima are ∞ if $\phi'(0) = -\infty$ and $c = 0$. For $c > 0$, the corrected maximum range lies in $[0, \infty)$. Thus, the value space $[-\theta^\phi_{\max,c}, \theta^\phi_{\max,c}]$ is always bounded for c-corrected ϕ-scaled odds ratio.

Lemma 3.3.3. Let $\phi \in \Phi$ be strictly convex and let $n_1, n_2 \in \mathbb{N}$ be the row counts in a binomial sampling scheme. Set $n = n_1 + n_2$ and let $c \geq 0$ then it holds

$$\max_{\substack{\mathbf{n} \\ n_{1+}=n_1 \\ n_{2+}=n_2}} \hat{\theta}^{\phi}_{\mathcal{B}}(\mathbf{n} + c\mathbf{1}_{2\times 2}) = \theta^{\phi}_{\mathcal{B}}(\mathbf{n}_{\max} + c\mathbf{1}_{2\times 2}) \tag{3.21}$$

$$= \phi'\left(\frac{(n+4c)(n_1+c)}{(n_1+2c)^2}\right) + \phi'\left(\frac{(n+4c)(n_2+c)}{(n_2+2c)^2}\right) - 2\phi'\left(\frac{c(n+4c)}{(n_1+2c)(n_2+2c)}\right). \tag{3.22}$$

The maximum is generated by the table $\mathbf{n}_{\max} = (n_1, 0, 0, n_2)$. The maximum (3.22) must not be finite, when $\phi'(0) = -\infty$ and $c = 0$.

Proof. Let $\mathbf{n}(n_{11}, n_{21})$ be the table with marginals n_1 and n_2, which is generated by n_{11} and n_{21}. The partial derivatives with respect to $n_{i1}, i = 1, 2$ are:

$$\begin{aligned}
&\frac{\partial \hat{\theta}^{\phi}(\mathbf{n}(n_{11}, n_{21}) + c\mathbf{1}_{2\times 2})}{\partial n_{i1}} \\
&\quad = (-1)^{i+1} \frac{(n+4c)(n_{(3-i)1}+c)}{(n_1+2c)(n_{12}+n_{21}+2c)^2} \phi''\left(\frac{(n+4c)(n_{11}+c)}{(n_1+2c)(n_{11}+n_{21}+2c)}\right) \\
&\quad + (-1)^{i+1} \frac{(n+4c)(n_{3-i}-n_{(3-i)1}+c)}{(n_1+2c)(n-n_{11}-n_{21}+2c)^2} \phi''\left(\frac{(n+4c)(n_1-n_{11}+c)}{(n_1+2c)(n-n_{11}-n_{21}+2c)}\right) \\
&\quad + (-1)^{i+1} \frac{(n+4c)(n_{(3-i)1}+c)}{(n_2+2c)(n_{11}+n_{21}+2c)^2} \phi''\left(\frac{(n+4c)(n_{21}+c)}{(n_2+2c)(n_{11}+n_{21}+2c)}\right) \\
&\quad + (-1)^{i+1} \frac{(n+4c)(n_{3-i}-n_{(3-i)1}+c)}{(n_2+2c)(n-n_{11}-n_{21}+2c)^2} \phi''\left(\frac{(n+4c)(n_2-n_{21}+c)}{(n_2+2c)(n-n_{11}-n_{21}+2c)}\right)
\end{aligned}$$

Due to the strict convexity, it holds $\phi'' > 0$ such that $\frac{\partial \hat{\theta}^{\phi}(\mathbf{n}(n_{11}, n_{21}) + c\mathbf{1}_{2\times 2})}{\partial n_{11}} > 0$ and $\frac{\partial \hat{\theta}^{\phi}(\mathbf{n}(n_{11}, n_{21}) + c\mathbf{1}_{2\times 2})}{\partial n_{21}} < 0$. Thus, the maximum is obtained, if n_{11} is large and if n_{21} is small. Since $n_{11} \in [0, n_1] \cap \mathbb{N}$ and $n_{21} \in [0, n_2] \cap \mathbb{N}$, the maximum is obtained for $n_{11} = n_1$ and $n_{21} = 0$, which is $\mathbf{n}_{\max} = (n_1, 0, 0, n_2)$ and formula (3.21) follows directly by calculating $\hat{\theta}^{\phi}(\mathbf{n}_{\max} + c\mathbf{1}_{2\times 2})$. □

3.3.2 Relative Length

Under the conditions of Lemma 3.3.1, θ^{ϕ} ($\hat{\theta}^{\phi}$) ranges from $-\theta^{\phi}_{\max}$ to $\theta^{\phi}_{\max}$ with $\theta^{\phi}_{\max} = \max_{\mathcal{S}}\{\theta^{\phi}(\boldsymbol{\pi}^*)\}$, where $\boldsymbol{\pi}^*$ denotes a 2×2 table with $\pi_{11} = \min\{\pi_{1+}, \pi_{+1}\}$. In case of independent binomial sampling scheme, Lemma 3.3.1 can be adjusted by Lemma 3.1.1. Meanwhile, the corresponding $\theta^{\phi}_{\max}$ for $\theta^{\phi}_{\mathcal{B}}$ is obtained by replacing $\boldsymbol{\pi}^*$ in $\theta^{\phi}_{\max}$ by $\boldsymbol{\pi}_d = \text{diag}(\frac{n_1}{n_1+n_2}, \frac{n_2}{n_1+n_2})$, where diag(·) stands for a diagonal matrix (Lemma 3.3.3).

The c-corrected estimator of θ^{ϕ} is $\hat{\theta}^{\phi}_c = \hat{\theta}^{\phi}(\mathbf{n} + c\mathbf{1}_{2\times 2})$, where $\mathbf{1}_{2\times 2}$ denotes a 2×2 table of ones, and is always finite for $c > 0$. It ranges from $-\theta^{\phi}_{\max,c}$ to $\theta^{\phi}_{\max,c}$, with $\theta^{\phi}_{\max,c} = \hat{\theta}^{\phi}(n\boldsymbol{\pi}^* + c\mathbf{1}_{2\times 2})$ (multinomial) or $\theta^{\phi}_{\max,c} = \hat{\theta}^{\phi}(\boldsymbol{n}_d + c\mathbf{1}_{2\times 2})$ (binomial), where $\boldsymbol{n}_d = \text{diag}(n_1, n_2)$. Let $[CI^c_L, CI^c_U]$ be the c-corrected asymptotic $100(1-\alpha)\%$ Wald CI of the ϕ-scaled odds ratio. Since CI^c_L and CI^c_U have to lie in the interval $[-\theta^{\phi}_{\max,c},\ \theta^{\phi}_{\max,c}]$, they are defined as $\max(\widetilde{CI}_L^c, -\theta^{\phi}_{max,c})$ and $\min(\widetilde{CI}_U^c, \theta^{\phi}_{max,c})$. $\widetilde{CI}_L^c$ and $\widetilde{CI}_U^c$ denote the lower and upper limits of the confidence interval $CI^{\phi}_{\alpha}(\mathbf{n} + c\mathbf{1}_{2\times 2})$, derived when substituting $\mathbf{n}$ in (3.19) by $\mathbf{n} + c\mathbf{1}_{2\times 2}$.

A standard criterion for comparing CIs is their length. In this set-up, due to the scale differences among the members of the θ^{ϕ} family, this comparison is useless. For this, CIs are compared in terms of their relative lengths (RLs) instead, which can be calculated as long as the corresponding θ^{ϕ} is bounded (see Lemma 3.3.1). For example, the RL of the c-corrected asymptotic Wald CI is

$$RL_c^{\phi}(\mathbf{n}) = \frac{CI_U^c - CI_L^c}{2\theta^{\phi}_{\max,c}}. \tag{3.23}$$

It is denoted as $RL^{\phi}_{c,\mathcal{B}}$ when it is defined in the product binomial sampling scheme. The relative length (3.23) can be calculated using the functions `CI.cov.percent.corrected` for the product binomial sampling scheme and `CI.cov.percent.corrected.multi` for the multinomial sampling scheme (see Appendix B.3).

Table 3.4: Categorization of the relative length $RL_{c,\mathcal{B}}$ of the $(n_1+1)(n_2+1) = 256$ tables in the binomial sampling scheme with $n_1 = 15 = n_2$ sample size in the power divergence family (Example 1.10.4) with parameter $\lambda \in \{-1/3, 0, 1/3, 2/3\}$. The continuity correction is small ($c = 0.001$) and standard ($c = 0.5$). The row *censored* gives the number of CIs that had to be truncated to fit into the value range. The percentage intervals $[10i\%, 10(i+1)\%), i = 0, ..., 9$ are closed on the left and open on the right side. Relative length has been calculated using the R-functions `get.LR.dist`, `is.RL.censored` and `count.RL.censoring` (see Appendix B.3).

	$c = 0.001$				$c = 0.5$			
λ	-1/3	0	1/3	2/3	-1/3	0	1/3	2/3
0-10%	196	132	0	2	0	0	0	2
10-20%	0	64	28	52	130	0	2	8
20-30%	0	0	162	96	60	186	148	62
30-40%	0	0	20	88	12	24	86	152
40-50%	0	0	20	14	26	44	14	24
50-60%	0	0	12	0	28	2	4	2
60-70%	0	0	8	2	0	0	0	4
70-80%	2	0	0	0	0	0	2	0
80-90%	0	0	4	0	0	0	0	0
90-100%	0	0	0	0	0	0	0	2
100%	58	60	2	2	0	0	0	0
censored	58	60	60	8	30	24	20	20

Table 3.4 shows the distribution of the relative length over all 256 CIs for the sampling scheme of the prednisolone data (binomial with $n_1 = 15 = n_2$), categorised in the intervals $[0.1i\%, 0.1(i+1)\%), i = 0, ..., 9$ and the value 100%. For $\lambda \in \{-1/3, 0\}$ and $c = 0.001$ the tables leading to CIs with a RL of 100% all included sampling zeros. In these cases, the tables can be perfectly separated in small RL tables (< 20%) and degenerate RL tables (100%). With the same correction and a positive λ, the tables give greater RL but are more spread around the range. For greater correction ($c = 0.5$), there are no RL values

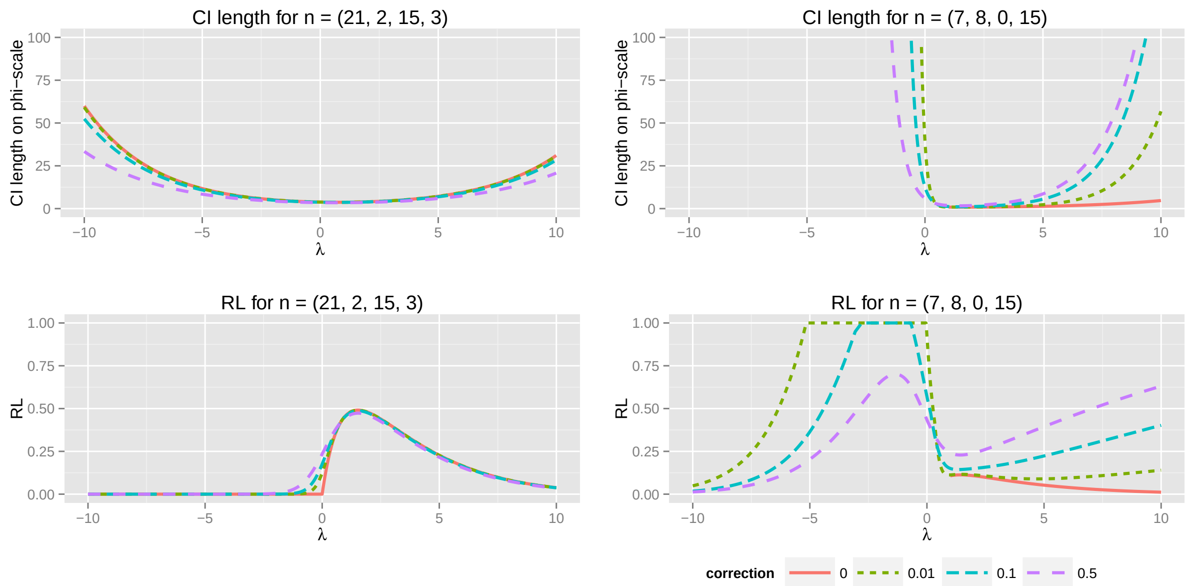


Figure 3.2: Plots for the absolute length $CI_U - CI_L$ and relative length (RL) of the ϕ-scaled odds ratio CIs for Example 1.2.1 (left) and Example 1.2.2 (right) for the power divergence $\lambda \mapsto \phi_\lambda$ and different corrections ($c \in \{0, 0.01, 0.1, 0.5\}$).

of 100% and the RLs are more spread around the medium values $(20-50\%)$. The rule of thumb is, that the RL increases when at least one cell count is small. This is explained by the term $(\frac{\pi_{ij}}{\pi_{i+}\pi_{+j}})^\lambda$ resp. $(\frac{\pi_{ij}}{\pi_{i+}\pi_{+j}})^{\lambda-1}$ (or their binomial equivalents) within the asymptotic variance formula, where small probabilities lead to great asymptotic variances if $\lambda < 1$ and therefore to long CIs. The separation and spreading effect takes place for other pairs of (n_1, n_2) (balanced and unbalanced). Thus, Table 3.4 represents the general behaviour of the RL under different λ and different corrections c.

Figure 3.2 shows the absolute and relative length of the CIs dependent on $\lambda \in [-10, 10]$ for different corrections $c \in \{0, 0.01, 0.1, 0.5\}$ for the prednisolone and larynx data (Table 1.1). When no sampling zero is in the table, the absolute and relative length of the CIs are hardly effected by the correction term c (Figure 3.2, left). This is in agreement with the analogue observation in Chapter 2 for the log-odds ratio. Therefore, the absolute length converges to infinity if $|\lambda| \to \infty$, where the minimum is reached within $\lambda \in [0, 2]$. The RL becomes large for small positive values of λ. This is explained by the boundedness of the value range which is smallest for λ in these regions. The correction effects the absolute CI length and the relative length heavily when a sampling zero is in the table. The RL reaches the maximum value of 1 for negative $\lambda \in [-5, 0]$ and small corrections $c < 0.1$ when the absolute length grows for $|\lambda| \to \infty$.

3.4 Invertibility

As already explained, the ϕ-scale odds ratio measures the divergence from independence on ϕ-scale. For interpretation purpose the log-odds ratios can be rescaled to the ordinary odds ratio by using the exponential function. Since the ϕ-scaled odds ratios are measured on different scales, such an interpretation is not straightforward and requires additional attention. A helpful approach is to invert the ϕ-scaled odds ratio to the log scale.

Lemma 3.4.1. Let $\phi \in \Phi$ be strictly convex and differentiable. Fix π_{1+} and $\pi_{+1} \in (0,1)$ and let $\Delta_{2\times 2}(\pi_{1+}, \pi_{+1})$ be the set of 2×2 probabilities with marginals π_{1+} and π_{+1}. Let

$$\begin{aligned} \theta^{\phi}_{\max} &= \sup\{\theta^{\phi}(\boldsymbol{\pi}) \mid \boldsymbol{\pi} \in \Delta_{2\times 2}(\pi_{1+}, \pi_{+1})\} \in (-\infty, \infty] \qquad \text{and} \\ \theta^{\phi}_{\min} &= \inf\{\theta^{\phi}(\boldsymbol{\pi}) \mid \boldsymbol{\pi} \in \Delta_{2\times 2}(\pi_{1+}, \pi_{+1})\} \in [-\infty, \infty). \end{aligned}$$

The function

$$\theta^{\phi} : \Delta_{2\times 2}(\pi_{1+}, \pi_{+1}) \to (\theta^{\phi}_{\min}, \theta^{\phi}_{\max}), \qquad \boldsymbol{\pi} \mapsto \theta^{\phi}(\boldsymbol{\pi})$$

is invertible.

Proof. Each element $\boldsymbol{\pi} \in \Delta_{2\times 2}(\pi_{1+}, \pi_{+1})$ has a representation as $\pi_{11} \in (\underline{\pi}_{11}, \overline{\pi}_{11})$ (recall the reasoning of Proposition 3.1.3 (ii)). The function $\pi_{11} \mapsto \theta^{\phi}(\pi_{11}|\pi_{1+}, \pi_{+1})$ is strictly increasing with maximum value $\theta^{\phi}_{\max}$ and minimum value $\theta^{\phi}_{\min}$ and thus invertible. $\square$

Lemma 3.4.1 shows the existence of an inversion function. Due to the complexity of the equation

$$\theta^{\phi}(\pi_{11}|\pi_{1+}, \pi_{+1}) = \theta_0 \in (\theta^{\phi}_{\min}, \theta^{\phi}_{\max}), \qquad \phi \in \Phi, \tag{3.24}$$

an analytical solution is not always possible. Positive counterexamples are the Pearson and Kullback-Leibler divergence, where inversion can be achieved by solving equation (3.24) by hand.

For the Pearson divergence (Example 1.10.3, $\phi_P(x) = 1/2(x-1)^2$) it holds

$$\begin{aligned} \theta^{\phi_P}_0 &= \theta^{\phi_P}(\pi_{11}|\pi_{1+}, \pi_{+1}) \\ &= \frac{\pi_{11}}{\pi_{1+}\pi_{+1}} - \frac{\pi_{1+} - \pi_{11}}{\pi_{1+}\pi_{+2}} - \frac{\pi_{+1} - \pi_{11}}{\pi_{2+}\pi_{+1}} + \frac{1 - (\pi_{1+} + \pi_{+1}) + \pi_{11}}{\pi_{2+}\pi_{+2}} \\ &= \pi_{11} \frac{\pi_{2+}\pi_{+2} + \pi_{2+}\pi_{+1} + \pi_{1+}\pi_{+2} + \pi_{1+}\pi_{+1}}{\pi_{1+}\pi_{2+}\pi_{+2}\pi_{+1}} \\ &\quad + \frac{\pi_{1+}\pi_{2+}\pi_{+1} + \pi_{+1}\pi_{1+}\pi_{+2} - \pi_{1+}\pi_{+1} + (\pi_{1+} + \pi_{+1})\pi_{1+}\pi_{+1}}{\pi_{1+}\pi_{2+}\pi_{+2}\pi_{+1}} \\ &= \pi_{11} \frac{1}{\pi_{1+}\pi_{2+}\pi_{+2}\pi_{+1}} + \frac{\pi_{1+}\pi_{2+}(\pi_{2+} + \pi_{+2} + \pi_{1+} + \pi_{+1}) - \pi_{1+}\pi_{+1}}{\pi_{1+}\pi_{2+}\pi_{+2}\pi_{+1}} \\ &= \pi_{11} \frac{1}{\pi_{1+}\pi_{2+}\pi_{+2}\pi_{+1}} + \frac{1}{\pi_{2+}\pi_{+2}}, \end{aligned}$$

such that the solution of (3.24) becomes

$$\pi_{11} = \theta^{\phi_P}_0 \pi_{1+}\pi_{2+}\pi_{+2}\pi_{+1} - \pi_{1+}\pi_{+1} = \pi_{1+}\pi_{+1}(\theta^{\phi_P}_0 \pi_{2+}\pi_{+2} - 1).$$

Using the same technique, the inverse for the Kullback-Leibler ($\phi_{KL}(x) = x \log x - x + 1$) can also be achieved. The solution to equation (3.24) is given by

$$\pi_{11} = \frac{(\pi_{1+} + \pi_{+1})(1 - \theta^{\phi_{KL}}_0) - 1}{2(1 - \theta^{\phi_{KL}}_0)} - \sqrt{\frac{((\pi_{1+} + \pi_{+1})(1 - \theta^{\phi_{KL}}_0) - 1)^2}{4(1 - \theta^{\phi_{KL}}_0)^2} - \frac{\pi_{1+}\pi_{+1}}{\theta^{\phi_{KL}}_0 - 1}}.$$

The inversions for other divergences lead to more complex equations and have to be solved through numerical techniques. This can be done by using the R function `rescale4` (Appendix B.3).

Inversion can also be done for the binomial sampling scheme:

Lemma 3.4.2. Let $\phi \in \Phi$ be strictly convex and twice differentiable, and let $n_1, n_2 \in \mathbb{N}$ be row marginal counts in a 2×2 table with $r = \frac{n_2}{n_1}$. Fix $\pi_{1+} = \frac{n_1}{n_1+n_2} = 1 - \pi_{2+} = \frac{1}{1+r^{-1}}$. For any $\pi \in (0,1)$ set

$$\begin{aligned}(\Delta_2 \times \Delta_2)(\pi) &= \{(\pi_1, \pi_2) \mid \pi_1\pi_{1+} + \pi_2\pi_{2+} = \pi\},\\ \theta^\phi_{\mathcal{B},\min} &= \inf\{\theta^\phi_{\mathcal{B}}(\pi_1, \pi_2, r) \in (\Delta_2 \times \Delta_2)(\pi)\}, \qquad \text{and}\\ \theta^\phi_{\mathcal{B},\max} &= \sup\{\theta^\phi_{\mathcal{B}}(\pi_1, \pi_2, r) \in (\Delta_2 \times \Delta_2)(\pi)\}.\end{aligned}$$

The function

$$\theta^\phi_{\mathcal{B}} : (\Delta_2 \times \Delta_2)(\pi_{+1}) \to (\theta^\phi_{\mathcal{B},\min}, \theta^\phi_{\mathcal{B},\max}), \qquad (\pi_1, \pi_2) \mapsto \theta^\phi_{\mathcal{B}}(\pi_1, \pi_2, r).$$

is invertable.

Proof. The identity $\pi_1\pi_{1+} + \pi_2\pi_{2+} = \pi$ shows that $\pi_2 = \pi_2(\pi_1|\pi_{+1}, r)$ is a function of π_1. For strictly convex ϕ, the function $\pi_1 \mapsto \theta^\phi_{\mathcal{B}}(\pi_1, \pi_2(\pi_1|\pi_{+1}, r))$ is strictly increasing with maximum value $\theta_{\mathcal{B},\max}$ and minimum value $\theta_{\mathcal{B},\min}$, such that it is invertible. □

Inversion in the binomial sampling scheme is done by solving the equation

$$\theta^\phi(\pi_1|\pi_{+1}, r) = \theta_0 \in (\theta^\phi_{\mathcal{B},\min}, \theta^\phi_{\mathcal{B},\max}), \qquad \phi \in \Phi. \tag{3.25}$$

3.5 Examples and Interpretation

The proposed θ^{ϕ_λ} will be illustrated on the two representative examples (Examples 1.2.1 and 1.2.2) and compared to the classical $\log\theta$. Note that both data sets have fixed row marginals (independent binomial sampling) whereas the prednisolone data additionally possesses a sampling zero. The $\hat{\theta}^{\phi_\lambda}$ for these two data sets, for the $\lambda = 0$ (log-odds ratio), 1/3 and 1, along with their standard errors, bounds, asymptotic 95% Wald CIs and their RLs are provided in Table 3.5 for the c-corrected and the standard ($c = 0$) MLEs. Although the c-corrected estimation is not required for the larynx data, it is applied for obtaining a finite bound for $\log\hat{\theta}$. Observe that for the larynx data, the $\hat{\theta}^{\phi_\lambda}$ values are almost not affected by the choice of λ, while this is the case for the prednisolone data.

The measures θ^ϕ and $\theta^\phi_{\mathcal{B}}$ cannot be interpreted as ratios of odds. Their interpretation, as a ϕ-scaled association, is linked to the intrinsic association parameter of the ϕ-association and the slope of the ϕ-divergence based regression model, as described in Section 3.1. A simple interpretation is possible through the relative association

$$\mathcal{A}^\phi_c = \frac{\theta^\phi(\mathbf{n} + c\mathbf{1}_{2\times2})}{\theta^\phi_{\max,c}}, \qquad c \geq 0,$$

Table 3.5: Estimates, bounds, standard errors, asymptotic 95% Wald CIs and relative length for θ^{ϕ_λ} ($\lambda = 0, 1/3, 1$) for the larynx and the prednisolone data, with and without the continuity correction c, for the binomial and multinomial distribution.

				Binomial				Multinomial			
Data Set	c	λ	$\hat{\theta}^{\phi_\lambda}$	SE	$\theta^{\phi_\lambda}_{\max,c,\mathcal{B}}$	95% CI	RL	SE	$\theta^{\phi_\lambda}_{\max,c}$	95% CI	RL
Larynx	0	0	0.742	0.973	∞	[-1.166, 2.650]	0.000	0.973	∞	[-1.166, 2.650]	0.000
Example 1.2.1		1/3	0.740	0.956	7.585	[-1.133, 2.613]	0.247	0.956	13.369	[-1.133, 2.614]	0.140
		1	0.744	0.950	4.060	[-1.118, 2.607]	0.459	0.951	42.025	[-1.119, 2.608]	0.044
	0.5	0	0.664	0.893	7.461	[-1.086, 2.413]	0.234	0.893	7.475	[-1.086, 2.413]	0.234
		1/3	0.663	0.881	5.355	[-1.065, 2.391]	0.323	0.882	6.774	[-1.065, 2.391]	0.255
		1	0.667	0.880	3.864	[-1.058, 2.391]	0.446	0.880	16.637	[-1.059, 2.392]	0.104
Prednisolone	0	0	∞	-	∞	-	-	-	∞	-	-
Example 1.2.2		1/3	4.399	-	7.560	-	-	-	12.356	-	-
		1	2.609	0.219	4.000	[2.179, 3.038]	0.107	0.397	31.034	[1.830, 3.387]	0.025
	0.5	0	3.309	1.522	6.868	[0.326, 6.291]	0.434	1.522	8.868	[0.326, 6.291]	0.434
		1/3	2.792	0.885	5.099	[1.059, 4.526]	0.340	0.891	6.054	[1.047, 4.538]	0.288
		1	2.333	0.460	3.750	[1.431, 3.235]	0.241	0.519	12.516	[1.316, 3.351]	0.081

with its sign indicating the direction (positive, negative) of the underlying association and $|\mathcal{A}_c^\phi|$ expressing the relative strength of association in the sample $\mathbf{n}$ for $c \geq 0$. For example, for the larynx data without correction ($c = 0$) in case of $\lambda = 1/3$, the association corresponds to $\mathcal{A}_0^{\phi_{1/3}} = 9.76\%$ of the maximum possible association and is non-significant. For the prednisolone data with correction $c = 0.5$ and the same $\lambda = 1/3$, the effect of prednisolone on normalisation is positive and it is significant. It corresponds to $\mathcal{A}_{0.5}^{\phi_{1/3}} = 54.76\%$ of the maximum possible association.

3.6 Evaluation Studies

Asymptotic Wald 95% CIs for the θ^{ϕ_λ} are evaluated, for different values of λ, in terms of their coverage probability (CP)

$$\mathrm{P}\left(\theta_0^{\phi_\lambda} \in CI_\alpha^\phi(\mathbf{N} + c\mathbf{1}_{2\times 2})\right) = \mathbb{E}(\mathbf{1}_{\{\theta_0^{\phi_\lambda} \in CI_\alpha^\phi(\mathbf{n}+c\mathbf{1}_{2\times 2})\}}) = \sum_{\mathbf{n}\in\mathcal{S}} \mathbf{1}_{\{\theta_0^{\phi_\lambda} \in CI_\alpha^\phi(\mathbf{n}+c\mathbf{1}_{2\times 2})\}} \mathrm{P}(\mathbf{N} = \mathbf{n}), \tag{3.26}$$

and their average relative length (ARL)

$$ARL_c = \mathbb{E}(RL_{c,\mathcal{B}}^\phi(\mathbf{N})) = \sum_{\mathbf{n}\in\mathcal{S}} RL_{c,\mathcal{B}}^\phi(\mathbf{n})\mathrm{P}(\mathbf{N} = \mathbf{n}) \; . \tag{3.27}$$

Here, $\theta_0^{\phi_\lambda}$ is the true value of the power divergence scaled odds ratio and $\mathcal{S}$ is the set of all possible 2×2 tables $\mathbf{n}$ of fixed sample size n or fixed row marginals n_1, n_2, for multinomial or independent binomial sampling scheme.

The adopted sampling scheme in this study is the independent binomial with focus lies on relatively small sample sizes ($n_i \leq 30, i = 1, 2$). CP (3.26) of the two-sided 95% Wald CIs for the ϕ-scaled odds ratio can be calculated using the R functions `cov.prob.phi` and `cov.prob.phi.multi` for the product binomial and multinomial sampling scheme, respectively (see Appendix B.3). In addition, this functions are also available for the log-scale (`cov.prob.log` and `cov.prob.log.multi`). In the evaluation study, the constant cell correction ($c = 0.5$) is used to enable contemplation of the power divergences with $\lambda < 1$. Otherwise, the function `cov.prob.phi.no.correction` calculates the CP und the condition of no sampling zeros. The ARL (3.27) can be calculated using `calc.ARL`.

The CP and ARL for tables of balanced ($n_1 = n_2 = 10, 15$) and unbalanced ($n_1 = 10,\ n_2 = 30$) row marginals, for selected combinations of binomial parameters $\pi_1,\ \pi_2$, and for $\lambda \in \{-1/3, 0, 1/3, 2/3, 1\}$ are given in Table 3.6. Since the sample sizes considered are relatively small, the CP and ARL are computed analytically by generating each time all possible **n** tables and applying (3.26) and (3.27) directly. Observe that, regarding the CP, $\lambda = -1/3$ gives unstable coverage probabilities that range from 0.826 to 1 and are mostly farther apart from the nominal level than the CPs of the CIs corresponding to the other values of λ. For the remaining λ's, there is no evidence for conclusions of absolute preference over all parameter scenarios and sample size combinations. Note that for balanced samples, the CPs for $\lambda = 1/3$ are closer to the nominal level than the classical log-odds ratio CI's ($\lambda = 0$). $\pi_1 >> \pi_2$ is an exception. In the unbalanced sample size scenarios, the better behaviour of $\lambda = 1/3$ over $\lambda = 0$ in terms of CP becomes clearer (exceptions are the cases $(\pi_2, \pi_2) \in \{(0.7, 01), (0.7, 0.05), (0.7, 0.025), (0.4, 0.025)\}$). With regard to the ARL, the classical log-odds ratio exhibits the shortest ARL. There are only few exceptions.

To have a more complete view over the whole parameter space, Figure 3.3 presents contour plots of the CP and ARL for $(\pi_1, \pi_2) \in (0, 1)^2$ for different choices of λ and $n_1 = n_2 = 15$. Here, π_1, π_2 range from 0.01 to 0.99 by increments of 0.01. These contour plots show that, in terms of coverage, the log-odds ratio is too conservative when both π_1 and π_2 are very small (or large), and degenerate at the borders of the parameter space. On the contrary, the $\lambda = 1/3$ case has a more uniform CP over the parameter space that is closer to the nominal level. The **HRL** separation, introduced in the context of log-odds ratio (see Figure 2.2), is detected also here. Agresti [1999] pointed out, that the log-odds ratios have degenerate coverage probabilities for $|\log \theta| > 4$ (i.e. $\theta > 54.598$ or $\theta < 0.0183$), which corresponds to the region $\mathbf{H}_{0,4}$. Focusing further on the separation for $k = 4$ and choosing $p = 0.1$, the summary statistics for the CP and ARL based on the corresponding contour plot data of Figure 3.3 are provided in Table 3.7. This is done for the whole parameter space as well as for the subspaces $\mathbf{R}_{0.1}$, $\mathbf{L}_{0.1,4}$ and $\mathbf{H}_{0.1,4}$. The summary statistics verify that the choice $\lambda = 1/3$ is closer to the nominal level with the cost of a slightly higher ARL. Note that for the subspace $\mathbf{H}_{0.1,4}$, the ARL for the $\lambda = 1/3$ case is lower than for the log-odds ratio. In addition, Table 3.8 shows the summary statistics for the more extreme high association setting $\mathbf{H}_{0,6}$. Further options for k and p support these observations as well.

To gain some further insight, line-plots of the CP and ARL for selected subspaces are shown in Figure 3.4. They are in agreement with the outcomes above. The scale effect of the parameter λ is visualized in Figure 3.5. Hereby, the curves of constant $\log \theta$ (=1,2,3,4) over the parameter space of (π_1, π_2) along with the curves of constant $\theta^{\phi_{1/3}}$ (=0.99, 1.93, 2.78, 3.53) are provided in solid (black) and dashed (blue) lines for balanced and unbalanced sample sizes. The values of constant $\theta^{\phi_{1/3}}$ are chosen so that $\theta^{\phi_{1/3}} = \log \theta$ on the line $\pi_1 = 1 - \pi_2$.

Furthermore, following Agresti [1999], a study is conducted for the CP and ARL of 95%

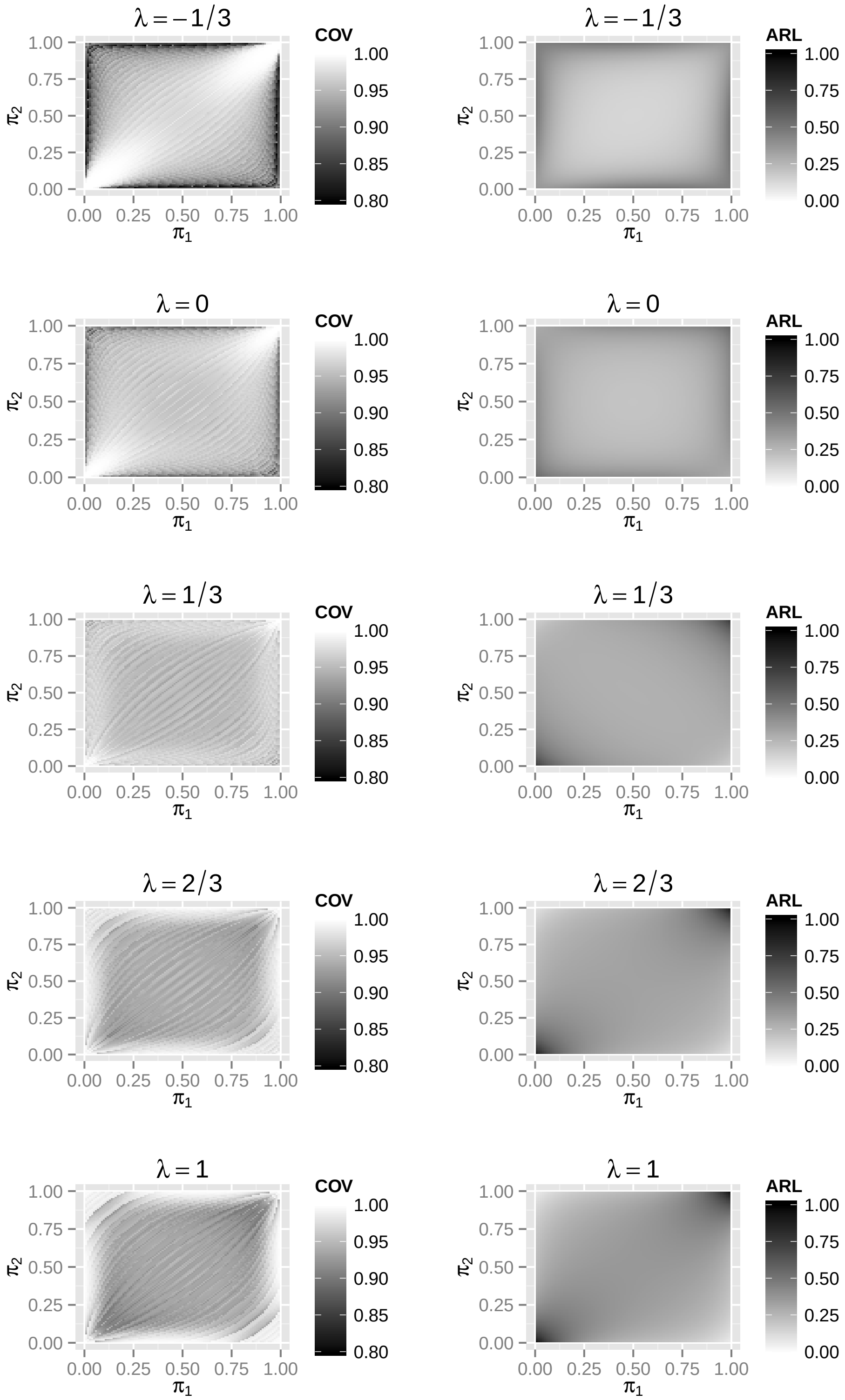


Figure 3.3: Contour plots for the CP (left) and the ARL (right) for the θ^{ϕ_λ}'s 95% CIs ($\lambda = -1/3, 0, 1/3, 1$) with continuity correction $c = 0.5$ under the independent binomial setting with $n_1 = n_2 = 15$ for π_1, π_2 ranging from 0.01 to 0.99 by 0.01.

Table 3.6: Coverage probabilities and average relative length for the 95% CIs of the θ^{ϕ_λ} ($\lambda \in \{-1/3, 0, 1/3, 2/3, 1\}$) under the independent binomial sampling scheme, for three choices of tables' sample sizes, using continuity correction $c = 0.5$.

		Coverage Probability														
		$(n_1, n_2) = (10, 10)$					$(n_1, n_2) = (15, 15)$					$(n_1, n_2) = (10, 30)$				
π_1	π_2	−1/3	0	1/3	2/3	1	−1/3	0	1/3	2/3	1	−1/3	0	1/3	2/3	1
0.700	0.700	0.997	0.980	0.963	0.923	0.923	0.993	0.973	0.960	0.944	0.919	0.986	0.972	0.950	0.937	0.925
	0.500	0.982	0.978	0.961	0.954	0.934	0.978	0.966	0.959	0.949	0.943	0.971	0.973	0.953	0.940	0.936
	0.300	0.958	0.971	0.962	0.932	0.932	0.961	0.959	0.943	0.950	0.949	0.967	0.971	0.950	0.941	0.945
	0.100	0.938	0.962	0.967	0.974	0.974	0.930	0.960	0.965	0.954	0.955	0.957	0.956	0.966	0.949	0.943
	0.050	0.897	0.952	0.978	0.967	0.975	0.903	0.942	0.966	0.979	0.974	0.943	0.956	0.972	0.974	0.970
	0.025	0.856	0.909	0.955	0.989	0.972	0.826	0.927	0.970	0.984	0.992	0.912	0.961	0.964	0.978	0.977
0.500	0.500	0.990	0.960	0.958	0.948	0.948	0.984	0.957	0.957	0.954	0.943	0.992	0.968	0.946	0.944	0.930
	0.300	0.982	0.978	0.961	0.954	0.934	0.978	0.966	0.959	0.949	0.943	0.989	0.967	0.949	0.942	0.928
	0.100	0.937	0.959	0.967	0.963	0.963	0.945	0.969	0.968	0.964	0.954	0.979	0.974	0.951	0.937	0.934
0.400	0.400	0.992	0.965	0.960	0.940	0.940	0.986	0.961	0.959	0.951	0.934	0.989	0.971	0.949	0.943	0.930
	0.200	0.988	0.979	0.938	0.917	0.917	0.972	0.969	0.946	0.930	0.924	0.993	0.973	0.951	0.944	0.934
	0.100	0.962	0.973	0.969	0.969	0.954	0.956	0.969	0.962	0.960	0.935	0.982	0.975	0.952	0.934	0.929
	0.050	0.890	0.947	0.978	0.980	0.974	0.906	0.945	0.971	0.982	0.982	0.955	0.978	0.966	0.966	0.942
	0.025	0.846	0.903	0.975	0.989	0.993	0.845	0.934	0.969	0.985	0.992	0.906	0.959	0.977	0.969	0.969
0.300	0.100	0.985	0.980	0.977	0.964	0.963	0.964	0.976	0.971	0.946	0.953	0.998	0.988	0.955	0.937	0.920
	0.050	0.962	0.964	0.972	0.971	0.974	0.933	0.962	0.968	0.979	0.974	0.988	0.983	0.972	0.943	0.927
0.200	0.200	0.999	0.995	0.970	0.913	0.913	0.998	0.989	0.957	0.933	0.904	0.985	0.971	0.956	0.927	0.906
	0.100	0.999	0.995	0.986	0.956	0.956	0.995	0.987	0.970	0.938	0.938	0.999	0.992	0.970	0.920	0.915
	0.050	0.973	0.984	0.984	0.983	0.983	0.959	0.978	0.979	0.977	0.971	1.000	0.997	0.968	0.947	0.907
	0.025	0.903	0.961	0.962	0.968	0.969	0.911	0.955	0.967	0.979	0.979	0.988	0.985	0.971	0.966	0.966
0.100	0.100	1.000	1.000	0.991	0.950	0.950	1.000	0.999	0.976	0.922	0.915	0.995	0.981	0.972	0.963	0.947
	0.050	1.000	0.999	0.995	0.969	0.962	1.000	0.997	0.989	0.957	0.957	1.000	0.999	0.979	0.959	0.917
	0.025	0.999	0.990	0.991	0.991	0.991	0.996	0.986	0.986	0.987	0.987	1.000	1.000	0.997	0.980	0.934
0.050	0.050	1.000	1.000	0.999	0.986	0.986	1.000	1.000	0.995	0.966	0.966	0.999	0.993	0.977	0.977	0.975
	0.025	1.000	1.000	0.999	0.985	0.984	1.000	1.000	0.997	0.975	0.975	1.000	1.000	0.994	0.994	0.956

		Average Relative Length														
		$(n_1, n_2) = (10, 10)$					$(n_1, n_2) = (15, 15)$					$(n_1, n_2) = (10, 30)$				
π_1	π_2	−1/3	0	1/3	2/3	1	−1/3	0	1/3	2/3	1	−1/3	0	1/3	2/3	1
0.700	0.700	0.248	0.322	0.393	0.450	0.486	0.161	0.231	0.301	0.359	0.395	0.161	0.222	0.275	0.303	0.302
	0.500	0.238	0.306	0.369	0.419	0.451	0.156	0.220	0.283	0.333	0.365	0.168	0.217	0.256	0.273	0.265
	0.300	0.268	0.318	0.362	0.393	0.412	0.177	0.231	0.279	0.315	0.335	0.180	0.222	0.255	0.272	0.271
	0.100	0.392	0.375	0.359	0.345	0.335	0.296	0.295	0.289	0.282	0.276	0.236	0.258	0.271	0.280	0.286
	0.050	0.458	0.403	0.356	0.322	0.301	0.384	0.338	0.294	0.263	0.246	0.309	0.293	0.275	0.261	0.255
	0.025	0.500	0.420	0.353	0.307	0.281	0.448	0.367	0.297	0.250	0.226	0.383	0.324	0.273	0.237	0.219
0.500	0.500	0.215	0.289	0.361	0.420	0.457	0.142	0.209	0.277	0.333	0.369	0.142	0.201	0.252	0.280	0.279
	0.300	0.238	0.306	0.369	0.419	0.451	0.156	0.220	0.283	0.333	0.365	0.146	0.206	0.260	0.293	0.299
	0.100	0.371	0.385	0.396	0.403	0.406	0.274	0.294	0.308	0.318	0.324	0.199	0.247	0.292	0.325	0.346
0.400	0.400	0.222	0.297	0.369	0.427	0.464	0.146	0.214	0.282	0.339	0.375	0.146	0.206	0.257	0.285	0.284
	0.200	0.275	0.337	0.394	0.439	0.467	0.183	0.244	0.303	0.349	0.378	0.154	0.219	0.281	0.323	0.338
	0.100	0.362	0.393	0.418	0.437	0.447	0.263	0.296	0.324	0.344	0.355	0.192	0.251	0.308	0.351	0.376
	0.050	0.438	0.439	0.435	0.429	0.425	0.352	0.349	0.340	0.332	0.327	0.267	0.299	0.329	0.354	0.373
	0.025	0.487	0.468	0.444	0.423	0.409	0.419	0.387	0.351	0.322	0.305	0.349	0.346	0.343	0.344	0.348
0.300	0.100	0.355	0.406	0.449	0.481	0.499	0.253	0.303	0.348	0.381	0.400	0.193	0.262	0.329	0.379	0.406
	0.050	0.423	0.452	0.472	0.484	0.488	0.333	0.355	0.369	0.376	0.379	0.260	0.310	0.358	0.397	0.422
0.200	0.200	0.300	0.376	0.447	0.502	0.535	0.197	0.271	0.344	0.402	0.439	0.193	0.258	0.313	0.343	0.342
	0.100	0.358	0.430	0.494	0.542	0.567	0.251	0.321	0.384	0.433	0.462	0.208	0.287	0.360	0.411	0.431
	0.050	0.412	0.475	0.526	0.562	0.576	0.316	0.370	0.415	0.445	0.460	0.259	0.332	0.402	0.452	0.478
	0.025	0.450	0.503	0.545	0.572	0.578	0.366	0.407	0.434	0.448	0.452	0.321	0.377	0.431	0.467	0.487
0.100	0.100	0.385	0.479	0.562	0.626	0.651	0.276	0.365	0.447	0.509	0.544	0.253	0.337	0.407	0.443	0.443
	0.050	0.418	0.520	0.610	0.679	0.699	0.316	0.410	0.495	0.558	0.589	0.278	0.377	0.468	0.522	0.540
	0.025	0.442	0.546	0.639	0.709	0.723	0.349	0.443	0.526	0.585	0.611	0.318	0.417	0.514	0.568	0.591
0.050	0.050	0.435	0.557	0.670	0.758	0.775	0.336	0.451	0.559	0.642	0.678	0.302	0.414	0.515	0.565	0.572
	0.025	0.448	0.582	0.706	0.804	0.818	0.355	0.481	0.601	0.693	0.730	0.326	0.452	0.573	0.635	0.657

CIs for θ^{ϕ_λ} with $\lambda \in \{-\frac{1}{3}, 0, \frac{1}{3}, \frac{2}{3}, 1\}$ under the binomial setting. In particular, $K = 1000$ random points $(\pi_1, \pi_2) \in (0,1)^2$ were generated. For each of them the associated CP and ARL were computed for tables with row marginal counts $n_1 = n_2 = 15$. The derived CP

Table 3.7: Summary statistics of the CP and the ARL of the contour plot data corresponding to the full space $(0,1)^2$ and the subspaces $\mathbf{R}_p, \mathbf{L}_{p,k}, \mathbf{H}_{p,k}, p = 0.1, k = 4$. The number of data points considered is 200, 8635 and 966 for the $\mathbf{R}_{0.1}$, $\mathbf{L}_{0.1,4}$ and $\mathbf{H}_{0.1,4}$ subspaces.

		Coverage						ARL					
	Space	Min.	1st Q.	Median	Mean	3rd Q.	Max.	Min.	1st Q.	Median	Mean	3rd Q.	Max.
$\lambda = -1/3$	whole	0.740	0.938	0.965	0.952	0.978	1.000	0.142	0.172	0.227	0.255	0.323	0.499
	$\mathbf{R}_{0.1}$	0.884	0.999	1.000	0.994	1.000	1.000	0.276	0.318	0.342	0.338	0.360	0.379
	$\mathbf{L}_{0.1,4}$	0.801	0.948	0.967	0.959	0.978	1.000	0.142	0.167	0.211	0.233	0.284	0.457
	$\mathbf{H}_{0.1,4}$	0.740	0.852	0.879	0.879	0.912	0.950	0.343	0.408	0.439	0.437	0.467	0.499
$\lambda = 0$	whole	0.740	0.957	0.964	0.961	0.970	1.000	0.209	0.233	0.271	0.285	0.324	0.550
	$\mathbf{R}_{0.1}$	0.919	0.995	0.999	0.994	1.000	1.000	0.365	0.416	0.451	0.450	0.479	0.550
	$\mathbf{L}_{0.1,4}$	0.886	0.959	0.965	0.964	0.971	0.999	0.209	0.229	0.260	0.274	0.305	0.462
	$\mathbf{H}_{0.1,4}$	0.740	0.902	0.927	0.923	0.942	0.964	0.299	0.323	0.339	0.349	0.372	0.417
$\lambda = 1/3$	whole	0.927	0.953	0.959	0.960	0.968	1.000	0.180	0.281	0.293	0.313	0.323	0.723
	$\mathbf{R}_{0.1}$	0.956	0.980	0.988	0.986	0.994	1.000	0.447	0.506	0.545	0.553	0.592	0.723
	$\mathbf{L}_{0.1,4}$	0.934	0.952	0.959	0.959	0.967	0.992	0.272	0.282	0.294	0.311	0.323	0.534
	$\mathbf{H}_{0.1,4}$	0.927	0.961	0.967	0.965	0.971	0.979	0.180	0.257	0.273	0.276	0.296	0.369
$\lambda = 2/3$	whole	0.911	0.938	0.948	0.953	0.968	0.999	0.111	0.300	0.330	0.335	0.354	0.864
	$\mathbf{R}_{0.1}$	0.922	0.962	0.974	0.973	0.987	0.999	0.509	0.575	0.617	0.631	0.679	0.864
	$\mathbf{L}_{0.1,4}$	0.911	0.937	0.946	0.950	0.958	0.995	0.243	0.311	0.332	0.340	0.356	0.583
	$\mathbf{H}_{0.1,4}$	0.940	0.978	0.984	0.982	0.989	0.997	0.111	0.204	0.231	0.227	0.249	0.331
$\lambda = 1$	whole	0.900	0.933	0.940	0.948	0.963	0.999	0.080	0.300	0.354	0.348	0.382	0.909
	$\mathbf{R}_{0.1}$	0.915	0.940	0.962	0.961	0.981	0.999	0.544	0.607	0.648	0.665	0.713	0.909
	$\mathbf{L}_{0.1,4}$	0.900	0.932	0.938	0.944	0.952	0.998	0.224	0.319	0.359	0.357	0.384	0.600
	$\mathbf{H}_{0.1,4}$	0.933	0.975	0.988	0.982	0.992	0.998	0.080	0.176	0.207	0.200	0.226	0.307

Table 3.8: Summary statistics for the coverage probability (a) and the ARL (b) of the contour plot data for the subspace defined by $\{|\log\theta| > 6, \pi_1 = 0.01, ..., 0.99, \pi_2 = 0.01, ..., 0.99\}$.

Divergence	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
$\lambda = -1/3$	0.740^a	0.868	0.880	0.880	0.898	0.931
	0.447^b	0.458	0.471	0.467	0.474	0.482
$\lambda = 0$	0.740	0.898	0.913	0.910	0.927	0.951
	0.299	0.317	0.324	0.328	0.338	0.362
$\lambda = 1/3$	0.927	0.951	0.959	0.957	0.968	0.978
	0.180	0.212	0.224	0.226	0.238	0.269
$\lambda = 1$	0.982	0.988	0.991	0.991	0.994	0.997
	0.080	0.115	0.131	0.132	0.147	0.180

and ARL are presented in Figure 3.6, as functions of $\log\theta$.

Observe that the log-odds ratio produces degenerate coverage probabilities for $|\log\theta| > 4$, as also pointed out by Agresti [1999]. For the $\lambda = 1/3$-scaled odds ratio, the coverage probabilities are more stable and they are degenerate for about $|\log\theta| > 8$. Hence, the effect described by Agresti, occurs for θ^{ϕ_λ} with $\lambda = 1/3$ as well, but for greater $|\log\theta|$ than for the log-odds ratio and even then, it is not that strong. Overall, the CPs for $\lambda = 1/3$ are closer to the nominal of 95% than for $\lambda = 0$. For $|\log\theta| > 4$ (≤ 4), the ARL tends to be decreasing (increasing) in λ (see also Table 3.5). Note that the plot of the ARL has a convex (concave) structure for $\lambda < 1/3$ ($> 1/3$). For $\lambda = 1/3$, the ARL values are more centred around a single value (which depends on the sample sizes) over the whole range of $\log\theta$. For $|\log\theta| < 4$ the ARL for $\lambda = 1/3$ is higher than for $\lambda = 0$.

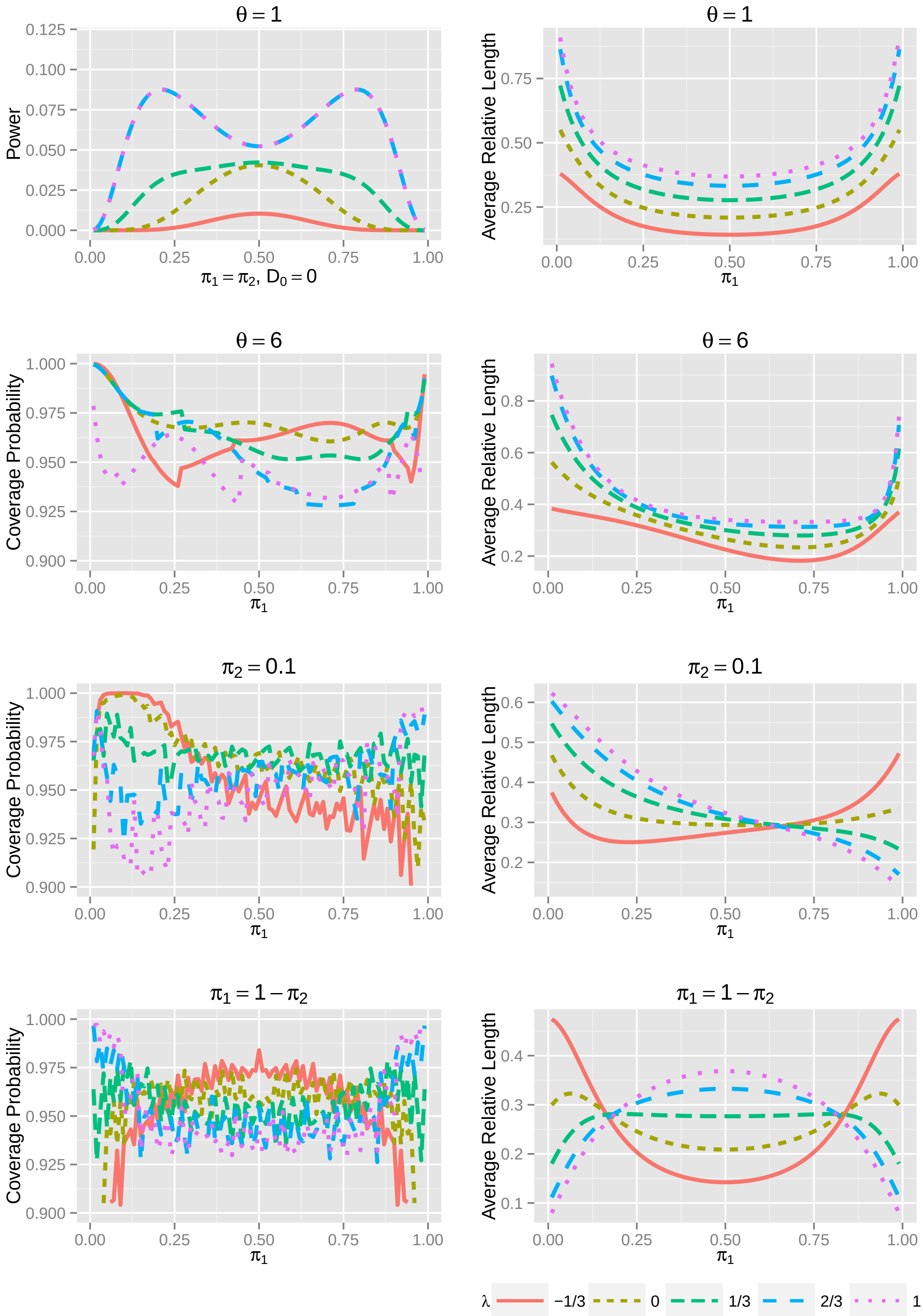


Figure 3.4: Line-plots for the CP (left) and the ARL (right) of 95% CIs for the θ^{ϕ_λ} ($\lambda \in \{-1/3, 0, 1/3, 2/3, 1\}$) in the binomial setting with row marginals $n_1 = n_2 = 15$ and continuity correction $c = 0.5$, when $\pi_1 = 0.01, ..., 0.99$ and π_2 such that $\theta = 1,\ 6$ (upper graphs), or $\pi_2 = 0.1$, $\pi_2 = 1 - \pi_1$ (lower graphs).

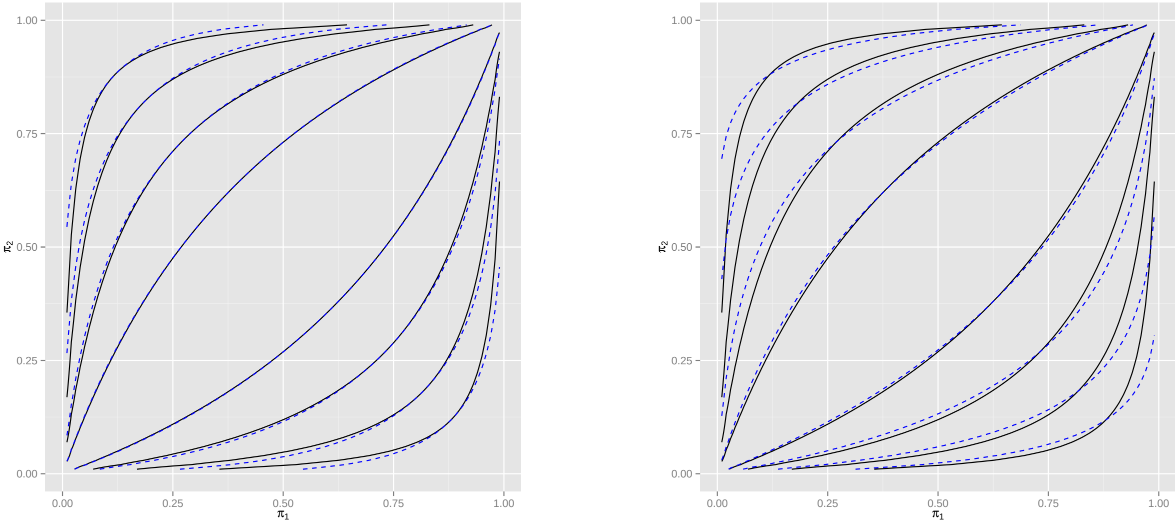


Figure 3.5: Curves of constant $\log\theta$ (=1,2,3,4) (solid) and constant $\theta^{\phi_{1/3}}$ (=0.99, 1.93, 2.78, 3.53) (dashed) for $n_1 = n_2$ (left) and $n_2 = 4n_1$ (right).

Table 3.9: Summary statistics for the CP and ARL of the 95% $c = 0.5$-corrected CIs for θ^{ϕ_λ} ($\lambda \in \{-1/3, 0, 1/3, 2/3, 1\}$) for 10,000 randomly (uniformly) generated pairs of binomial distributions $\pi_1, \pi_2 \in (0,1)$ and $n_1, n_2 \in \{5, 10, 15, \ldots, 50\}$.

	CP					ARL				
Power. Div., λ	$-1/3$	0	1/3	2/3	1	$-1/3$	0	1/3	2/3	1
Min.	0.000	0.000	0.663	0.884	0.828	0.046	0.088	0.106	0.058	0.046
1st Qu.	0.940	0.954	0.951	0.942	0.935	0.105	0.153	0.188	0.201	0.204
Median	0.960	0.960	0.955	0.947	0.943	0.172	0.211	0.228	0.239	0.243
Mean	0.938	0.953	0.959	0.953	0.947	0.206	0.229	0.248	0.261	0.263
3rd Qu.	0.972	0.968	0.966	0.963	0.956	0.286	0.288	0.287	0.298	0.300
Max.	1.000	1.000	1.000	1.000	1.000	0.627	0.757	0.815	0.913	0.933
mean |COV-0.95|	0.040	0.022	0.009	0.013	0.016	-	-	-	-	-
P(COV<0.92)	0.172	0.059	0.001	0.004	0.056	-	-	-	-	-
min for										
$\lvert\log\theta\rvert \le \log(5)$	0.000	0.000	0.935	0.884	0.828	0.046	0.088	0.133	0.063	0.046
$\lvert\log\theta\rvert \le \log(10)$	0.000	0.000	0.935	0.884	0.828	0.046	0.088	0.133	0.058	0.046
$\lvert\log\theta\rvert \le \log(20)$	0.000	0.000	0.892	0.884	0.828	0.046	0.088	0.133	0.058	0.046

For an overall evaluation, a simulation study was carried out along the lines of an analogous study of Agresti [1999] for the log-odds ratio. $K = 10,000$ randomly generated pairs of independent binomial distributions $\mathcal{B}(n_i, \pi_i)$, $i = 1, 2$, with π_i being uniform distributed in $(0,1)$ and n_i discrete uniform in $\{5, 10, 15, \ldots, 50\}$ were sampled. The associated CP and ARL were computed for the 95% CIs for the θ^{ϕ_λ} ($\lambda \in \{-1/3, 0, 1/3, 2/3, 1\}$) with a continuity correction $c = 0.5$. The summary statistics, provided in Table 3.9, indicate that with respect to the CP, the $\lambda = 1/3$ case is preferable over the other considered choices. Indeed, its CP is more concentrated around the nominal level. In addition, it has the smallest probability of falling below 0.92, and a minimum value of 0.935, when $|\log\theta| \le \log(10)$. Its CP-minimum of $\log(\theta)$ is 0. On the other hand, in terms of ARL, the log-odds ratio is slightly better than the measure corresponding to $\lambda = 1/3$.

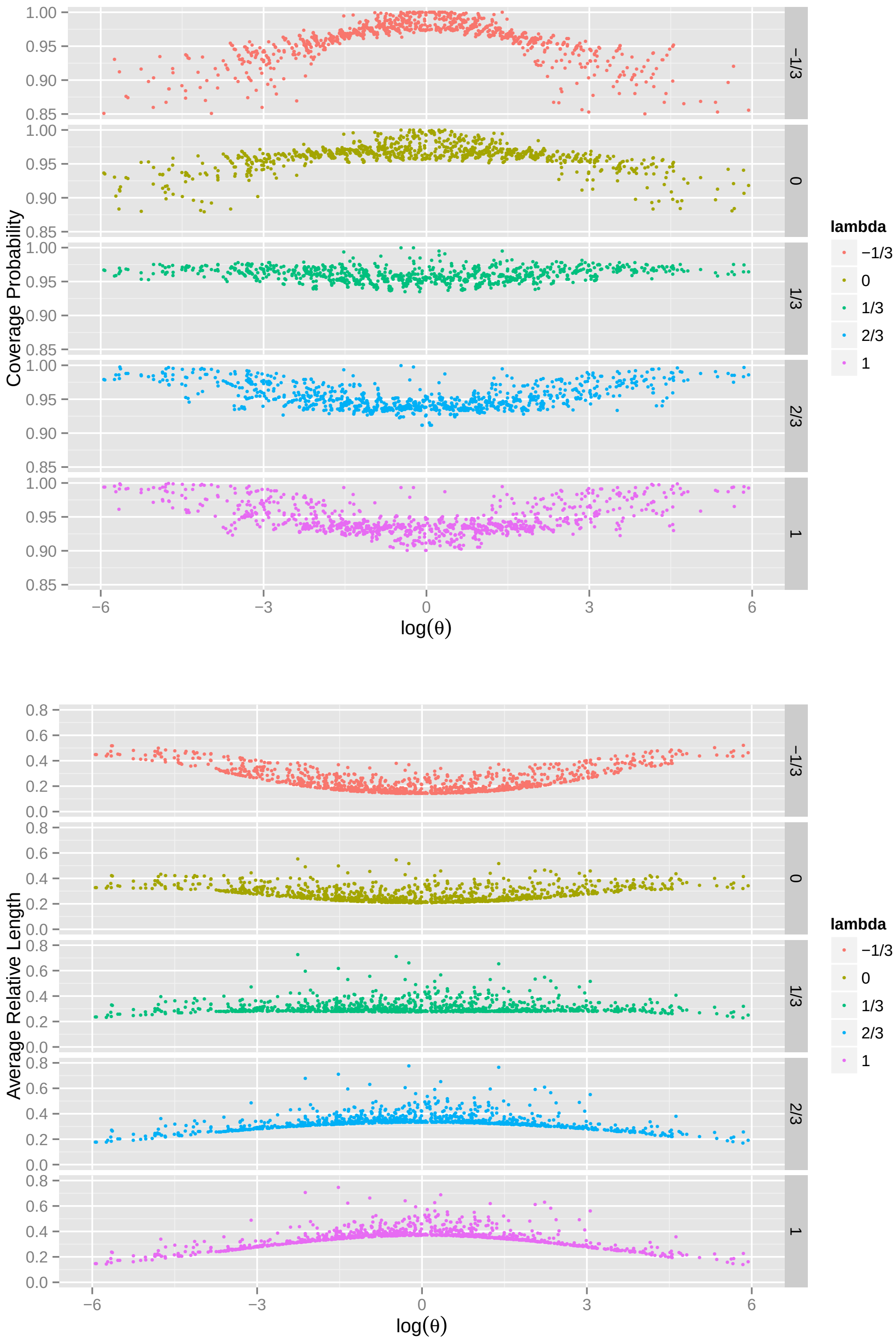


Figure 3.6: Plot of (i) CPs and (ii) ARLs of the 95% CIs for the θ^{ϕ_λ} ($\lambda \in \{-1/3, 0, 1/3, 2/3, 1\}$) against $\log\theta$ in the binomial setting with row marginals $n_1 = n_2 = 15$ and continuity correction $c = 0.5$ for $K = 1000$ randomly generated points (π_1, π_2).

To summarise, it was verified that in $\mathbf{H}_{0.1,4}$ scenarios the $\lambda = 1/3$-scaled odds ratio with $c = 0.5$ leads to asymptotic CIs that perform better than the log-odds ratio in terms of CP and ARL. In all other settings, its ARL is increased in comparison to that of the log-odds ratio $\log(\theta)$. Thus, the use of $\theta^{\phi_{1/3}}$ is suggested whenever $\log(\theta)$ is expected to be larger than four. Note that $\theta^{\phi_{1/3}}$ is not able to solve the degeneration problem of the CIs for extremely small probabilities ($\pi_1, \pi_2 < 0.05$), which gives CPs near 100% in these situations. In general, the region of 'bad coverage' within the $\mathbf{R}_{0.1}$ setting becomes smaller in comparison to $\log(\theta)$ at the cost of higher ARL. The use of $\lambda = 1/3$ is suggested at the cost of higher ARL, if CP should not fall below the nominal level. Use of the classical log-odds ratio is suggested ($\lambda = 0$), if nominal compliance is not important. Finally, to use the classical $\log(\theta)$ with constant cell correction $c = 0.5$ is suggested in $\mathbf{L}_{0.1,4}$ settings, though it is more conservative due to the lower ARL.

Asymptotic tests for testing $H_0 : \theta^{\phi}(\boldsymbol{\pi}) = D_0$ vs $H_1 : \theta^{\phi}(\boldsymbol{\pi}) \neq D_0, D_0 \in \mathbb{R}$ are evaluated on basis of the type I error rate (TIER). Let $Z^{\phi}_{\mathcal{B}} = Z^{\phi}_{\mathcal{B}}(\mathbf{n})$ be the test statistic based on the ϕ-scaled odds ratio as in (3.18) for a realisation $\mathbf{n} = (n_{ij}) \in \mathbb{N}^4_0$. Let $u_{1-\alpha/2}$ be the $(1-\alpha/2)$-quantile of the standard normal distribution such that the test has α-level rejection region $\mathbf{R} = \{\mathbf{n} \mid |Z^{\phi}_{\mathcal{B}}| > u_{1-\alpha/2}\}$. The type I error rate is given by

$$\mathbf{P}_{D_0}(\mathbf{R}) = \sum_{\mathbf{n} \in \mathcal{S}} \mathbf{1}_{\{|Z^{\phi}_{\mathcal{B}}| > u_{1-\alpha/2}\}} \mathbf{P}(\mathbf{N} = \mathbf{n}), \tag{3.28}$$

where $\mathbf{P}_{D_0}$ is the probability under H_0. For $D_0 = 0$ the underlying probability is the independence probability. In the binomial sampling scheme, one can write $\mathbf{P}_0 = \mathbf{P}_{\pi_1}$ as it holds $\pi_1 = \pi_2$ in the case of independence. The TIER is calculated using the R function `eval.error0` (see Appendix B.3).

Plots $\pi_1 \mapsto \mathbf{P}_{\pi_1}(\mathbf{R})$ have been made for the sample sizes $(n_1, n_2) \in \{5, 10, ..., 50\}$. Figure 3.7 shows some selected examples. Step-width is $\pi_1 = 0.01, 0.02, ..., 0.99$. Based on these data points, boxplots (Figure 3.8) and summary statistics (Table 3.10) are given on the subspace $n_1 \leq n_2$, where the cases $n_1 > n_2$ are cancelled due to symmetry.

The plots indicate a tendency for higher λ values to lead to higher type I errors of the corresponding independence test. This tendency fails for high unbalanced samples ($n_1 = 5, n_2 \geq 20$ or $n_1 = 10, n_2 \geq 35$) in the case of rare probabilities ($\pi_1 < 0.25$ or $\pi_1 > 0.75$), which have a non-monotonic ordering of the type I error rates. The test based on log-scale is very conservative, especially at the rare probability scenarios. Thus, an increase of λ reduces the conservatism. The $\lambda = 1/3$-scale is an improvement for the log-scale. It is more stable at the $\alpha = 95\%$-level and breaks the $\alpha = 95\%$-level only a few times at a low size. The use of ϕ-scaled odds ratio to test independence must be rejected when the nominal limit is important. The $\lambda = 1/3$-scale with c-correction ($c = 0.5$) should be used as basis for an independence test if a small violation of the α-level is acceptable. The boxplots and summary statistics support this suggestion.

3.7 Discussion

A family of association measures for 2×2 tables is proposed, which are scaled by the ϕ-divergence. The classical log-odds ratio is a member of this family. The new measures share the same properties as the odds ratio. Furthermore, asymptotic CIs and tests are constructed. Focusing on the measures that correspond to the power divergence of Cressie and Read [1984], extensive evaluation studies were conducted with regard to the CP of the associated 95% asymptotic CIs and their ARL, together with the type I error rates

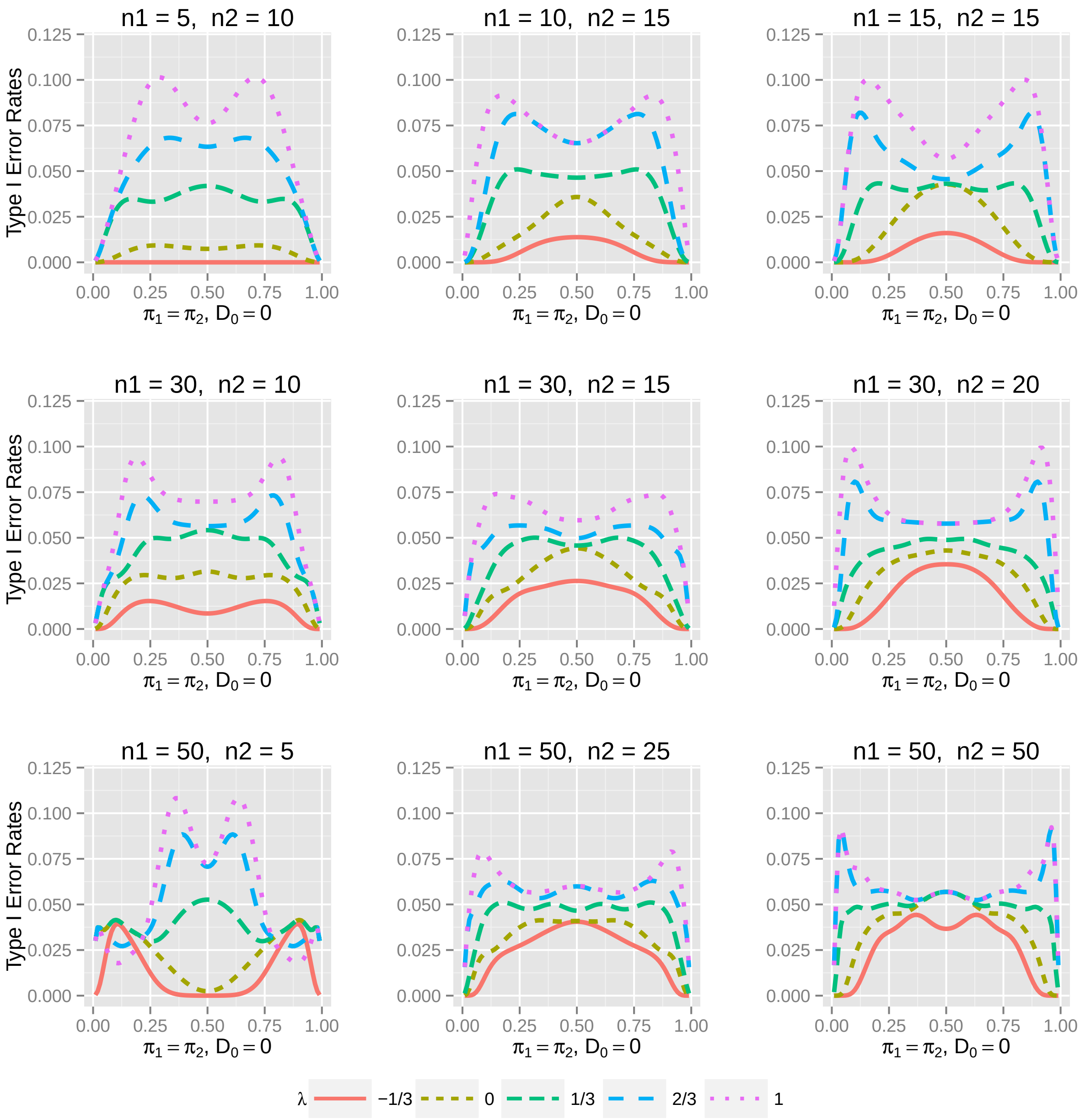


Figure 3.7: Type I error of the $\alpha = 95\%$ independence test based on the c-corrected, $c = 0.5$, ϕ-scaled odds ratios for the two-sided hypothesis $H_0 : \theta^{\phi}_{\mathcal{B}} = 0 =: D_0$ vs $H_1 : \theta^{\phi}_{\mathcal{B}} \neq 0$. The ϕ-scale is the power divergence family, $\lambda = -1/3, 0, 1/3, 2/3, 1$. Plot stepwidth is $\pi_1 = 0.01, 0.02, ..., 0.99$.

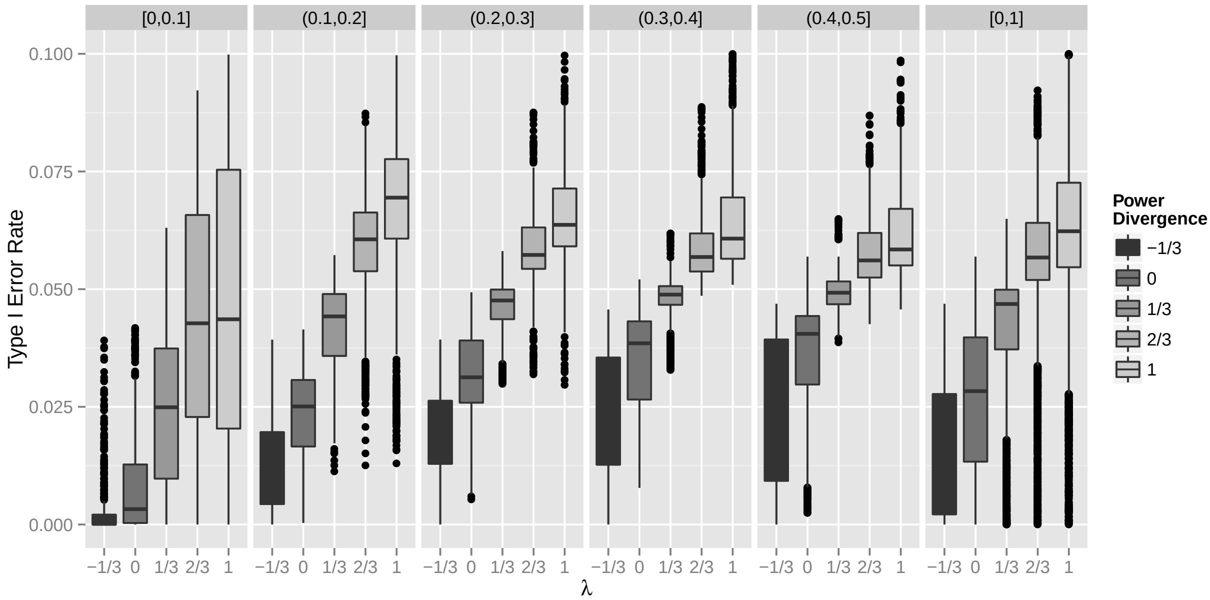


Figure 3.8: Boxplots for the type I error rate of the c-corrected, $c = 0.5$, ϕ-scaled odds ratio test for independence ($\alpha = 95\%$) taken over all scenarios $n_1 \leq n_2, (n_1, n_2) \in \{5, 10, ..., 50\}^2$ in the power divergence family (Example 1.10.4). Step-width is $\pi_1 = 0.01, 0.02, ..., 0.99$, where the single boxplots are restricted to special value ranges $\pi_1 \in (0, 0.1], ..., (0.4, 0.5]$ and the full range $\pi_1 \in [0, 1]$.

of the corresponding independence tests. The CIs are compared in terms of their relative length, since due to their scale difference CIs for different ϕ-scaled odds ratios cannot be compared in terms of their average length. In case of relatively small sample sizes, the CPs and ARLs have been computed analytically. The parameter space separation into subspaces of high, low association or rare probabilities, identified in the evaluation studies, provides further insight into the pattern of association strength underlying the (π_1, π_2) space. Agresti's statement about the low coverage of the log-odds ratio CIs in situations of extremely high association ($|\log \theta| > 4$) was verified. In such cases, the use of $\lambda = 1/3$ is suggested, since the corresponding CI improves the coverage when approaching the borders of the parameter space. Overall, the $\lambda = 1/3$-odds ratio CI is less conservative than the classical log-odds ratio CI, but gives in most cases higher ARL. The type I error rates for the asymptotic tests of independence based on the ϕ-scaled odds ratio support the observation of less conservative tests closer to the nominal for $\lambda = 1/3$. All studies presented here were performed under the independent binomial sampling. Trials for multinomial sampling led to similar results.

For the examples and the evaluation studies, the constant cell correction ($c = 0.5$) was adopted. This type of correction was supported among others by Anscombe [1956], Haldane [1956] and Gart [1962]. Alternatively, single cell correction could be used (Plackett [1962]), or the independence smoothed correction (cf. Agresti [1999]). A comparison of some classical asymptotic confidence intervals (with and without correction) are provided in Lui and Lin [2003] and Fagerland et al. [2015] or in the extensive correction study given in Chapter 2 of this work, which also includes different CI types as Score, LR or CR CIs.

In these studies, the focus lies on the power divergence family for known λ. Alternatively,

Table 3.10: Summary statistics for the type I error rate of the c-corrected, $c = 0.5$, ϕ-scaled odds ratio test for independence ($\alpha = 95\%$) taken over all scenarios $n_1 \leq n_2, (n_1, n_2) \in \{5, 10, ..., 50\}^2$ in the power divergence family (Example 1.10.4) with $\lambda = -1/3, 0, 1/3, 2/3, 1$. Step-width is $\pi_1 = 0.01, 0.02, ..., 0.99$, where the single box-plots are restricted to special value ranges $\pi_1 \in (0, 0.1], ..., (0.4, 0.5]$ and the full range $\pi_1 \in [0, 1]$.

Szenario	Power Divergence	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
$\pi_1 \in [0, 1]$	$\lambda = -1/3$	0.0000	0.0022	0.0149	0.0166	0.0277	0.0469
	$\lambda = 0$	0.0000	0.0134	0.0283	0.0263	0.0398	0.0569
	$\lambda = 1/3$	0.0000	0.0372	0.0468	0.0418	0.0499	0.0649
	$\lambda = 2/3$	0.0000	0.0520	0.0567	0.0554	0.0641	0.0922
	$\lambda = 1$	0.0000	0.0548	0.0627	0.0621	0.0739	0.1105
$\pi_1 \in [0, 0.1]$	$\lambda = -1/3$	0.0000	0.0000	0.0002	0.0030	0.0021	0.0391
	$\lambda = 0$	0.0000	0.0003	0.0033	0.0082	0.0127	0.0418
	$\lambda = 1/3$	0.0000	0.0097	0.0249	0.0244	0.0374	0.0630
	$\lambda = 2/3$	0.0000	0.0228	0.0428	0.0430	0.0657	0.0922
	$\lambda = 1$	0.0000	0.0214	0.0453	0.0494	0.0787	0.1051
$\pi_1 \in (0.1, 0.2]$	$\lambda = -1/3$	0.0000	0.0043	0.0115	0.0126	0.0196	0.0392
	$\lambda = 0$	0.0004	0.0166	0.0251	0.0233	0.0307	0.0414
	$\lambda = 1/3$	0.0113	0.0358	0.0442	0.0422	0.0490	0.0572
	$\lambda = 2/3$	0.0125	0.0538	0.0606	0.0578	0.0663	0.0873
	$\lambda = 1$	0.0130	0.0612	0.0696	0.0661	0.0783	0.1066
$\pi_1 \in (0.2, 0.3]$	$\lambda = -1/3$	0.0000	0.0129	0.0215	0.0199	0.0263	0.0393
	$\lambda = 0$	0.0053	0.0259	0.0313	0.0310	0.0391	0.0493
	$\lambda = 1/3$	0.0298	0.0436	0.0476	0.0457	0.0499	0.0581
	$\lambda = 2/3$	0.0319	0.0543	0.0573	0.0586	0.0631	0.0876
	$\lambda = 1$	0.0297	0.0592	0.0637	0.0656	0.0717	0.1014
$\pi_1 \in (0.3, 0.4]$	$\lambda = -1/3$	0.0000	0.0127	0.0247	0.0234	0.0355	0.0457
	$\lambda = 0$	0.0078	0.0266	0.0385	0.0341	0.0432	0.0521
	$\lambda = 1/3$	0.0328	0.0467	0.0488	0.0479	0.0506	0.0619
	$\lambda = 2/3$	0.0486	0.0537	0.0568	0.0595	0.0618	0.0888
	$\lambda = 1$	0.0510	0.0566	0.0614	0.0665	0.0711	0.1105
$\pi_1 \in (0.4, 0.5]$	$\lambda = -1/3$	0.0000	0.0093	0.0303	0.0247	0.0393	0.0469
	$\lambda = 0$	0.0024	0.0297	0.0405	0.0353	0.0443	0.0569
	$\lambda = 1/3$	0.0387	0.0468	0.0492	0.0494	0.0516	0.0649
	$\lambda = 2/3$	0.0426	0.0525	0.0561	0.0580	0.0619	0.0869
	$\lambda = 1$	0.0457	0.0551	0.0586	0.0627	0.0676	0.1094

one could consider to estimate λ from the data, as in Kateri and Agresti [2010]. However, for the 2×2 table and for given λ, the corresponding models (association or regression type) are saturated and thus cannot serve for the additional estimation of λ. This could be possible in case of additional explanatory variables. Beyond the power divergence, other ϕ-divergences can be used to build new measures of association. The divergence of Kateri et al. [2015] gives families of strictly convex scale functions with finite $\phi'(0), \phi''(0)$, and is thus compatible with sampling zeros. Evaluation results are given for this divergence (Section A.2). The ϕ-divergences that are based on non-strictly convex functions, like that of Balakrishnan and Sanghvi [1968], are not suitable for this set-up.

Note that some results of this chapter on ϕ-scaled odds ratios have been published in Espendiller and Kateri [2016].

The great flexibility of the ϕ-divergence induced scale change can be used to generalise other measures in categorical data analysis. As a brief outlook, the ϕ-scaled relative risk can been introduced, which is closely related to the ϕ-scaled odds ratio. Next to some structural properties, its asymptotic distribution is given for the product binomial and multinomial sampling scheme. Some properties, like boundedness, and further application for inference (CIs, hypothesis tests) as well as proofs have been excluded from the outlook.

3.7.1 Outlook: ϕ-scaled Relative Risk

The relative risk (1.31) for 2×2 table with underlying product binomial sampling scheme (cf. Section 1.8) can be expressed in log-scale as

$$\log \mathcal{R} = \log \frac{\pi_1}{\pi_2} = \log \frac{\pi_1/\pi}{\pi_2/\pi} = \log \pi_1/\pi - \log \pi_2/\pi = \phi'_{KL}(\pi_1/\pi) - \phi'_{KL}(\pi_2/\pi),$$

where $\pi = \frac{n_1\pi_1+n_2\pi_2}{n_1+n_2} = \frac{\pi_1+r\pi_2}{1+r}$ for $r = n_2/n_1$. Thus, the log-relative risk measures the divergence of the success probabilities to the hypothesis of equal success probabilities (or independence between explanatory and response variables) on ϕ_{KL}-scale in the Kullback-Leibler divergence (Example 1.10.2). A scale change can be conducted by replacing the Kullback-Leibler by the ϕ-divergence (1.68). For the product binomial sampling scheme, the generalised ϕ-scaled relative risk for success $\mathcal{R}^\phi_\mathcal{B}$ and failure $\mathcal{R}^\phi_{\mathcal{B},f}$ for a differentiable function $\phi \in \Phi$ are defined as

$$\mathcal{R}^\phi_\mathcal{B}(\pi_1, \pi_2, r) = \phi'\left(\frac{\pi_1}{\pi}\right) - \phi'\left(\frac{\pi_2}{\pi}\right) \quad \text{and} \quad \mathcal{R}^\phi_{\mathcal{B},f}(\pi_1, \pi_2, r) = \phi'\left(\frac{1-\pi_1}{1-\pi}\right) - \phi'\left(\frac{1-\pi_2}{1-\pi}\right). \tag{3.29}$$

A similar definition can be made for the multinomial sampling scheme. Let $\boldsymbol{\pi} = (\pi_{ij}) \in \Delta_{I,J}$ be a 2×2 multinomial probability vector and let $\mathbf{N} = (N_{ij}) \sim \mathcal{M}(n, \boldsymbol{\pi})$. Then the generalised ϕ-scaled relative risks for success $\mathcal{R}^\phi$ and failure $\mathcal{R}^\phi_f$ become

$$\mathcal{R}^\phi(\boldsymbol{\pi}) = \phi'\left(\frac{\pi_{11}}{\pi_{1+}\pi_{+1}}\right) - \phi'\left(\frac{\pi_{21}}{\pi_{2+}\pi_{+1}}\right) \quad \text{and} \quad \mathcal{R}^\phi_f(\boldsymbol{\pi}) = \phi'\left(\frac{\pi_{12}}{\pi_{1+}\pi_{+2}}\right) - \phi'\left(\frac{\pi_{22}}{\pi_{2+}\pi_{+2}}\right). \tag{3.30}$$

The structural properties given in the following Proposition are proven by following the techniques used to prove Proposition 3.1.3 for the ϕ-scaled odds ratio:

Proposition 3.7.1. Let $\phi \in \Phi$ be differentiable and strictly convex. Let $\mathbf{S} = [0, 1; 1, 0]$ the 2×2 matrix with zeros in the diagonal and ones in the off-diagonal. Let $\boldsymbol{\pi} = [\pi_{ij}]$ be a multinomial probability as 2×2 matrix and $\boldsymbol{\pi}^I = [\pi_{i+}\pi_{+j}]$ be the corresponding probability matrix of independence. The ϕ-scaled relative risk fulfils the following properties:

1. It holds $\mathcal{R}^\phi(\boldsymbol{\pi}) = 0$ if and only if $\boldsymbol{\pi} = \boldsymbol{\pi}^I$.
2. For $\boldsymbol{\pi} \in \Delta_{2\times2}$ and for $\epsilon > 0$ define

$$\boldsymbol{\pi}^\epsilon = \boldsymbol{\pi} + (\epsilon, 0, -\epsilon, 0), \qquad \boldsymbol{\pi}^{\epsilon,1} = \boldsymbol{\pi} + (\epsilon, -\epsilon, 0, 0), \qquad \boldsymbol{\pi}^{\epsilon,2} = \boldsymbol{\pi} + (0, 0, \epsilon, -\epsilon).$$

   It holds

$$\mathcal{R}^\phi(\boldsymbol{\pi}) > \mathcal{R}^\phi(\boldsymbol{\pi}^\epsilon), \qquad \mathcal{R}^\phi(\boldsymbol{\pi}) > \mathcal{R}^\phi(\boldsymbol{\pi}^{\epsilon,1}), \qquad \mathcal{R}^\phi(\boldsymbol{\pi}) > \mathcal{R}^\phi(\boldsymbol{\pi}^{\epsilon,2}).$$

3. $\mathcal{R}^\phi$ is asymmetric.
4. In general, $\mathcal{R}^\phi$ is not invariant under row or column scaling.
5. $\mathcal{R}^\phi$ is antisymmetric under row permutation, e.g. $\mathcal{R}^\phi(\mathbf{Sn}) = -\mathcal{R}^\phi(\mathbf{n})$.
6. The following relation holds for column permutation

$$\mathcal{R}^\phi(\boldsymbol{\pi}\mathbf{S}) = \mathcal{R}_f^\phi(\boldsymbol{\pi}).$$

7. It holds $\mathcal{R}^\phi(\mathbf{S}\boldsymbol{\pi}\mathbf{S}) = -\mathcal{R}^\phi(\boldsymbol{\pi}\mathbf{S}) = -\mathcal{R}_f^\phi(\boldsymbol{\pi})$.
8. $\mathcal{R}^\phi$ is not null invariant.
9. $\mathcal{R}^\phi$ is sampling scheme invariant.

The proof is not given here, since it resembles the proof of Proposition 3.1.3.

Next to some characteristics for table transformation, Proposition 3.7.1 1. fixes the measurement basis at $0 \in \mathbb{R}$ such that 2. can specify the direction and the effect size of the risk. Notice that each $\boldsymbol{\pi}^\epsilon, \boldsymbol{\pi}^{\epsilon,1}$ and $\boldsymbol{\pi}^{\epsilon,2}$ corresponds to an increase in risk.

The $\mathcal{R}^\phi$ is connected to the ϕ-scaled odds ratio θ^ϕ (3.1) by

$$\theta^\phi(\boldsymbol{\pi}\mathbf{S}) = \mathcal{R}^\phi(\boldsymbol{\pi}) - \mathcal{R}^\phi(\boldsymbol{\pi}\mathbf{S}) = \mathcal{R}^\phi(\boldsymbol{\pi}) - \mathcal{R}_f^\phi(\boldsymbol{\pi}) = -\theta^\phi(\boldsymbol{\pi}).$$

Therefore θ^ϕ is antisymmetric under row permutation. $\mathcal{R}^\phi$ has to be recalculated, when the success and failure variables are changed. Thus, the well-known property for the log-odds ratio and the log-relative risk is preserved.

The estimators are $\hat{\mathcal{R}}^\phi(\mathbf{N}) = \mathcal{R}^\phi(\hat{\boldsymbol{\pi}})$ and $\hat{\mathcal{R}}_f^\phi(\mathbf{N}) = \mathcal{R}_f^\phi(\hat{\boldsymbol{\pi}})$ for the multinomial sampling scheme and $\hat{\mathcal{R}}_{\mathcal{B}}^\phi(\mathbf{N}) = \mathcal{R}_{\mathcal{B}}^\phi(\hat{\pi}_1, \hat{\pi}_2, r)$ and $\hat{\mathcal{R}}_{\mathcal{B},f}^\phi(\mathbf{N}) = \mathcal{R}_{\mathcal{B},f}^\phi(\hat{\pi}_1, \hat{\pi}_2, r)$ for the binomial sampling scheme. The ϕ-relative risk is sampling scheme invariant. This is obvious when using $\pi_i = \pi_{i1}/\pi_{i+}, i = 1, 2$, such that $\hat{\mathcal{R}}^\phi(\mathbf{N}) = \hat{\mathcal{R}}_{\mathcal{B}}^\phi(\mathbf{N})$ and $\hat{\mathcal{R}}_f^\phi(\mathbf{N}) = \hat{\mathcal{R}}_{\mathcal{B},f}^\phi(\mathbf{N})$.

Remark 3.7.2. (Sampling Scheme for the Relative Risk)

Note that the relative risk is usually applied to data in the product binomial sampling scheme to compare the success probabilities. The ϕ-scaled relative risk is sampling scheme invariant by Proposition 3.7.1 9.. Thus, the analysis of $\mathcal{R}^\phi$'s theoretical properties can be done in the multinomial sampling scheme. One exception is the asymptotic behaviour as in the case of the ϕ-scaled odds ratio (cf. Section 3.2.1).

The ϕ-scaled relative risk is calculated with the R function `RR.phi` (see Appendix B.3).

Following the technique for the ϕ-scaled odds ratios (Theorem 3.2.1 and Theorem 3.2.2), the asymptotic distribution for the ϕ-relative risk is deduced.

Theorem 3.7.3. Let $\phi \in \Phi$ be two times differentiable and strictly convex. Let $n \in \mathbb{N}$ and $\boldsymbol{\pi} = (\pi_{ij}) \in \Delta_{2\times 2}$ be the parameters of a multinomial distribution $\mathbf{N} \sim \mathcal{M}(n, \boldsymbol{\pi})$ and let $\hat{\boldsymbol{\pi}} = \mathbf{N}/n$ be the MLE of $\boldsymbol{\pi}$. Then $\mathcal{R}^\phi(\hat{\boldsymbol{\pi}})$ is asymptotically normal distributed with mean $\mathcal{R}^\phi(\boldsymbol{\pi})$ and asymptotic variance

$$\sigma^2_{\mathcal{R}^\phi}(\boldsymbol{\pi}) = Aa_{11}^2 + Ba_{11}a_{21} + Ca_{21}^2 - (Da_{11} + Ea_{21})^2, \tag{3.31}$$

where

$$a_{i1} = \phi''\left(\frac{\pi_{i1}}{\pi_{i+}\pi_{+1}}\right), \qquad i = 1, 2,$$

and

$$\begin{aligned}
A &= [\pi_{11}(\pi_{1+}\pi_{+1} - \pi_{11}(\pi_{1+} + \pi_{+1}))^2 + \pi_{11}^2\pi_{12}\pi_{+1}^2 + \pi_{21}\pi_{11}^2\pi_{1+}^2](\pi_{1+}\pi_{+1})^{-4}\\
B &= [(\pi_{1+}\pi_{+1} - \pi_{11}(\pi_{1+} + \pi_{+1}))\pi_{21}\pi_{2+}\\
&\quad + \pi_{21}\pi_{11}\pi_{1+}(\pi_{2+}\pi_{+1} - \pi_{21}(\pi_{2+} + \pi_{+1}))](\pi_{1+}\pi_{+1}\pi_{2+}\pi_{+1})^{-2}\\
C &= [\pi_{21}(\pi_{2+}\pi_{+1} - \pi_{21}(\pi_{2+} + \pi_{+1}))^2 + \pi_{21}^2\pi_{11}\pi_{2+}^2 + \pi_{22}\pi_{21}^2\pi_{+1}^2](\pi_{2+}\pi_{+1})^{-4}\\
D &= [\pi_{11}\pi_{1+}\pi_{+1} - \pi_{11}^2\pi_{1+} - \pi_{11}^2\pi_{+1} - \pi_{11}\pi_{12}\pi_{+1} - \pi_{11}\pi_{21}\pi_{1+}](\pi_{1+}\pi_{+1})^{-2}\\
E &= [\pi_{11}\pi_{21}\pi_{2+} - \pi_{21}\pi_{2+}\pi_{+1} + \pi_{21}^2\pi_{2+} + \pi_{21}^2\pi_{+1} + \pi_{21}\pi_{22}\pi_{+1}](\pi_{1+}\pi_{+1})^{-2}.
\end{aligned}$$

Thus, it holds:

$$\sqrt{n}(\mathcal{R}^\phi(\hat{\boldsymbol{\pi}}) - \mathcal{R}^\phi(\boldsymbol{\pi})) \xrightarrow{d} \mathcal{N}(0, \sigma^2_{\mathcal{R}^\phi}(\boldsymbol{\pi})) \qquad \text{for} \qquad n \to \infty.$$

Proof. Let $\boldsymbol{\pi} = (\pi_{ij})$ be the 2×2 multinomial probability vector. By Lemma 1.4.2, the MLE $\hat{\boldsymbol{\pi}}$ of $\boldsymbol{\pi}$ obtains asymptotic normality

$$\sqrt{n}(\hat{\boldsymbol{\pi}} - \boldsymbol{\pi}) \xrightarrow{d} \mathcal{N}(0, \boldsymbol{\Sigma}(\boldsymbol{\pi})) \qquad \text{for } n \to \infty,$$

with $\boldsymbol{\Sigma}(\boldsymbol{\pi}) = \operatorname{diag}(\boldsymbol{\pi}) - \boldsymbol{\pi}^T\boldsymbol{\pi}$. Since $\phi \in \Phi$ is two times differentiable, the partial derivatives $\frac{\partial \mathcal{R}^\phi(\boldsymbol{\pi})}{\partial \pi_{ij}}$, $i, j = 1, 2$, exist:

$$\frac{\partial \mathcal{R}^\phi(\boldsymbol{\pi})}{\partial \pi_{11}} = \frac{\pi_{1+}\pi_{+1} - \pi_{11}(\pi_{1+} + \pi_{+1})}{(\pi_{1+}\pi_{+1})^2}\phi''\left(\frac{\pi_{11}}{\pi_{1+}\pi_{+1}}\right) + \frac{\pi_{21}\pi_{2+}}{(\pi_{2+}\pi_{+1})^2}\phi''\left(\frac{\pi_{21}}{\pi_{2+}\pi_{+1}}\right) \tag{3.32}$$

$$\frac{\partial \mathcal{R}^\phi(\boldsymbol{\pi})}{\partial \pi_{12}} = -\frac{\pi_{11}\pi_{+1}}{(\pi_{1+}\pi_{+1})^2}\phi''\left(\frac{\pi_{11}}{\pi_{1+}\pi_{+1}}\right) \tag{3.33}$$

$$\frac{\partial \mathcal{R}^\phi(\boldsymbol{\pi})}{\partial \pi_{21}} = -\frac{\pi_{11}\pi_{1+}}{(\pi_{1+}\pi_{+1})^2}\phi''\left(\frac{\pi_{11}}{\pi_{1+}\pi_{+1}}\right) - \frac{\pi_{2+}\pi_{+1} - \pi_{21}(\pi_{2+} + \pi_{+1})}{(\pi_{2+}\pi_{+1})^2}\phi''\left(\frac{\pi_{21}}{\pi_{2+}\pi_{+1}}\right) \tag{3.34}$$

$$\frac{\partial \mathcal{R}^\phi(\boldsymbol{\pi})}{\partial \pi_{22}} = \frac{\pi_{21}\pi_{+1}}{(\pi_{2+}\pi_{+1})^2}\phi''\left(\frac{\pi_{21}}{\pi_{2+}\pi_{+1}}\right) \tag{3.35}$$

Since ϕ is strictly convex, $\phi'' > 0$ and so $\nabla\mathcal{R}^\phi(\boldsymbol{\pi}) \neq 0$. The delta method (Theorem 1.4.3) gives asymptotic normality

$$\sqrt{n}(\mathcal{R}^\phi(\hat{\boldsymbol{\pi}}) - \mathcal{R}^\phi(\boldsymbol{\pi})) \xrightarrow{d} \mathcal{N}(0,$$

with

$$\sigma^2_{\mathcal{R}^\phi}(\boldsymbol{\pi}) = \nabla\mathcal{R}^\phi(\boldsymbol{\pi})^T\boldsymbol{\Sigma}(\boldsymbol{\pi})\nabla\mathcal{R}^\phi(\boldsymbol{\pi}) = \sum_{i,j=1}^{2} \pi_{ij}\left(\frac{\partial \mathcal{R}^\phi(\boldsymbol{\pi})}{\partial \pi_{ij}}\right)^2 - \left(\sum_{i,j=1}^{2} \pi_{ij}\left(\frac{\partial \mathcal{R}^\phi(\boldsymbol{\pi})}{\partial \pi_{ij}}\right)\right)^2.$$

Inserting the derivatives gives the result (3.31). $\square$

The asymptotic result in the binomial sampling scheme can be obtained following the techniques in Theorem 3.2.2:

Theorem 3.7.4. Let $\phi \in \Phi$ be two times differentiable and strictly convex. Let $n_i \in \mathbb{N}$ and $\pi_i \in (0,1)$ be the parameters of the product binomial sampling scheme $N_{i1} \sim \mathcal{B}(n_i, \pi_i), i = 1, 2$, and fix $r = \frac{n_1}{n_2}$. Then $\mathcal{R}^{\phi}_{\mathcal{B}}(\hat{\pi}_1, \hat{\pi}_2)$ is asymptotically normal distributed with mean $\mathcal{R}^{\phi}_{\mathcal{B}}(\pi_1, \pi_2)$ and asymptotic variance

$$\begin{aligned}\sigma^2_{\mathcal{R}^{\phi}_{\mathcal{B}}}(\pi_1, \pi_2; r) &= \left(\frac{\pi_2\pi_{2+}}{\pi^2}\phi''\left(\frac{\pi_1}{\pi}\right) + \frac{\pi_2\pi_{1+}}{\pi^2}\phi''\left(\frac{\pi_2}{\pi}\right)\right)^2 \pi_1(1-\pi_1) \\ &\quad + \left(\frac{\pi_1\pi_{2+}}{\pi^2}\phi''\left(\frac{\pi_1}{\pi}\right) + \frac{\pi_1\pi_{1+}}{\pi^2}\phi''\left(\frac{\pi_2}{\pi}\right)\right)^2 \frac{\pi_2(1-\pi_2)}{r}\end{aligned} \tag{3.36}$$

where

$$\pi = \frac{n_1\pi_1 + n_2\pi_2}{n_1 + n_2} = \frac{\pi_1 + r\pi_2}{1 + r} \qquad \text{and} \qquad 1 - \pi_{2+} = \pi_{1+} = \frac{n_1}{n_1 + n_2} = \frac{1}{1 + r^{-1}}.$$

Thus, it holds:

$$\sqrt{n}(\mathcal{R}^{\phi}_{\mathcal{B}}(\hat{\boldsymbol{\pi}}) - \mathcal{R}^{\phi}_{\mathcal{B}}(\boldsymbol{\pi})) \xrightarrow{d} \mathcal{N}(0, \sigma^2_{\mathcal{R}^{\phi}_{\mathcal{B}}}(\pi_1, \pi_2; r)) \quad \text{for} \quad n \to \infty.$$

Proof. For fixed $n_i, i = 1, 2$, the marginal probabilities $\pi_{i+} = \frac{n_i}{n_1+n_2}, i = 1, 2$, are also fixed, thus $\pi = \pi_1\pi_{1+} + \pi_2\pi_{2+}$. Hence,

$$\frac{\partial\mathcal{R}^{\phi}_{\mathcal{B}}(\pi_1, \pi_2, r)}{\partial\pi_1} = \frac{\pi - \pi_1\pi_{1+}}{\pi^2}\phi''\left(\frac{\pi_1}{\pi}\right) + \frac{\pi_2\pi_{1+}}{\pi}\phi''\left(\frac{\pi_2}{\pi^2}\right) \tag{3.37}$$

$$\frac{\partial\theta^{\phi}_{\mathcal{B}}(\pi_1, \pi_2, r)}{\partial\pi_2} = -\frac{\pi_1\pi_{2+}}{\pi^2}\phi''\left(\frac{\pi_1}{\pi}\right) - \frac{\pi - \pi_2\pi_{2+}}{\pi^2}\phi''\left(\frac{\pi_2}{\pi}\right). \tag{3.38}$$

Furthermore, $\frac{\partial\mathcal{R}^{\phi}_{\mathcal{B}}(\pi_1,\pi_2,r)}{\partial\pi_1} > 0$ and $\frac{\partial\mathcal{R}^{\phi}_{\mathcal{B}}(\pi_1,\pi_2,r)}{\partial\pi_2} < 0$ since $\phi'' > 0$ (ϕ is strictly convex). Based on the asymptotic normality of $(\hat{\pi}_1, \hat{\pi}_2)$ given in (3.10) and since $\nabla\mathcal{R}^{\phi}_{\mathcal{B}} \neq 0$, the asymptotic variance of $\mathcal{R}^{\phi}_{\mathcal{B}}(\hat{\pi}_1, \hat{\pi}_2, r)$ is derived by the delta method (Theorem 1.4.3) as

$$\begin{aligned}\sigma^2_{\mathcal{R}^{\phi}_{\mathcal{B}}}(\pi_1, \pi_2, r) &= \nabla\mathcal{R}^{\phi}_{\mathcal{B}}(\pi_1, \pi_2, r)^T \begin{pmatrix}\pi_1(1-\pi_1) & 0 \\ 0 & \frac{\pi_2(1-\pi_2)}{r}\end{pmatrix} \nabla\mathcal{R}^{\phi}_{\mathcal{B}}(\pi_1, \pi_2, r) \\ &= \left(\frac{\partial\mathcal{R}^{\phi}_{\mathcal{B}}(\pi_1, \pi_2, r)}{\partial\pi_1}\right)^2 \pi_1(1-\pi_1) + \left(\frac{\partial\mathcal{R}^{\phi}_{\mathcal{B}}(\pi_1, \pi_2, r)}{\partial\pi_2}\right)^2 \frac{\pi_2(1-\pi_2)}{r}.\end{aligned}$$

Using the identities

$$\begin{aligned}\pi - \pi_1\pi_{1+} &= \pi_1\pi_{1+} + \pi_2\pi_{2+} - \pi_1\pi_{1+} = \pi_2\pi_{2+} \\ \pi - \pi_2\pi_{2+} &= \pi_1\pi_{1+} + \pi_2\pi_{2+} - \pi_2\pi_{2+} = \pi_1\pi_{1+}\end{aligned}$$

in the derivatives $\frac{\partial\mathcal{R}^{\phi}_{\mathcal{B}}(\pi_1,\pi_2,r)}{\partial\pi_i}$ gives (3.36). □

The partial derivatives for the ϕ-scaled relative risk for the multinomial are calculated using the R function `RR.nabla`, while the partial derivatives (3.37) and (3.38) for the product binomial sampling scheme are calculated using `RR.nabla.binomial`. The corresponding standard error estimated are calculated using `RR.SE` and `RR.SE.binomial` (see Appendix B.3).

For the Kullback-Leibler divergence (Example 1.10.2), the ϕ_{KL}-scaled relative risk is reduced to the log-relative risk, $\mathcal{R}_{\mathcal{B}}^{\phi_{KL}}(\pi_1, \pi_2) = \log\left(\frac{\pi_1}{\pi_2}\right)$. Its standard error estimated becomes

$$\mathrm{SE}(\mathcal{R}_{\mathcal{B}}^{\phi_{KL}}(\hat{\pi}_1, \hat{\pi}_2)) = \sqrt{\frac{\sigma^2_{\mathcal{R}_{\mathcal{B}}^{\phi_{KL}}}(\hat{\pi}_1, \hat{\pi}_2, r)}{n_1}} = \sqrt{\frac{1-\hat{\pi}_1}{n_1\hat{\pi}_1} + \frac{1-\hat{\pi}_2}{n_1 r \hat{\pi}_2}} = \sqrt{\frac{1-\hat{\pi}_1}{n_1\hat{\pi}_1} + \frac{1-\hat{\pi}_2}{n_2\hat{\pi}_2}},$$

which is the known result for the log-relative risk (Agresti [2013, p. 71, Section 3.1.3]). More general, for the power divergence (Example 1.10.4, $\lambda \neq -1, 0$) it holds

$$\mathcal{R}_{\mathcal{B}}^{\phi_\lambda}(\pi_1, \pi_2) = \frac{1}{\lambda}\left(\left(\frac{\pi_1}{\pi}\right)^\lambda - \left(\frac{\pi_2}{\pi}\right)^\lambda\right) = \frac{1}{\lambda}\left(\frac{\pi_1^\lambda - \pi_2^\lambda}{\pi^\lambda}\right),$$

with

$$\sigma^2_{\mathcal{R}_{\mathcal{B}}^{\phi_\lambda}}(\pi_1, \pi_2, r) = \left(\pi_{2+}\left(\frac{\pi_1}{\pi}\right)^{\lambda-1} + \pi_{1+}\left(\frac{\pi_2}{\pi}\right)^{\lambda-1}\right)^2 \frac{\pi_1\pi_2}{\pi^4}\left(\pi_2(1-\pi_1) + \frac{\pi_1(1-\pi_2)}{r}\right),$$

where again $\pi = \frac{n_1\pi_1 + n_2\pi_2}{n_1+n_2}$.

Hypothesis tests, confidence intervals and other statistical properties are derived in the same way as for the ϕ-scaled odds ratio using the asymptotic behaviour and are therefore not given here. The two-sided 95% Wald confidence intervals for the ϕ-scaled relative risk are calculated using the R function `RR.CI.phi` and `RR.CI.phi.multi` for the product binomial and multinomial sampling scheme (see Appendix B.3).

Remark 3.7.5. In the case of Kullback-Leibler divergence, i.e. for the log-relative risk, the standard error estimated coincides for both sampling schemes, multinomial and binomial. This result does not hold in general. For example, consider the Minimum Discrimination Information ($\phi(x) = -\log x + x - 1$). For $\mathcal{R}^\phi$ and $\mathcal{R}_{\mathcal{B}}^\phi$ based on the data

$$\mathbf{n} = \begin{pmatrix} 9 & 41 \\ 20 & 29 \end{pmatrix},$$

it holds $\mathcal{R}^\phi(\hat{\boldsymbol{\pi}}) = 0.03980822 = \mathcal{R}_{\mathcal{B}}^\phi(\hat{\pi}_1, \hat{\pi}_2, r)$ with

$$\mathrm{SE}(\mathcal{R}^\phi(\hat{\boldsymbol{\pi}})) = 0.4686603 \neq 0.4673058 = \mathrm{SE}(\mathcal{R}_{\mathcal{B}}^\phi(\hat{\pi}_1, \hat{\pi}_2)).$$

So, the sampling scheme influences the asymptotic variance estimates, which is also the case for the ϕ-scaled odds ratio (cf. Section 3.2.1).

Chapter 4

Models and Model-based Measures

Association measures for categorical data, like Pearson's contingency coefficient Goodman and Kruskal's γ and Camér's V, are a simple, parsimonious and easy to communicate way to express the association structure of contingency tables by a scalar parameter. However, they reduce the information on association structures drastically. Recall that in an $I \times J$ table, there are $(I-1)(J-1)$ parameters required to describe the underlying association exactly (Remark 1.9.1).

The generalised odds ratios presented in Section 1.9 occur naturally as association parameters and are a fundamental part of models for contingency tables. When all $(I-1)(J-1)$ local odds ratio are known, the saturated model (1.46) is obtained. The connection between models and the generalised odds ratio has already been highlighted in Section 1.9.4. Association models (1.55) can be used to define models with less than $(I-1)(J-1)$ association parameters. For example, the uniform association model $\theta_{ij}^{L} = c, i = 1, ..., I-1, j = 1, ..., J-1$, reduces the number of association parameters to one. Other well known examples are the row-effect model (1.57) or the column-effect model (1.58), which reduce the number of parameters to $I-1$ or $J-1$, respectively. Therefore, association models can give a compromise between narrowness and simplicity of models.

The idea arises to use this increased flexibility of models to construct new measures of association. This approach has already been adopted by Clayton [1974], who used uniform global association models to define measures between two ordered categorical variables. In detail, for the global association model $\theta_{ij}^{G} = c, i = 1, ..., I-1, j = 1, ..., J-1$, he suggests a weighted mean estimator ($\hat{\theta}^{G,WM}$) and a Mantel-Haenszel type estimator ($\hat{\theta}^{G,MH}$) of the uniform global log-odds ratio. Measures of association following the idea of Yule's Q then are $(\hat{\theta}^{G,t} - 1)/(\hat{\theta}^{G,t} + 1), t = \text{WM, MH}$. Therefore, association model parameter estimates can be used to construct new measures of association for $I \times J$ contingency tables.

Beh and Farver [2009] defined three other closed-form estimators for the local uniform association parameter in linear-by-linear association models, including scores for the row and column variable. A non-model based approach to define generalised odds ratio for ordinal data can be found in Agresti [1980], which is similar to the measure proposed by Goodman and Kruskal [1954]. Connection between other categorical data models and categorical measures occur in other cases too, for example Fleiss and Cohen [1973] give a connection between ANOVA models and its intraclass correlation coefficient to Kappa-type measures in rater agreement.

Association models are a widely used and developed tool in association analysis of contingency tables, but require knowledge of modelling and computation effort in the fitting process. The fitting procedures are iterative based on the ideas of Newton-Raphson and can have disadvantages. Beh and Farver [2009, p. 350] mention:

> *"Iterative procedures, such as Newton's method, can suffer from many problems that arise because of the very nature of the algorithm - poor initial values, slow convergence and even lack of convergence. Although these may arise only rarely in practice, they can nonetheless occur."*

Therefore, the introduction of model-based measures that can be estimated in closed-form, next to their simplicity, can solve the problems occurring during the iterative procedures. On the other hand, researchers often use association measures as a first hand data analysis and model parameters, which are estimated in closed-form, can be used as such measures.

The aim of this chapter is to propose new measures of association generalising the results of Clayton [1974] for global odds ratios. His results are extended for the generalised odds ratios (Section 1.9). After an evaluation study, the generalised results are used to define easy to apply closed-form estimators for the uniform association and row-effect model parameters, which are compared to the estimators introduced in Beh and Farver [2009] and the standard MLE.

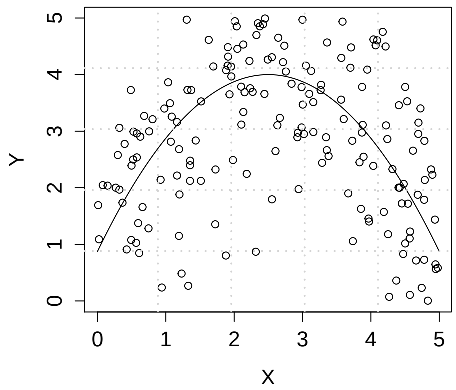
A motivating example is introduced in Section 4.1 which shows the negative effect of information loss when using association measures. Some data sets that will be used in the following, are given and studied in Section 4.2. Section 4.3 presents different estimation techniques for the uniform association model like the non-iterative estimators for the local linear-by-linear model by Beh and Farver [2009] and the weighted mean and Mantel-Haenszel type estimators by Clayton [1974] for the generalised odds ratios. To study the quality of the different estimators, a simulation study is presented in Section 4.4, comparing the closed-form estimators to MLE methods for uniform association. The results are then extended from the uniform association model to the row-effect model (R-model), giving closed-form weighted mean and Mantel-Haenszel type estimators for the row-effect association parameter in Section 4.5 together with theoretical results and an analysis of scores based on these closed-form parameter estimates. The quality of the newly introduced estimates is commented in Section 4.6 using examples and a simulation study. The new introduced estimates correspond to models, whose fit can be analysed via the non-iterative goodness-of-fit tests introduced in Section 4.7. The results of this chapter are summarized in Section 4.8.

4.1 Motivation

A simple example can be used to show the usefulness and problems of parameter reduction: Assume X describes the dose of a medicine for a disease with status Y, which are real valued and are later categorized. Assume the following distribution:

$$X \sim R[0,5] \qquad \text{and} \qquad Y = -\frac{1}{2}(X-2.5)^2 + 4 + \mathcal{N}(0, 1.25). \tag{4.1}$$

Thus, the health status Y is related to the dose X parabolically. This fictional example is comprehensible on the grounds of factual connection, as a low dose will not improve the health status, whereas a high dose will effect it in a negative way due to the side effects. Let the categories for X and Y be $[0,1],(1,2],(2,3],(3,4],(4,5]$. 175 points of the underlying model (4.1) have been simulated, where points with $Y \notin [0,5]$ were replaced. Figure 4.1 (left) shows the simulated points and the resulting table of the categorization is given in Figure 4.1 (right).



Dose (X)	Health Status (Y) 1	2	3	4	5
1	3	9	13	5	0
2	3	3	10	9	8
3	1	2	4	12	17
4	0	5	10	9	8
5	11	10	12	6	5

Figure 4.1: The 175 simulated pairs (X, Y) of the underlying model (4.1) and expected outcome $\mathbb{E}(Y|X)$ (solid line) (left) with contingency table resulting from the categorization of the 175 simulated pairs (right).

To show the problem caused by information loss, when describing association structures by association measures, the γ of Goodman and Kruskal [1954] is calculated. It is defined by $\gamma = \frac{n_c - n_d}{n_c + n_d}$, where n_c is the number of concordant pairs in the table, while n_d is the number of discordant pairs in the table (cf. Agresti [2013, Section 2.4.5, p. 57]). For the contingency table in Figure 4.1 (right) it has a 95% confidence interval of $[-0.202, 0.110]$ and thus indicates non-significant association in the data at the 5% level.

Dose (X)	Health Status (Y) 1/2	2/3	3/4	4/5
1/2	1.20	0.85	0.78	0.93
2/3	0.41	0.35	0.32	0.41
3/4	-0.42	-0.35	-0.32	-0.41
4/5	-1.19	-0.86	-0.77	-0.93

Dose (X)	Health Status (Y) 1/2	2/3	3/4	4/5
1/2	-1.10	0.84	0.85	∞
2/3	0.69	-0.51	1.20	0.47
3/4	∞	-0.00	-1.20	-0.47
4/5	$-\infty$	-0.51	-0.59	-0.06

Table 4.1: Estimated values of $\log \boldsymbol{\theta}^L$ based on 10^7 simulations of model (4.1) (left) and estimates $\log \hat{\boldsymbol{\theta}}^L$ based on the sample data of Figure 4.1 (right).

To calculate $\log \boldsymbol{\theta}^L$ of model (4.1), $K = 10^7$ random samples $(X_1, Y_1), ..., (X_K, Y_K)$ based on the model have been used to calculate the estimate $\hat{\boldsymbol{\pi}} = (\hat{\pi}_{ij}), \hat{\pi}_{ij} = n_{ij}/K$, where n_{ij} is the number of outcomes $(X, Y) = (i, j)$ of the discretized random sample, which is shown in Table 4.1 (left). Table 4.1 (right) shows the sample estimates of the categorized sample of Figure 4.1 (right). The local odds ratios in Table 4.1 (left) represent the parabolic structure of the model, giving ordered local odds ratios $\theta^L_{1j} > \theta^L_{2j} > \theta^L_{3j} > \theta^L_{4j}, j = 1, ..., 4$, where local odds ratios in the same row give quite similar values. This clearly indicates an association structure in the data not captured by Goodman and Kruskal's γ. Thus, unidimensional association measures are not able to cover more complex structures and can give misleading interpretations.

One way to cover more complex association in a parsimonious manner without using all $(I-1)(J-1)$ local odds ratios in this data set are association models. Since the odds ratios in every row have similar values, the R-model seems to be a good candidate to fit the data. The R model assumes $\log \theta^L_{ij} = c_{1i}, i = 1, ..., I-1, j = 1, ..., J-1$, which is a model using four parameters. The R-model for Figure 4.1 (right) has MLEs

$$\hat{c}_{11} = 0.5916211 \qquad \hat{c}_{12} = 0.6148828 \qquad \hat{c}_{13} = -0.5057322 \qquad \hat{c}_{14} = -0.7230531$$

and fits the data on a significance level of $\alpha = 5\%$ (p.val $= 0.189, G^2 = 16.05, df = 12$). While the association in the data cannot be measured using a single parameter, the structure can be covered by the R-model using four parameters, which can be regarded as R-type association measures.

4.2 Examples

Examples of data sets are given, for which different models are appropriate. These data will be used in the following study. The following data sets are used:

Table 4.2 cross-classifies the size of tonsils with carriers and noncarriers of Streptococcus pyogenes among children aged two weeks till 14 years from different social levels in three boroughs in Middlesex (UK) (cf. Clayton [1974]). Table 4.3 shows results of a survey among students from the University of Ioannina in Greece from 1995 (cf Marselos et al. [1997]). The students were asked for the frequency of alcohol and cannabis consumption on the basis of a four and three-level ordinal variable, respectively. Table 4.4 shows the responds of married couples in Arizona, who were asked how often sex is fun. Table 4.5 shows the categorized income cross-classified with the job satisfaction in the United States measured on an ordinal scale (cf. Agresti [2002, p. 57, Table 2.8]). Table 4.6 shows the responder's opinion on the amount of national spending for welfare cross-classified with their highest degree obtained. This data is based on the General Social Survey (GGS, U.S., 2008). Table 4.7 presents 8,310 teenage girls cross-classified by their educational level after 4 years of second-level education and their test for intellectual capacity (TIC) score in Holland (cf. Kateri [2014, p. 184]).

Table 4.8 shows the model fits of the independence (I) model (1.2), the uniform association (U) model ($\theta^h_{ij} = c, i = 1, ..., I-1, j = 1, ..., J-1$), the row-effect (R) model ($\theta^h_{ij} = c_{i1}, i = 1, ..., I-1, j = 1, ..., J-1$) and the column-effect (C) model ($\theta^h_{ij} = c_{1j}, i = 1, ..., I-1, j = 1, ..., J-1$) for the local ($h = L$) and global ($h = G$) odds ratios. These models can be defined using the HLP models introduced in Section 1.9.5. The model (1.59) becomes

$$\mathbf{C}^h \log(\mathbf{A}^h \boldsymbol{\pi}) = \mathbf{X}\boldsymbol{\beta}, \qquad h = L, G,$$

for an appropriate choice of the design matrix $\mathbf{X}$ and the matrices $\mathbf{A}^h$ and $\mathbf{C}^h$ from (1.43). The model fit can then be calculated using Lang's R function `mph.fit` based on Lang [2004, 2005], which is used in the R function `model.fits` (see Appendix B.4).

Table 4.2: Size of tonsils of carriers and noncarriers of Streptococcus pyogenes, from Holmes and Williams [1954] as cited by Clayton [1974].

Tonsils	not enlarged	enlarged	greatly enlarged
Carriers	24	29	19
Noncarriers	269	560	497

Table 4.3: Students' survey about cannabis use at the University of Ioannia, Greece (1995).

Alcohol consumption	I tried cannabis ... Never	Once or twice	More often
At mostonce/moth	204	6	1
Twice/moth	211	13	5
Twice/week	357	44	38
More often	92	34	49

Table 4.4: Responds of married couples in Arizona of their sexual fun rating. Source: Hout et al. [1987] cited by [Agresti, 2002, p. 65, Table 2.19].

	Wife's Rating of Sexual Fun			
Husband's Rating	Never or Occasionally	Fairly Often	Very Often	Almost Always
Never or occasionally	7	7	2	3
Fairly often	2	8	3	7
Very often	1	5	4	9
Almost always	2	8	9	14

Table 4.5: Cross-classification of Income and Job Satisfacation based on the 1996 General Social Survey, National Opinion Research Center.

	Job Satisfacation			
Income (dollars)	Very Dissatisfied	Little Dissatisfied	Moderately Satisfied	Very Satisfied
$< 15k$	1	3	10	6
$15k - 25k$	2	3	10	7
$25k - 40k$	1	6	14	12
$> 40k$	0	1	9	11

Table 4.6: Respondents' cross-classification by educational level and their opinion about national spending for walfare (GSS 2008)

	Highest degree obtained				
Welfare Spending	LT hight school	High school	Junior college	Bachelor	Graduate
too little	45	116	19	48	23
About right	40	167	33	68	41
Too much	47	185	34	63	26

Table 4.7: Cross-classification of 8,310 teenage girls in Holland by their educational levels after 4 years of second-level education, their test for intellectual capacity (TIC) score (c.f [Kateri, 2014, p.184]).

	Test for intellectual capacity (TIC)						
Education	1	2	3	4	5	6	7
DO	51	60	115	123	78	56	9
LBO	144	223	382	370	290	107	26
MAVO	60	134	288	424	442	266	72
MBO	75	167	320	458	428	258	72
HAVO	26	68	211	373	450	402	169
VWO	5	9	77	183	307	326	209

Education-level scale is dropped out (DO), junior level of education for professions (LBO), medium level of general education (MAVO), senior level of education for professions (MBO), high level of general education (HAVO) and general education preparing for university (VWO).

The model appropriate to describe the streptococcus data (Table 4.2) is the uniform association model, which for $2 \times J$ tables (as here) is equivalent to the row-effect model. Analogously, one can verify that the model adopted for Tables 4.3 (cannabis), 4.4 (sexual fun) and 4.5 (job satisfacation) is the uniform association model and for Table 4.6 (welfare spending) the appropriate model is the row-effect model. While for these tables, the models based on $\boldsymbol{\theta}^L$ and $\boldsymbol{\theta}^G$ hold, the appropriate model for Table 4.7 is the local row-

Table 4.8: Model fit (G^2, df and p-value) for selected data sets and for different association models: Independence (I), uniform association (U), R and C model based on the local and global log-odds ratio.

		Local			Global		
	I	U	R	C	U	R	C
	Streptococcus (2 × 3) - Table 4.2						
G^2	7.321	0.237	0.237	0	0.302	0.302	0
df	2	1	1	0	1	1	0
p-val	0.026	0.626	0.626	1	0.583	0.583	1
	Cannabis (4 × 3) - Table 4.3						
G^2	152.793	1.469	1.296	1.100	6.029	5.248	4.215
df	6	5	3	4	5	3	4
p-val	0	0.917	0.730	0.894	0.303	0.154	0.378
	Sexual Fun (4 × 4) - Table 4.4						
G^2	15.486	5.005	2.274	2.552	5.754	2.576	2.878
df	9	8	6	6	8	6	6
p-val	0.078	0.757	0.893	0.863	0.675	0.860	0.824
	Job Satisfication (4 × 4) - Table 4.5						
G^2	6.764	3.723	1.774	3.589	3.927	2.754	3.591
df	9	8	6	6	8	6	6
p-val	0.662	0.881	0.939	0.732	0.864	0.839	0.732
	Welfare Spending (3 × 5) - Table 4.6						
G^2	10.363	10.213	5.570	5.789	10.341	5.526	5.804
df	8	7	6	4	7	6	4
p-val	0.240	0.177	0.473	0.215	0.170	0.478	0.214
	Intellectual Capacity (Girls) (6 × 7) - Table 4.7						
G^2	1288.285	201.513	28.649	183.142	231.269	90.219	197.338
df	30	29	25	24	29	25	24
p-val	0	0	0.279	0	0	0	0

effect model, while no model based on the association model with global odds ratio fits the data. Thus, the data sets have been chosen to cover different association models.

4.3 Estimation of Uniform Association

The different types of closed-form estimators for the uniform association model parameter are presented: The non-iterative estimators by Beh and Davy [2004]; Beh and Farver [2009], the weighted mean and Mantel-Heanszel type estimators motivated by Clayton [1974].

4.3.1 Non-iterative Estimators

Let $\mathbf{N} \sim \mathcal{M}(n, \boldsymbol{\pi})$ be an multinomial random variable for an $I \times J$ contingency table with ordinal classification variables. Consider the linear-by-linear model (1.55) with known scores $\{\mu_i\}$ and $\{\nu_j\}$ for the rows and columns, respectively. Assume that the scores satisfy the sum-to-zero constraint (1.56) with marginal weights. These constraints can be used to get a closed-form estimator for φ. Multiplying (1.55) by $\pi_{i+}\mu_i$ and summing over $i = 1, ..., I$ gives

$$\sum_{i=1}^{I} \pi_{i+}\mu_i \log \pi_{ij} = \left(\sum_{i=1}^{I} \pi_{i+}\mu_i \lambda_i^X \right) + \varphi\nu_j. \tag{4.2}$$

Multiplying further (4.2) by $\pi_{+j}\nu_j$ and summing over $j = 1, ..., J$ gives finally

$$\varphi = \sum_{i,j=1}^{I,J} \pi_{i+}\pi_{+j}\mu_i\nu_j \log(\pi_{ij}). \tag{4.3}$$

Notice that (4.3) is equivalent to

$$\varphi = \sum_{i,j=1}^{I,J} \pi_{i+}\pi_{+j}\mu_i\nu_j \log \left(\frac{\pi_{ij}}{\pi_{+}\pi_{+j}} \right) \tag{4.4}$$

since

$$\begin{aligned}
&\sum_{i,j=1}^{I,J} \pi_{i+}\pi_{+j}\mu_i\nu_j [\log \pi_{ij} - \log \pi_{+} - \log \pi_{+j}] \\
&\quad = \sum_{i,j=1}^{I,J} \pi_{i+}\pi_{+j}\mu_i\nu_j \log \pi_{ij} - \left(\sum_{i=1}^{I} \pi_{i+}\mu_i \log \pi_{i+} \right) \underbrace{\left(\sum_{j=1}^{J} \pi_{+j}\nu_j \right)}_{=0} \\
&\qquad - \left(\sum_{j=1}^{J} \pi_{+j}\nu_j \log \pi_{+j} \right) \underbrace{\left(\sum_{i=1}^{I} \pi_{i+}\mu_i \right)}_{=0} \\
&\quad = \sum_{i,j=1}^{I,J} \pi_{i+}\pi_{+j}\mu_i\nu_j \log \pi_{ij}.
\end{aligned}$$

Identity (4.4) can be used to define a closed-form estimator for φ by

$$\hat{\varphi}_{\text{LogNI}} = \sum_{i,j=1}^{I,J} \hat{\pi}_{i+}\hat{\pi}_{+j}\mu_i\nu_j \log \left(\frac{\hat{\pi}_{ij}}{\hat{\pi}_{i+}\hat{\pi}_{+j}} \right), \tag{4.5}$$

where $\hat{\boldsymbol{\pi}} = (\hat{\pi}_{ij})$ is the MLE of $\boldsymbol{\pi}$. Estimator (4.5) has already been suggested and analysed by Beh and Davy [2004] and Beh and Farver [2009] and is called *non-iterative* (LogNI) *estimator*. Focus will lie on the uniform association model in the following and thus the normalized equidistant scores

$$\mu_i = \frac{i - \sum_{i=1}^{I} i\pi_{i+}}{\sqrt{\sum_{i=1}^{I} i^2\pi_{i+} - \left(\sum_{i=1}^{I} i\pi_{i+} \right)^2}}, \qquad \nu_j = \frac{j - \sum_{j=1}^{J} j\pi_{+j}}{\sqrt{\sum_{j=1}^{J} j^2\pi_{+j} - \left(\sum_{j=1}^{J} j\pi_{+j} \right)^2}},$$

for $i = 1, ..., I, j = 1, ..., J$ are used, which fulfil the sum-to-zero constraints (1.56) with marginal weights.

Since $\hat{\varphi}_{\text{LogNI}}$ is not compatible with sampling zeros, Beh and Davy [2004] suggest the use of the first order Taylor approximation

$$\log\left(\frac{\pi_{ij}}{\pi_{i+}\pi_{+j}}\right) \approx \frac{\pi_{ij}}{\pi_{i+}\pi_{+j}} - 1 \tag{4.6}$$

to get an alternative estimator:

$$\hat{\varphi}_{\text{BDNI}} = \sum_{i,j=1}^{I,J} \hat{\pi}_{i+}\hat{\pi}_{+j}\mu_i\nu_j \left(\frac{\hat{\pi}_{ij}}{\hat{\pi}_{i+}\hat{\pi}_{+j}} - 1\right) = \sum_{i,j=1}^{I,J} \hat{\pi}_{ij}\mu_i\nu_j. \tag{4.7}$$

Note that the estimator (4.7) can equivalently be introduced based on the correlation model (1.77) using similar techniques that result in the φ_{LogNI} estimator.

Since (4.6) is only suitable for $\frac{\pi_{ij}}{\pi_{i+}\pi_{+j}} \approx 1$, which is not the case, when the true model has a high divergence from independence, Beh and Farver [2009] also considered the approximation

$$\log\left(\frac{\pi_{ij}}{\pi_{i+}\pi_{+j}}\right) \approx 2\left(\frac{\pi_{ij} - \pi_{i+}\pi_{+j}}{\pi_{ij} + \pi_{i+}\pi_{+j}}\right)$$

based on

$$\log x = 2\left(\frac{x-1}{x+1} + \frac{1}{3}\left(\frac{x-1}{x+1}\right)^3 + \frac{1}{5}\left(\frac{x-1}{x+1}\right)^5 + ...\right),$$

leading to a *modified non-iterative* (MNI) estimator for φ:

$$\hat{\varphi}_{\text{MNI}} = 2\sum_{i,j=1}^{I,J} \hat{\pi}_{i+}\hat{\pi}_{+j}\mu_i\nu_j \left(\frac{\hat{\pi}_{ij} - \hat{\pi}_{i+}\hat{\pi}_{+j}}{\hat{\pi}_{ij} + \hat{\pi}_{i+}\hat{\pi}_{+j}}\right). \tag{4.8}$$

Since equidistant scores were used, $\hat{\varphi}_{\text{LogNI}}, \hat{\varphi}_{\text{BDNI}}$ and $\hat{\varphi}_{\text{MNI}}$ are estimators for the uniform association model $\theta^L_{ij} = \varphi, i = 1, ..., I-1, j = 1, ..., J-1$. These non-iterative estimators can be calculated using the R functions `logNI`, `BDNI` and `MNI` (see Appendix B.4).

4.3.2 Weighted Mean and Mantel-Haenszel

Another approach to get closed-form estimates of the uniform association parameter for ordinal classification variables goes back to Clayton [1974]. The uniform association model ($\theta^h_{ij} = \theta^h_0 = \exp(\varphi)$, $h = L, CO, C, G$) assumes the same association strength for every generalised odds ratio. Estimation of θ^h_0 is normally done by maximum likelihood estimation and requires numerical techniques (e.g. Newton-Raphson). Clayton [1974] defined two closed-form estimators, a weighted mean estimator, $\hat{\theta}^{h,WM}$, and a Mantel-Haenszel estimator, $\hat{\theta}^{h,MH}$, for the global uniform association model ($h = G$). Let $\mathbf{n} = (n_{ij})$ be the sample frequencies of an $I \times J$ table. Let $a^h_{ij}(\mathbf{n}), b^h_{ij}(\mathbf{n}), c^h_{ij}(\mathbf{n}), d^h_{ij}(\mathbf{n})$ be the essential sums defined in (1.41), and let $\boldsymbol{\omega} = (\omega_{ij})$ and $\tilde{\boldsymbol{\omega}} = (\tilde{\omega}_{ij})$ be weights such that $\sum_{i,j=1}^{I-1,J-1} \omega_{ij} = 1$. The closed-form estimators are defined as

$$\log \hat{\theta}^{h,WM} = \sum_{i,j=1}^{I-1,J-1} \omega_{ij} \log \left(\frac{a_{ij}^h(\mathbf{n}) d_{ij}^h(\mathbf{n})}{b_{ij}^h(\mathbf{n}) c_{ij}^h(\mathbf{n})} \right) = \boldsymbol{\omega}^T \log \hat{\boldsymbol{\theta}}^h \qquad \text{(Weighted Mean)}, \qquad (4.9)$$

$$\hat{\theta}^{h,MH} = \frac{\sum_{i,j=1}^{I-1,J-1} \tilde{\omega}_{ij} a_{ij}^h(\mathbf{n}) d_{ij}^h(\mathbf{n})}{\sum_{i,j=1}^{I-1,J-1} \tilde{\omega}_{ij} b_{ij}^h(\mathbf{n}) c_{ij}^h(\mathbf{n})} \qquad \text{(Mantel-Haenszel)}. \quad (4.10)$$

Notice that both estimators generalise from the known estimators for $2 \times 2 \times K$ tables. Woolf [1955] has already introduced weighted mean (WM) estimators, where the (non-normalized) weights are given by the inverse of the standard error estimated for the single log-odds ratios in the 2×2 tables ($\omega_{ij} = \left(\pi_{ij}^{-1} + \pi_{ij+1}^{-1} + \pi_{i+1j}^{-1} + \pi_{i+1j+1}^{-1}\right)^{-1}$). Cochran [1954] and Mantel and Haenszel [1959] introduced the Mantel-Haenszel (MH) estimator, also called Cochran-Mantel-Haenszel (CMH) estimator. The choice of weights can be arbitrary. The original MH estimator uses constant weights ($\tilde{\omega}_{ij} = 1$), whereas Clayton [1974] chose variance minimal weights. Clearly, $\tilde{\boldsymbol{\omega}}$ does not have to be normalized, since the normalisation constant cancels out in (4.10). Before discussing the choice of the weights, it is verified that both estimators give adequate estimates for the uniform association parameter θ_0^h:

Lemma 4.3.1. Let $\boldsymbol{\omega} = (\omega_{ij})$ and $\tilde{\boldsymbol{\omega}} = (\tilde{\omega}_{ij})$ be weights such that $\sum_{i,j=1}^{I-1,J-1} \omega_{ij} = 1$. The weighted mean estimator $\hat{\theta}^{h,WM}$ (4.9) and the Mantel-Haenszel estimator $\hat{\theta}^{h,MH}$ (4.10) are consistent under $\theta_{ij}^h = \theta_0^h$.

Proof. Since $\boldsymbol{\pi} \mapsto \theta^h(\boldsymbol{\pi})$ is a continuous function and $\hat{\boldsymbol{\pi}} \xrightarrow{P} \boldsymbol{\pi}$ for $n \to \infty$ (Lemma 1.4.1), it holds $\hat{\theta}_{ij}^h - \theta_{ij}^h \xrightarrow{P} 0$, $i = 1, ..., I-1, j = 1, ..., J-1$, such that under $\theta_{ij}^h = \theta_0^h$ it holds $\hat{\theta}_{ij} \xrightarrow{P} \theta_0^h$. One concludes

$$\log \hat{\theta}^{h,WM} = \sum_{i,j=1}^{I-1,J-1} \omega_{ij} \log \hat{\theta}_{ij}^h \xrightarrow{P} \sum_{i,j=1}^{I-1,J-1} \omega_{ij} \log \theta_0^h = \log \theta_0^h \qquad \text{for} \qquad n \to \infty$$

and the weighted mean estimator is consistent. For the Mantel-Haenszel estimator it holds for the observed essential sums (1.41), that $b_{ij}^h(\mathbf{n})/n \xrightarrow{P} b_{ij}^h(\boldsymbol{\pi})$ and $c_{ij}^h(\mathbf{n})/n \xrightarrow{P} c_{ij}^h(\boldsymbol{\pi})$ for $n \to \infty$ and all $i = 1, ..., I-1, j = 1, ..., J-1$. Since convergence in probability is compatible for products of random variables, it holds

$$n^{-2} \sum_{i,j=1}^{I-1,J-1} \tilde{\omega}_{ij} \hat{\theta}_{ij}^h b_{ij}^h(\mathbf{n}) c_{ij}^h(\mathbf{n}) \xrightarrow{P} \sum_{i,j=1}^{I-1,J-1} \tilde{\omega}_{ij} \theta_{ij}^h b_{ij}^h(\boldsymbol{\pi}) c_{ij}^h(\boldsymbol{\pi})$$

and

$$n^{-2} \sum_{i,j=1}^{I-1,J-1} \tilde{\omega}_{ij} b_{ij}^h(\mathbf{n}) c_{ij}^h(\mathbf{n}) \xrightarrow{P} \sum_{i,j=1}^{I-1,J-1} \tilde{\omega}_{ij} b_{ij}^h(\boldsymbol{\pi}) c_{ij}^h(\boldsymbol{\pi})$$

for $n \to \infty$. Thus under $\theta_{ij}^h = \theta_0^h$ it holds

$$\hat{\theta}^{h,MH} = \frac{n^{-2}}{n^{-2}} \frac{\sum_{i,j=1}^{I-1,J-1} \tilde{\omega}_{ij} \hat{\theta}_{ij}^h b_{ij}^h(\mathbf{n}) c_{ij}^h(\mathbf{n})}{\sum_{i,j=1}^{I-1,J-1} \tilde{\omega}_{ij} b_{ij}(\mathbf{n}) c_{ij}(\mathbf{n})}$$

$$\xrightarrow{P} \frac{\sum_{i,j=1}^{I-1,J-1} \tilde{\omega}_{ij}\theta_{ij}^h b_{ij}^h(\boldsymbol{\pi})c_{ij}^h(\boldsymbol{\pi})}{\sum_{i,j=1}^{I-1,J-1} \tilde{\omega}_{ij} b_{ij}^h(\boldsymbol{\pi})c_{ij}^h(\boldsymbol{\pi})} = \frac{\sum_{i,j=1}^{I-1,J-1} \tilde{\omega}_{ij}\theta_0^h b_{ij}^h(\boldsymbol{\pi})c_{ij}^h(\boldsymbol{\pi})}{\sum_{i,j=1}^{I-1,J-1} \tilde{\omega}_{ij} b_{ij}^h(\boldsymbol{\pi})c_{ij}^h(\boldsymbol{\pi})} = \theta_0^h$$

for $n \to \infty$ and the Mantel-Haenszel estimator is consistent. □

Since both estimators are consistent, the weights should be chosen to minimize the variance in order to increase the quality of the estimators. The optimal weights can easily be calculated:

Lemma 4.3.2. Let $\mathbf{X} = (X_1, ..., X_k)$ be a random variable with positive definite variance-covariance matrix $\boldsymbol{\Sigma} = (\sigma_{ij})$. Let $\boldsymbol{\omega} = (\omega_1, ..., \omega_k)$ be weights with $\sum_{i=1}^k \omega_i = 1$. The weights

$$\boldsymbol{\omega}^* = \frac{\boldsymbol{\Sigma}^{-1}\mathbf{1}}{\mathbf{1}^T\boldsymbol{\Sigma}^{-1}\mathbf{1}}$$

minimize the variance of $\boldsymbol{\omega}^T\mathbf{X}$, i.e. it holds $\mathbb{V}\text{ar}\,(\boldsymbol{\omega}^T\mathbf{X}) \geq \mathbb{V}\text{ar}\,((\boldsymbol{\omega}^*)^T\mathbf{X})$ for every weight $\boldsymbol{\omega} = (\omega_1, ..., \omega_k)$ with $\sum_{i=1}^k \omega_i = 1$.

Proof. The result can directly be obtained by applying the results of Harville [1998, Section 19.3, p. 464] for minimizing the quadratic form $\boldsymbol{\omega}^T\boldsymbol{\Sigma}\boldsymbol{\omega}$ under linear constraints. □

The asymptotic normality of $\log \hat{\boldsymbol{\theta}}^h$ has already been proven in Theorem 1.9.6. The asymptotic variance-covariance matrix is $\boldsymbol{\Sigma}^h(\boldsymbol{\pi})$ as given in (1.45), which can be calculated using the R function `log.genOR` from Appendix B.1. Together with the results of Lemma 4.3.1, the proof by Clayton [1974] for the optimal weights can be extended to generalised association models:

Corollary 4.3.3. Let $\mathbf{N} \sim \mathcal{M}(n, \boldsymbol{\pi})$ the multinomial random variable underlying an $I \times J$ contingency table $\mathbf{n} = (n_{ij})$. Let $\boldsymbol{\Sigma}^h(\boldsymbol{\pi})$ be the variance-covariance matrix of $\log \hat{\boldsymbol{\theta}}^h$ as given in (1.45). The variance-minimal weights of the weighted mean estimator $\log \hat{\theta}^{h,WM}$ (4.9) are given by

$$\boldsymbol{\omega} = \frac{(\boldsymbol{\Sigma}^h(\boldsymbol{\pi}))^{-1}\mathbf{1}}{\mathbf{1}^T(\boldsymbol{\Sigma}^h(\boldsymbol{\pi}))^{-1}\mathbf{1}}.$$

Lemma 4.3.4. Let $\boldsymbol{\omega} = (\omega_{ij})$ be the variance-minimal weights of the weighted mean estimator $\hat{\theta}^{h,WM}$ (Corollary 4.3.3). Then the variance-minimal weights $\tilde{\boldsymbol{\omega}} = (\tilde{\omega}_{ij})$ for the Mantel-Haenszel estimator $\hat{\theta}^{h,MH}$ are $\tilde{\omega}_{ij} = \omega_{ij}/(b_{ij}^h(\mathbf{n})c_{ij}^h(\mathbf{n}))$ for an $I \times J$ contingency table $\mathbf{n}$, where b_{ij}^h and c_{ij}^h are the essential sums defined in (1.41).

In this case, the asymptotic variances for $\hat{\theta}^{h,WM}$ and $\hat{\theta}^{h,MH}$ coincide.

Proof. Let $\mathbf{N} \sim \mathcal{M}(n, \boldsymbol{\pi})$ be multinomial distributed and let $\mathbf{n}$ be a realisation of $\mathbf{N}$. Let $\boldsymbol{\Sigma}^h(\boldsymbol{\pi})$ be the variance-covariance matrix of $\log \hat{\boldsymbol{\theta}}^h = \log \boldsymbol{\theta}^h(\hat{\boldsymbol{\pi}})$ as given in (1.45) and let $a_{ij}^h, b_{ij}^h, c_{ij}^h, d_{ij}^h$ be the essential sums defined in (1.41). By Theorem 1.9.6 it holds

$$\sqrt{n}(\log \hat{\boldsymbol{\theta}}^h - \log \boldsymbol{\theta}^h) \xrightarrow{d} \mathcal{N}(0, \boldsymbol{\Sigma}^h(\boldsymbol{\pi})), \qquad \text{for} \qquad n \to \infty.$$

Let $\mathbf{f}(\log \boldsymbol{\theta}^h) = (\exp \log \theta^h_{11}, ..., \exp \log \theta^h_{I-1J-1}) = (\theta^h_{11}, ..., \theta^h_{I-1J-1}) = \boldsymbol{\theta}^h$, such that

$$\mathbf{D} := \frac{\partial \mathbf{f}(\log \boldsymbol{\theta}^h)}{\partial \log \boldsymbol{\theta}^h} = \text{diag}(\exp(\log(\boldsymbol{\theta}^h))) = \text{diag}(\boldsymbol{\theta}^h) \neq \mathbf{0}.$$

Applying the delta method (Theorem 1.4.3) on $\mathbf{f}(\log \boldsymbol{\theta}^h)$, it holds

$$\sqrt{n}(\hat{\boldsymbol{\theta}}^h - \boldsymbol{\theta}^h) \xrightarrow{d} \mathcal{N}(0, \text{diag}(\boldsymbol{\theta}^h)\boldsymbol{\Sigma}^h(\boldsymbol{\pi})\text{diag}(\boldsymbol{\theta}^h)), \qquad \text{for} \qquad n \to \infty. \tag{4.11}$$

Let $\tilde{\boldsymbol{\alpha}}(\mathbf{N}) = (\tilde{\omega}_{ij} b^h_{ij}(\mathbf{N}) c^h_{ij}(\mathbf{N}))$ and write

$$\hat{\theta}^{h,MH} = \frac{\sum_{i,j=1}^{I-1,J-1} \tilde{\omega}_{ij} \hat{\theta}^h_{ij} b^h_{ij}(\mathbf{N}) c^h_{ij}(\mathbf{N})}{\sum_{i,j=1}^{I-1,J-1} \tilde{\omega}_{ij} b^h_{ij}(\mathbf{N}) c^h_{ij}(\mathbf{N})} = \frac{(\tilde{\boldsymbol{\alpha}}(\mathbf{N}))^T \hat{\boldsymbol{\theta}}^h}{(\tilde{\boldsymbol{\alpha}}(\mathbf{N}))^T \mathbf{1}} = \frac{(\tilde{\boldsymbol{\alpha}}(\mathbf{N}))^T \hat{\boldsymbol{\theta}}^h / n^2}{(\tilde{\boldsymbol{\alpha}}(\mathbf{N}))^T \mathbf{1} / n^2} \tag{4.12}$$

Remark that by (4.12) the estimator $\hat{\theta}^{h,MH}$ is invariant under weight multiplication $\tilde{\boldsymbol{\omega}} \mapsto k\tilde{\boldsymbol{\omega}}, k \neq 0$. It can be assumed without loss of generality that $(\tilde{\boldsymbol{\alpha}}(\mathbf{N}))^T \mathbf{1}/n^2 = 1$ by setting $k = ((\tilde{\boldsymbol{\alpha}}(\mathbf{N}))^T \mathbf{1})^{-1}$ and it holds for the asymptotic variance

$$\mathbb{V}\text{ar } (\hat{\theta}^{h,WM}) = \mathbb{V}\text{ar } ((\tilde{\boldsymbol{\alpha}}(\mathbf{N}))^T \hat{\boldsymbol{\theta}}^h / n^2) = \mathbb{V}\text{ar } ((\tilde{\boldsymbol{\alpha}}(\hat{\boldsymbol{\pi}}))^T \hat{\boldsymbol{\theta}}^h).$$

Let $\tilde{\boldsymbol{\alpha}}(\boldsymbol{\pi}) = (\tilde{\omega}_{ij} b^h_{ij}(\boldsymbol{\pi}) c^h_{ij}(\boldsymbol{\pi}))$. Since $\tilde{\boldsymbol{\alpha}}$ is a continuous function of $\boldsymbol{\pi}$ and $\hat{\boldsymbol{\pi}} \xrightarrow{P} \boldsymbol{\pi}$, it holds $n^{-2}\tilde{\boldsymbol{\alpha}}(\mathbf{N}) = \tilde{\boldsymbol{\alpha}}(\hat{\boldsymbol{\pi}}) \xrightarrow{P} \tilde{\boldsymbol{\alpha}}(\boldsymbol{\pi})$, such that with (4.11) and by Slutsky's theorem it holds

$$\sqrt{n}(\tilde{\boldsymbol{\alpha}}(\hat{\boldsymbol{\pi}}))^T (\hat{\boldsymbol{\theta}}^h - \boldsymbol{\theta}^h) \xrightarrow{d} \mathcal{N}(0, (\tilde{\boldsymbol{\alpha}}(\boldsymbol{\pi}))^T \text{diag}(\boldsymbol{\theta}^h)\boldsymbol{\Sigma}^h(\boldsymbol{\pi})\text{diag}(\boldsymbol{\theta}^h)\tilde{\boldsymbol{\alpha}}(\boldsymbol{\pi})), \qquad \text{for} \qquad n \to \infty.$$

Thus the asymptotic variance is:

$$\mathbb{V}\text{ar } ((\tilde{\boldsymbol{\alpha}}(\hat{\boldsymbol{\pi}}))^T \hat{\boldsymbol{\theta}}^h) = (\tilde{\boldsymbol{\alpha}}(\boldsymbol{\pi}))^T \text{diag}(\boldsymbol{\theta}^h)\boldsymbol{\Sigma}^h(\boldsymbol{\pi})\text{diag}(\boldsymbol{\theta}^h)\tilde{\boldsymbol{\alpha}}(\boldsymbol{\pi}),$$

which gives the asymptotic variance

$$\mathbb{V}\text{ar } \left(\log\left((\tilde{\boldsymbol{\alpha}}(\hat{\boldsymbol{\pi}}))^T \hat{\boldsymbol{\theta}}^h\right)\right) = \frac{(\tilde{\boldsymbol{\alpha}}(\boldsymbol{\pi}))^T \text{diag}(\boldsymbol{\theta}^h)}{(\tilde{\boldsymbol{\alpha}}(\boldsymbol{\pi}))^T \boldsymbol{\theta}^h} \boldsymbol{\Sigma}^h(\boldsymbol{\pi}) \frac{\text{diag}(\boldsymbol{\theta}^h)\tilde{\boldsymbol{\alpha}}(\boldsymbol{\pi})}{(\tilde{\boldsymbol{\alpha}}(\boldsymbol{\pi}))^T \boldsymbol{\theta}^h}, \tag{4.13}$$

by applying the delta method (Theorem 1.4.3) on $f(x) = \log(x)$.

Let $\boldsymbol{\omega} = (\omega_{ij})$ be the variance-minimal weights of the weighted mean estimator, i.e. the weights that minimize $\mathbb{V}\text{ar } (\boldsymbol{\omega}^T \log \hat{\boldsymbol{\theta}}^h) = \boldsymbol{\omega}^T \boldsymbol{\Sigma}^h(\boldsymbol{\pi})\boldsymbol{\omega}$. Set $k(\tilde{\boldsymbol{\omega}}) = (\tilde{\boldsymbol{\alpha}}(\boldsymbol{\pi}))^T \boldsymbol{\theta}^h$, then (4.13) is minimal, if

$$\begin{aligned}
\frac{(\tilde{\boldsymbol{\alpha}}(\boldsymbol{\pi}))^T \text{diag}(\boldsymbol{\theta}^h)}{k(\tilde{\boldsymbol{\omega}})} = \boldsymbol{\omega} \quad &\Leftrightarrow \quad \tilde{\omega}_{ij} b^h_{ij}(\boldsymbol{\pi}) c^h_{ij}(\boldsymbol{\pi}) \theta^h_{ij} = k(\tilde{\boldsymbol{\omega}})\omega_{ij}, \quad i = 1, ..., I-1, j = 1, ..., J-1 \\
&\Leftrightarrow \quad \tilde{\omega}_{ij} a^h_{ij}(\boldsymbol{\pi}) d^h_{ij}(\boldsymbol{\pi}) = k(\tilde{\boldsymbol{\omega}})\omega_{ij}, \quad i = 1, ..., I-1, j = 1, ..., J-1 \\
&\Leftrightarrow \quad \tilde{\omega}_{ij} = \frac{k(\tilde{\boldsymbol{\omega}})\omega_{ij}}{a^h_{ij}(\boldsymbol{\pi}) d^h_{ij}(\boldsymbol{\pi})}, \quad i = 1, ..., I-1, j = 1, ..., J-1,
\end{aligned}$$

which can be estimated by $\tilde{\omega}_{ij} = \frac{k(\tilde{\boldsymbol{\omega}}) n^2 \omega_{ij}}{a^h_{ij}(\mathbf{n}) d^h_{ij}(\mathbf{n})}$. Since by (4.12) the estimator is unaffected when the weights are multiplied by $k(\tilde{\boldsymbol{\omega}})n^2$, an appropriate choice for the weights is $\tilde{\omega}_{ij} = \frac{\omega_{ij}}{a^h_{ij}(\mathbf{n}) d^h_{ij}(\mathbf{n})}$. □

Remark 4.3.5. When applying Corollary 4.3.3 or Lemma 4.3.4 to calculate the optimal weights for the weighted mean resp. Mantel-Haenszel estimator, the variance-covariance matrix $\mathbf{\Sigma}^h(\boldsymbol{\pi})$ is unknown and has to be replaced by an estimate. Two commonly used estimates are the sample estimate $\mathbf{\Sigma}^h(\hat{\boldsymbol{\pi}})$ and the estimate under independence $\mathbf{\Sigma}^h(\hat{\boldsymbol{\pi}}^I)$. When the sample estimate is used the weights are denoted by $\boldsymbol{\omega} = \boldsymbol{\omega}(S)$ and by $\boldsymbol{\omega} = \boldsymbol{\omega}(I)$ when the variance-covariance matrix is estimated using the independence probability. There are clearly other estimates, such as the uniform association model estimate or any other model estimate. However, calculation of these estimates requires iterative techniques contradicting thus the aim of the chapter to get closed-form estimates avoiding complex fitting methods. In the cases S and I, the weights are available in closed form, but complex. Clayton [1974] gives closed formulas for the weights $\boldsymbol{\omega}(I)$ and $\tilde{\boldsymbol{\omega}}(I)$ in the case of global log-odds ratios. Another example is the following:

4.3.3 Example

The variance-covariance of $\log \hat{\boldsymbol{\theta}}^L$ has already been calculated as given in Remark 1.9.7. For a 3×3 table, the variance-covariance becomes

$$\mathbf{\Sigma}^L(\boldsymbol{\pi}) =$$

$$\begin{pmatrix} \frac{1}{\pi_{11}}+\frac{1}{\pi_{12}}+\frac{1}{\pi_{21}}+\frac{1}{\pi_{22}} & -\frac{1}{\pi_{11}}-\frac{1}{\pi_{22}} & -\frac{1}{\pi_{21}}-\frac{1}{\pi_{22}} & \frac{1}{\pi_{22}} \\ -\frac{1}{\pi_{12}}-\frac{1}{\pi_{22}} & \frac{1}{\pi_{12}}+\frac{1}{\pi_{13}}+\frac{1}{\pi_{22}}+\frac{1}{\pi_{23}} & \frac{1}{\pi_{22}} & -\frac{1}{\pi_{22}}-\frac{1}{\pi_{23}} \\ -\frac{1}{\pi_{21}}-\frac{1}{\pi_{22}} & \frac{1}{\pi_{22}} & \frac{1}{\pi_{21}}+\frac{1}{\pi_{22}}+\frac{1}{\pi_{31}}+\frac{1}{\pi_{32}} & -\frac{1}{\pi_{22}}-\frac{1}{\pi_{32}} \\ \frac{1}{\pi_{22}} & -\frac{1}{\pi_{21}}-\frac{1}{\pi_{23}} & -\frac{1}{\pi_{22}}-\frac{1}{\pi_{32}} & \frac{1}{\pi_{22}}+\frac{1}{\pi_{23}}+\frac{1}{\pi_{32}}+\frac{1}{\pi_{33}} \end{pmatrix},$$

and for $\boldsymbol{\pi} = \boldsymbol{\pi}^I = (\pi_{i+}\pi_{+j})$, the probability table under independence, it holds:

$$\left(\mathbf{\Sigma}^L(\boldsymbol{\pi}^I)\right)^{-1}\mathbf{1} = \begin{pmatrix} \pi_{1+}\pi_{+1}(\pi_{2+}\pi_{+2}+2\pi_{2+}\pi_{+3}+2\pi_{3+}\pi_{+2}+4\pi_{3+}\pi_{+3}) \\ \pi_{1+}\pi_{+3}(2\pi_{2+}\pi_{+1}+\pi_{2+}\pi_{+2}+4\pi_{3+}\pi_{+1}+2\pi_{3+}\pi_{+2}) \\ \pi_{3+}\pi_{+1}(2\pi_{1+}\pi_{+2}+4\pi_{1+}\pi_{+3}+\pi_{2+}\pi_{+2}+2\pi_{2+}\pi_{+3}) \\ \pi_{3+}\pi_{+3}(4\pi_{1+}\pi_{+1}+2\pi_{1+}\pi_{+2}+2\pi_{2+}\pi_{+1}+\pi_{2+}\pi_{+2}) \end{pmatrix}.$$

The normalized weights can then be calculated by

$$\boldsymbol{\omega} = \left(\mathbf{\Sigma}^L(\boldsymbol{\pi}^I)\right)^{-1}\mathbf{1}/(\mathbf{1}^T\left(\mathbf{\Sigma}^L(\boldsymbol{\pi}^I)\right)^{-1}\mathbf{1}).$$

This simple example shows that calculation of the optimal weights just requires inverting a $k \times k$ matrix ($k = (I-1)(J-1)$). Thus the WM and MH estimators are derived by simple matrix operations. Due to the complexity of the variance-covariance matrices $\mathbf{\Sigma}^h(\boldsymbol{\pi})$, the explicit formulas for the various types of generalised odds ratios are not provided here.

Lemma 4.3.6. Let $\boldsymbol{\omega}$ ($\tilde{\boldsymbol{\omega}}$) be the variance-minimal weights for the WM (MH) estimator and let $\mathbf{\Sigma}^h(\boldsymbol{\pi})$ be the asymptotic variance-covariance matrix of $\log \hat{\boldsymbol{\theta}}^h$. Then the asymptotic variance for both estimators becomes

$$\mathbb{V}\text{ar }(\log \hat{\theta}^{h,t}) = \left(\mathbf{1}^T\left(\mathbf{\Sigma}^h(\boldsymbol{\pi})\right)^{-1}\mathbf{1}\right)^{-1}, \qquad t = WM, MH. \tag{4.14}$$

Proof. Using

$$\boldsymbol{\omega} = \frac{\left(\boldsymbol{\Sigma}^h(\boldsymbol{\pi})\right)^{-1}\mathbf{1}}{\mathbf{1}^T\left(\boldsymbol{\Sigma}^h(\boldsymbol{\pi})\right)^{-1}\mathbf{1}}$$

from Corollary 4.3.3, and using the short notation $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}^h(\boldsymbol{\pi})$, it follows

$$\mathbb{V}\text{ar}\ (\log\hat{\theta}^{h,WM}) = \mathbb{V}\text{ar}\ (\boldsymbol{\omega}\log\hat{\boldsymbol{\theta}}^h) = \boldsymbol{\omega}^T\boldsymbol{\Sigma}\boldsymbol{\omega} = \frac{1}{(\mathbf{1}^T\boldsymbol{\Sigma}^{-1}\mathbf{1})^2}\mathbf{1}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{-1}\mathbf{1} = \left(\mathbf{1}^T\boldsymbol{\Sigma}^{-1}\mathbf{1}\right)^{-1}.$$

By Lemma 4.3.4 the WM and MH have the same asymptotic variance under variance-minimal weights. □

The function `WM.calc.log` and `MH.calc.log` (see Appendix B.4) can be used for the calculation of the weighted mean and Mantel-Haenszel estimators and their standard error estimated.

Remark 4.3.7. Estimators WM (S) and MH (S) are not defined for combinations of sampling zeros leading to essential sums (1.41) $a_{ij}^h(\mathbf{n}), b_{ij}^h(\mathbf{n}), c_{ij}^h(\mathbf{n}), d_{ij}^h(\mathbf{n})$ equal to zero, as the estimated variance-covariance $\boldsymbol{\Sigma}^h(\hat{\boldsymbol{\pi}})$ for calculating the weights is not defined in this case. The weights calculated based on the independence estimate, i.e. $\boldsymbol{\Sigma}^h(\boldsymbol{\pi}^I)$, are always defined, but WM (I) can give infinite estimates, if $\log\hat{\theta}_{ij}^h = \pm\infty$ occurs, whereas it is undefined, if one $\log\hat{\theta}_{ij}^h = +\infty$ and one $\log\hat{\theta}_{i'j'}^h = -\infty$, $i \neq i'$ or $j \neq j'$. The estimator MH (I) does not have these problems. Notice that the weights $\boldsymbol{\omega}$ and $\tilde{\boldsymbol{\omega}}$ are not naturally restricted to the interval $[0,1]$, enabling negative weights, which sometimes leads to negative MH estimators ($\theta^{h,MH} < 0$), dissenting the very nature of the odds ratio which has values in $[0,\infty]$.

Table 4.9 shows the presented closed-form estimators for the local and global uniform association parameter on log-scale for different data sets (Streptotoccus, Table 4.2; Cannabis, Table 4.3; Sexual Fun, Table 4.4; Job Satisfication, Table 4.5; Welfare Spending, Table 4.6; Intellectual Capacity Girls, Table 4.7). MLEs are calculated using Lang's *mph.fit*. Variance-covariance matrix cannot be calculated for the Job Satisfication data due to sampling zeros and sample based estimators for WH (S) and MH (S) do not exist. In addition it shows the estimated sampling zero probability for the corresponding data set under MLE model fit for the saturated and independence model and the local and global uniform association model. Estimated sampling zero probability can be estimated using the R function `zero.cell.prob`, MLEs can be calculated using `get.MLE.log` (see Appendix B.4).

The closed-form estimators are in most cases able to give adequate model parameter estimates close to the MLEs. BDNI and MNI seem to underestimate the local uniform association parameter for the cannabis data. In addition, the same holds for the Mantel-Haenszel type estimators for the welfare data and the sexual fun data for both, local and global, uniform association. Since the job satisfaction data has a sampling zero, not all estimator are defined and the weighted mean estimator gives an estimate of ∞.

	Strepto.	Cannabis	Sexual Fun	Job Satisf.	Welfare Spending	IC Girls
	n=1398	n=1054	n=91	n=96	n=955	n=8310
	2×3	4×3	4×4	4×4	3×5	6×7
Model	Estimated sampling zero probability					
Saturated	0.00%	37.18%	63.70%	100.00%	0.00%	0.73%
Independence	0.00%	0.00%	27.44%	87.98%	0.00%	0.00%
Local	0.00%	42.78%	38.38%	91.83%	0.00%	0.29%
Global	0.00%	2.99%	36.15%	89.34%	0.00%	0.00%
	Local uniform association (U^L)					
MLE	0.4286	0.8026	0.2885	0.2138	-0.0131	0.1786
WM (S)	0.4283	0.8029	0.2888	-	-0.0086	0.1718
WM (I)	0.4091	0.7817	0.2705	∞	-0.0179	0.1903
MH (S)	0.4211	0.7781	0.1903	-	-0.0321	0.1429
MH (I)	0.4021	0.7697	0.1808	0.1648	-0.0407	0.1562
logNI	0.4091	0.7817	0.2705	∞	-0.0179	0.1903
BDNI	0.4363	0.5732	0.2680	0.2092	-0.0132	0.1586
MNI	0.4038	0.6051	0.2580	0.2583	-0.0179	0.1687
	Global uniform association (U^G)					
MLE	0.6026	1.8622	1.0928	0.5599	-0.0165	1.0960
WM (S)	0.6022	1.8692	1.0667	-	-0.0152	1.1031
WM (I)	0.5726	1.9495	1.0658	∞	-0.0212	1.1642
MH (S)	0.5989	1.8713	0.9101	-	-0.0381	1.0880
MH (I)	0.5696	1.9242	0.9760	0.5899	-0.0430	1.1315

Table 4.9: MLEs and estimates for the different closed-form estimators of the local and global uniform association parameter on log-scale. The estimation of the probability of occurrence of a sampling zero under the saturated and independence model as well as under the estimated local and global uniform association model is based on $K = 100,000$ simulations.

4.3.4 Non-Iterative Confidence Intervals

Confidence intervals for the weighted mean and Mantel-Haenszel estimators under variance-optimal weight choice can be constructed using the formula for the asymptotic variance in Lemma 4.3.6. Using the corresponding variance estimate by inserting $\hat{\boldsymbol{\pi}}$ (for the independence or saturated model), the standard error estimated becomes:

$$SE(\log \hat{\theta}^{h,t}) = \sqrt{\left(\mathbf{1}^T \left(\boldsymbol{\Sigma}^h(\hat{\boldsymbol{\pi}})\right)^{-1} \mathbf{1}\right)^{-1} /n}, \qquad t = WM, MH.$$

The $(1-\alpha)\%$ Wald confidence interval then becomes

$$\left[\hat{\theta}^{h,t} - z_{1-\alpha/2} SE(\log \hat{\theta}^{h,t}), \hat{\theta}^{h,t} + z_{1-\alpha/2} SE(\log \hat{\theta}^{h,t})\right], \qquad t = WM, MH,$$

where $z_{1-\alpha/2}$ is the $(1-\alpha/2)$-quantile of a standard normal distribution.

Table 4.10 shows the 95% confidence intervals for the WM and MH estimator for the

	Strepto.	Cannabis	Sexual Fun	Job Satisf.	Welfare Spending	IC Girls
	n=1398	n=1054	n=91	n=96	n=955	n=8310
	2×3	4×3	4×4	4×4	3×5	6×7
	Local uniform association					
MLE	[0.1122, 0.7450]	[0.6492, 0.9561]	[0.1033, 0.4737]	[-0.0315, 0.4591]	[-0.0797, 0.0535]	[0.1671, 0.1902]
WM (S)	[0.1189, 0.7377]	[0.6524, 0.9534]	[0.1036, 0.4739]	-	[-0.0761, 0.0589]	[0.1600, 0.1835]
WM (I)	[0.0903, 0.7280]	[0.6818, 0.8816]	[0.1055, 0.4356]	-	[-0.0846, 0.0487]	[0.1806, 0.2000]
MH (S)	[0.1117, 0.7305]	[0.6276, 0.9286]	[0.0051, 0.3754]	-	[-0.0996, 0.0354]	[0.1311, 0.1547]
MH (I)	[0.0832, 0.7210]	[0.6699, 0.8696]	[0.0157, 0.3458]	[-0.0714, 0.4010]	[-0.1073, 0.0259]	[0.1465, 0.1659]
	Global uniform association					
MLE	[0.1615, 1.0438]	[1.5505, 2.1738]	[0.4210, 1.7645]	[-0.1142, 1.2340]	[-0.2335, 0.2006]	[1.0306, 1.1614]
WM (S)	[0.1630, 1.0413]	[1.5613, 2.1770]	[0.3964, 1.7369]	-	[-0.2326, 0.2021]	[1.0389, 1.1672]
WM (I)	[0.1311, 1.0141]	[1.6650, 2.2341]	[0.3946, 1.7370]	-	[-0.2382, 0.1959]	[1.0973, 1.2311]
MH (S)	[0.1598, 1.0381]	[1.5634, 2.1791]	[0.2399, 1.5803]	-	[-0.2555, 0.1792]	[1.0238, 1.1521]
MH (I)	[0.1281, 1.0111]	[1.6397, 2.2088]	[0.3048, 1.6472]	[-0.0840, 1.2637]	[-0.2601, 0.1740]	[1.0646, 1.1985]

Table 4.10: 95% confidence intervals (log-scale) for the different closed-form estimators of the local and global uniform association parameter on log-scale. Compare to Table 4.9.

uniform association model. It turns out, that the intervals are of similar length and position for all analysed data sets.

4.4 Simulation Study

To compare the quality of the different estimators for the uniform association parameter (MLE, WM (I,S), MH (I,S), LogNI, BDNI, MNI), a simulation study has been conducted based on the data sets used in Table 4.9. Since some estimators have problems with sampling zeros, the simulation was done (1) conditional on 'no sampling zeros' and (2) using the single cell correction with $c = 0.5$ to enable comparability.

The MLEs and marginal distributions of the data were used to define a probability model $\mathbf{N} \sim \mathcal{M}(n, \boldsymbol{\pi})$ underlying the simulation study, where n is the sample size of the data and $\boldsymbol{\pi} \in \Delta_{I,J}$ is $\hat{\boldsymbol{\pi}}$ (saturated) or such that $\theta_{ij}^h(\boldsymbol{\pi}) = \hat{\theta}_0^h, h = L, G$ (uniform association) with the same marginal distributions as in the data. Then K samples $\mathbf{n}_1, ..., \mathbf{n}_K$ have been generated to calculate the estimates $\hat{\theta}_1^h = \hat{\theta}^h(\mathbf{n}_1), ..., \hat{\theta}_K^h = \hat{\theta}^h(\mathbf{n}_K)$ for the different estimation types. The summary statistics and mean square error $K^{-1} \sum_{i=1}^K (\hat{\theta}_i - \theta_0^h)^2$ give insights in the behaviour of the different estimators. In addition, the proportions of samples which do fit the uniform association model are given. Notice that samples not fitting the model on an $\alpha = 95\%$ level are excluded from the summary statistics.

The results can be found in the Tables 4.11 and 4.12 for the local association and in Tables 4.13 and 4.14 for the global uniform association. For the local association the estimators LogNI, BDNI and MNI have been added. It turns out, that most estimators give adequate estimates compared to the MLE. The BDNI and MNI are below the MLE for the cannabis data, while the MH (S) estimator gives much smaller values for the sexual fun, job satisfaction and welfare spending data, even changing the sign for the first two mentioned data sets. This is due to a degenerate weights estimation (cf. Remark 4.3.7). Thus, use of the MH (S) estimator or the BDNI or MNI for the local association is not recommended. A similar observation can be made for the global association parameter,

		Conditional on 'no sampling zeros'							using single cell correction ($c = 0.5$)						
Data		Summary Statistics						Sqr.	Summary Statistics						Sqr.
source	Type	Min	1Q	Med	Mean	3Q	Max	Error	Min	1Q	Med	Mean	3Q	Max	Error
Streptococcus	MLE	-0.132	0.315	0.423	0.427	0.539	0.964	0.029	-0.128	0.333	0.437	0.438	0.546	0.922	0.026
$n = 1398$	WM (S)	-0.126	0.315	0.422	0.426	0.538	0.959	0.028	-0.128	0.333	0.435	0.437	0.544	0.920	0.026
2×3	WM (I)	-0.177	0.304	0.405	0.414	0.521	1.039	0.029	-0.124	0.322	0.421	0.425	0.534	0.949	0.027
	MH (S)	-0.210	0.289	0.394	0.400	0.507	0.943	0.030	-0.132	0.307	0.409	0.409	0.519	0.911	0.027
	MH (I)	-0.281	0.274	0.374	0.389	0.497	1.020	0.032	-0.129	0.288	0.394	0.398	0.514	0.878	0.030
	LogNI	-0.177	0.304	0.405	0.414	0.521	1.039	0.029	-0.124	0.322	0.421	0.425	0.534	0.949	0.027
	BDNI	-0.130	0.321	0.431	0.433	0.548	0.957	0.029	-0.127	0.339	0.444	0.445	0.555	0.903	0.027
Fit: 0.939/0.924	MNI	-0.173	0.301	0.398	0.406	0.511	0.960	0.027	-0.124	0.318	0.415	0.416	0.523	0.886	0.025
Cannabis	MLE	0.552	0.741	0.791	0.791	0.840	1.082	0.006	0.562	0.745	0.801	0.804	0.854	1.048	0.006
$n = 1054$	WM (S)	0.544	0.722	0.776	0.775	0.822	1.050	0.006	0.543	0.731	0.787	0.788	0.843	1.013	0.006
4×3	WM (I)	0.522	0.700	0.756	0.749	0.800	0.910	0.008	0.482	0.733	0.805	0.801	0.875	1.039	0.009
	MH (S)	-0.216	0.621	0.684	0.672	0.747	0.997	0.031	0.046	0.631	0.692	0.690	0.760	0.963	0.024
	MH (I)	-0.563	0.543	0.638	0.600	0.711	0.857	0.069	-0.368	0.574	0.681	0.647	0.759	0.956	0.055
	LogNI	0.522	0.700	0.756	0.749	0.800	0.910	0.008	0.482	0.733	0.805	0.801	0.875	1.039	0.009
	BDNI	0.455	0.545	0.567	0.567	0.589	0.678	0.057	0.452	0.549	0.572	0.572	0.595	0.673	0.054
Fit: 0.892/0.902	MNI	0.465	0.570	0.595	0.592	0.616	0.684	0.045	0.442	0.581	0.606	0.605	0.630	0.724	0.041
Sexual Fun	MLE	-0.009	0.209	0.273	0.274	0.332	0.560	0.009	0.015	0.228	0.285	0.291	0.358	0.626	0.009
$n = 91$	WM (S)	-0.010	0.203	0.263	0.264	0.319	0.557	0.008	0.014	0.214	0.268	0.275	0.334	0.612	0.008
4×4	WM (I)	-0.017	0.203	0.271	0.272	0.337	0.561	0.010	-0.006	0.224	0.286	0.293	0.371	0.689	0.011
	MH (S)	-1.229	-0.326	-0.127	-0.175	0.018	0.354	0.284	-1.697	-0.376	-0.175	-0.220	-0.020	0.427	0.342
	MH (I)	-1.060	-0.308	-0.123	-0.163	0.024	0.347	0.274	-1.680	-0.355	-0.162	-0.200	0.006	0.448	0.322
	LogNI	-0.017	0.203	0.271	0.272	0.337	0.561	0.010	-0.006	0.224	0.286	0.293	0.371	0.689	0.011
	BDNI	-0.009	0.202	0.255	0.251	0.301	0.438	0.007	0.015	0.219	0.265	0.265	0.318	0.480	0.006
Fit: 0.874/0.782	MNI	-0.016	0.192	0.254	0.252	0.313	0.481	0.008	-0.004	0.212	0.267	0.267	0.333	0.550	0.008
Job	MLE	-0.188	0.076	0.143	0.143	0.210	0.397	0.014	-0.093	0.110	0.179	0.181	0.248	0.522	0.012
Satisfication	WM (S)	-0.171	0.072	0.136	0.136	0.203	0.377	0.014	-0.139	0.103	0.171	0.173	0.241	0.558	0.013
$n = 96$	WM (I)	-0.222	0.085	0.149	0.152	0.219	0.489	0.014	-0.098	0.127	0.202	0.202	0.280	0.643	0.013
4×4	MH (S)	-1.272	-0.439	-0.270	-0.307	-0.109	0.264	0.343	-1.570	-0.419	-0.229	-0.281	-0.095	0.314	0.319
	MH (I)	-1.394	-0.430	-0.252	-0.292	-0.098	0.262	0.325	-2.240	-0.462	-0.243	-0.305	-0.093	0.311	0.363
	LogNI	-0.222	0.085	0.149	0.152	0.219	0.489	0.014	-0.098	0.127	0.202	0.202	0.280	0.643	0.013
	BDNI	-0.184	0.076	0.141	0.140	0.205	0.360	0.013	-0.093	0.110	0.177	0.175	0.240	0.454	0.011
Fit: 0.961/0.934	MNI	-0.212	0.080	0.144	0.145	0.211	0.420	0.013	-0.090	0.118	0.189	0.187	0.257	0.523	0.011
Welfare	MLE	-0.114	-0.035	-0.015	-0.012	0.013	0.085	0.001	-0.100	-0.036	-0.013	-0.014	0.009	0.066	0.001
Spending	WM (S)	-0.110	-0.034	-0.012	-0.010	0.016	0.087	0.001	-0.097	-0.033	-0.012	-0.012	0.010	0.068	0.001
$n = 955$	WM (I)	-0.121	-0.038	-0.017	-0.014	0.010	0.079	0.001	-0.102	-0.038	-0.016	-0.016	0.007	0.064	0.001
3×5	MH (S)	-0.225	-0.081	-0.052	-0.054	-0.024	0.059	0.003	-0.203	-0.086	-0.056	-0.060	-0.031	0.042	0.004
	MH (I)	-0.223	-0.086	-0.057	-0.058	-0.028	0.055	0.004	-0.198	-0.089	-0.060	-0.064	-0.035	0.042	0.004
	LogNI	-0.121	-0.038	-0.017	-0.014	0.010	0.079	0.001	-0.102	-0.038	-0.016	-0.016	0.007	0.064	0.001
	BDNI	0.113	-0.035	-0.015	-0.012	0.013	0.084	0.001	-0.100	-0.036	-0.013	-0.014	0.009	0.066	0.001
Fit: 0.382/0.379	MNI	-0.119	-0.037	-0.017	-0.014	0.010	0.079	0.001	-0.101	-0.038	-0.016	-0.016	0.007	0.063	0.001
Intellectual	MLE	0.159	0.175	0.178	0.179	0.183	0.201	0.000	0.162	0.175	0.178	0.178	0.183	0.196	0.000
Capacity (Girls)	WM (S)	0.150	0.167	0.171	0.171	0.175	0.194	0.000	0.153	0.167	0.170	0.171	0.174	0.189	0.000
$n = 8310$	WM (I)	0.169	0.187	0.193	0.193	0.198	0.234	0.000	0.172	0.187	0.192	0.192	0.197	0.226	0.000
6×7	MH (S)	0.084	0.124	0.132	0.131	0.139	0.164	0.002	0.075	0.123	0.132	0.131	0.138	0.165	0.002
	MH (I)	0.036	0.136	0.146	0.144	0.154	0.184	0.001	0.040	0.136	0.145	0.144	0.153	0.184	0.001
	LogNI	0.169	0.187	0.193	0.193	0.198	0.234	0.000	0.172	0.187	0.192	0.192	0.197	0.226	0.000
	BDNI	0.144	0.156	0.159	0.159	0.162	0.174	0.000	0.147	0.156	0.158	0.158	0.161	0.170	0.000
Fit: 0.000/0.000	MNI	0.152	0.166	0.169	0.169	0.172	0.189	0.000	0.156	0.165	0.169	0.169	0.172	0.185	0.000

Table 4.11: Simulation study for the different estimation techniques for the **local** uniform association parameter θ_0^L with $K = 1,000$ simulations under the **saturated** model for the sample data. Tables are generated conditional on 'no sampling zeros' (left) and using single cell correction with $c = 0.5$ (right). Summary statistics are based on the tables fitting the local uniform association model. Squared errors to the true model parameter are given. The proportions of samples fitting the model (*Fit*) is given based on the conditional samples (left) and the corrected samples (right). Notice that for the job satisfication data the saturated model under single cell correction has been used, to prevent structural zeros. Since for the intellectual capacity (Girls) data, the uniform association model never holds, summary statistics are given without conditioning on model fit.

		Conditional on 'no sampling zeros'							using single cell correction ($c = 0.5$)						
Data source	Type	Summary Statistics Min	1Q	Med	Mean	3Q	Max	Sqr. Error	Summary Statistics Min	1Q	Med	Mean	3Q	Max	Sqr. Error
Streptococcus	MLE	-0.110	0.322	0.436	0.435	0.542	1.011	0.028	-0.072	0.322	0.423	0.429	0.534	1.083	0.026
$n = 1398$	WM (S)	-0.107	0.322	0.434	0.434	0.540	1.007	0.028	-0.072	0.322	0.423	0.428	0.533	1.083	0.026
2×3	WM (I)	-0.122	0.316	0.438	0.437	0.551	1.113	0.031	-0.077	0.318	0.432	0.435	0.536	1.082	0.029
	MH (S)	-0.130	0.300	0.409	0.412	0.518	0.987	0.029	-0.110	0.295	0.406	0.405	0.507	1.083	0.027
	MH (I)	-0.145	0.295	0.416	0.415	0.524	1.031	0.031	-0.124	0.296	0.412	0.412	0.517	1.082	0.029
	LogNI	-0.122	0.316	0.438	0.437	0.551	1.113	0.031	-0.077	0.318	0.432	0.435	0.536	1.082	0.029
	BDNI	-0.109	0.327	0.444	0.441	0.552	1.000	0.029	-0.072	0.327	0.433	0.436	0.544	1.055	0.026
Fit: 0.940/0.947	MNI	-0.121	0.314	0.430	0.428	0.538	1.003	0.028	-0.077	0.316	0.426	0.426	0.525	0.994	0.026
Cannabis	MLE	0.571	0.734	0.787	0.786	0.835	1.038	0.006	0.569	0.748	0.799	0.804	0.853	1.054	0.006
$n = 1054$	WM (S)	0.545	0.720	0.770	0.768	0.818	1.008	0.007	0.548	0.735	0.784	0.787	0.838	1.015	0.006
4×3	WM (I)	0.538	0.700	0.758	0.751	0.803	0.950	0.008	0.544	0.751	0.821	0.811	0.875	1.035	0.008
	MH (S)	-0.219	0.613	0.687	0.669	0.750	0.949	0.034	-0.341	0.643	0.708	0.698	0.776	1.007	0.028
	MH (I)	-0.757	0.518	0.630	0.583	0.719	0.846	0.087	-0.964	0.584	0.700	0.651	0.779	0.957	0.068
	LogNI	0.538	0.700	0.758	0.751	0.803	0.950	0.008	0.544	0.751	0.821	0.811	0.875	1.035	0.008
	BDNI	0.463	0.542	0.566	0.564	0.588	0.679	0.058	0.460	0.550	0.573	0.573	0.596	0.672	0.054
Fit: 0.959/0.974	MNI	0.484	0.568	0.594	0.593	0.618	0.702	0.045	0.482	0.583	0.607	0.606	0.631	0.697	0.040
Sexual Fun	MLE	-0.025	0.218	0.275	0.279	0.334	0.568	0.008	-0.078	0.225	0.288	0.294	0.359	0.662	0.009
$n = 91$	WM (S)	-0.026	0.207	0.263	0.267	0.322	0.543	0.008	-0.072	0.212	0.273	0.278	0.339	0.622	0.009
4×4	WM (I)	-0.023	0.226	0.292	0.290	0.351	0.556	0.009	-0.086	0.231	0.305	0.310	0.375	0.692	0.012
	MH (S)	-1.358	-0.299	-0.102	-0.151	0.047	0.336	0.266	-1.831	-0.362	-0.131	-0.197	0.037	0.390	0.339
	MH (I)	-1.293	-0.279	-0.083	-0.133	0.063	0.368	0.251	-1.800	-0.334	-0.115	-0.173	0.054	0.460	0.315
	LogNI	-0.023	0.226	0.292	0.290	0.351	0.556	0.009	-0.086	0.231	0.305	0.310	0.375	0.692	0.012
	BDNI	-0.025	0.208	0.258	0.255	0.302	0.441	0.006	-0.077	0.215	0.268	0.267	0.321	0.490	0.006
Fit: 0.970/0.964	MNI	-0.023	0.216	0.275	0.271	0.327	0.503	0.007	-0.085	0.222	0.285	0.285	0.346	0.543	0.008
Job	MLE	-0.159	0.088	0.152	0.155	0.220	0.453	0.013	-0.151	0.123	0.196	0.198	0.275	0.649	0.013
Satisfication	WM (S)	-0.148	0.083	0.144	0.144	0.208	0.402	0.013	-0.140	0.112	0.185	0.185	0.255	0.604	0.012
$n = 96$	WM (I)	-0.145	0.091	0.156	0.158	0.226	0.496	0.014	-0.155	0.125	0.200	0.204	0.284	0.649	0.015
4×4	MH (S)	-1.310	-0.333	-0.164	-0.206	-0.037	0.232	0.231	-1.547	-0.370	-0.174	-0.230	-0.029	0.315	0.282
	MH (I)	-1.271	-0.311	-0.155	-0.191	-0.022	0.260	0.217	-1.892	-0.353	-0.158	-0.214	-0.009	0.330	0.269
	LogNI	-0.145	0.091	0.156	0.158	0.226	0.496	0.014	-0.155	0.125	0.200	0.204	0.284	0.649	0.015
	BDNI	-0.155	0.088	0.151	0.150	0.216	0.390	0.013	-0.150	0.122	0.193	0.190	0.263	0.558	0.011
Fit: 0.986/0.984	MNI	-0.144	0.088	0.151	0.151	0.217	0.429	0.013	-0.153	0.120	0.190	0.192	0.270	0.517	0.013
Welfare	MLE	-0.120	-0.037	-0.015	-0.014	0.010	0.121	0.001	-0.121	-0.034	-0.012	-0.012	0.010	0.101	0.001
Spending	WM (S)	-0.114	-0.036	-0.015	-0.014	0.009	0.122	0.001	-0.121	-0.034	-0.012	-0.012	0.011	0.099	0.001
$n = 955$	WM (I)	-0.119	-0.036	-0.014	-0.014	0.009	0.121	0.001	-0.118	-0.034	-0.011	-0.012	0.011	0.103	0.001
3×5	MH (S)	-0.246	-0.080	-0.050	-0.055	-0.026	0.059	0.004	-0.270	-0.080	-0.051	-0.054	-0.023	0.077	0.004
	MH (I)	-0.247	-0.080	-0.049	-0.055	-0.026	0.060	0.004	-0.272	-0.081	-0.051	-0.054	-0.023	0.080	0.004
	LogNI	-0.119	-0.036	-0.014	-0.014	0.009	0.121	0.001	-0.118	-0.034	-0.011	-0.012	0.011	0.103	0.001
	BDNI	-0.119	-0.037	-0.015	-0.014	0.010	0.119	0.001	-0.121	-0.034	-0.012	-0.012	0.010	0.099	0.001
Fit: 0.950/0.947	MNI	-0.118	-0.036	-0.014	-0.014	0.009	0.120	0.001	-0.117	-0.033	-0.011	-0.012	0.011	0.102	0.001
Intellectual	MLE	0.157	0.175	0.178	0.179	0.182	0.198	0.000	0.163	0.175	0.179	0.179	0.183	0.197	0.000
Capacity (Girls)	WM (S)	0.157	0.174	0.177	0.178	0.181	0.196	0.000	0.161	0.174	0.178	0.178	0.182	0.196	0.000
$n = 8310$	WM (I)	0.158	0.176	0.180	0.181	0.185	0.204	0.000	0.161	0.176	0.180	0.181	0.185	0.204	0.000
6×7	MH (S)	0.145	0.164	0.168	0.168	0.173	0.188	0.000	0.144	0.164	0.169	0.169	0.173	0.186	0.000
	MH (I)	0.146	0.164	0.169	0.169	0.174	0.191	0.000	0.143	0.165	0.169	0.169	0.174	0.190	0.000
	LogNI	0.158	0.176	0.180	0.181	0.185	0.204	0.000	0.161	0.176	0.180	0.181	0.185	0.204	0.000
	BDNI	0.143	0.156	0.158	0.158	0.161	0.172	0.000	0.147	0.156	0.159	0.159	0.162	0.172	0.000
Fit: 0.955/0.943	MNI	0.149	0.163	0.167	0.167	0.170	0.181	0.000	0.153	0.164	0.167	0.167	0.170	0.183	0.000

Table 4.12: Simulation study for the different estimation techniques for the **local** association parameter θ_0^L with $K = 1,000$ simulations under the **local uniform association** model for the sample data. Tables are generated conditional on 'no sampling zeros' (left) and using single cell correction with $c = 0.5$ (right). Summary statistics are based on the tables fitting the local uniform association model. Squared errors to the true model parameter are given. The proportions of samples fitting the model (*Fit*) is given based on the conditional samples (left) and the corrected samples (right).

		Conditional on 'no sampling zeros'							using single cell correction ($c = 0.5$)						
Data source	Type	Summary Statistics Min	1Q	Med	Mean	3Q	Max	Sqr. Error	Summary Statistics Min	1Q	Med	Mean	3Q	Max	Sqr. Error
Streptococcus	MLE	-0.141	0.450	0.616	0.619	0.785	1.371	0.098	-0.194	0.448	0.601	0.603	0.765	1.340	0.084
$n = 1398$	WM (S)	-0.140	0.450	0.614	0.617	0.782	1.365	0.097	-0.186	0.448	0.598	0.602	0.762	1.340	0.083
2×3	WM (I)	-0.148	0.440	0.599	0.604	0.776	1.448	0.092	-0.177	0.429	0.580	0.583	0.731	1.401	0.076
	MH (S)	-0.143	0.442	0.607	0.609	0.772	1.365	0.094	-0.207	0.440	0.588	0.592	0.754	1.340	0.080
Fit: 0.920/0.923	MH (I)	-0.151	0.432	0.592	0.596	0.773	1.441	0.090	-0.198	0.420	0.575	0.574	0.721	1.396	0.073
Cannabis	MLE	1.371	1.754	1.861	1.861	1.973	2.332	1.145	1.351	1.765	1.869	1.872	1.979	2.380	1.168
$n = 1054$	WM (S)	1.368	1.752	1.860	1.860	1.974	2.317	1.143	1.350	1.765	1.869	1.872	1.984	2.381	1.167
4×3	WM (I)	1.409	1.801	1.917	1.932	2.051	2.567	1.310	1.361	1.837	1.964	1.973	2.104	2.553	1.406
	MH (S)	1.368	1.747	1.849	1.854	1.967	2.280	1.130	1.347	1.760	1.864	1.867	1.976	2.381	1.157
Fit: 0.692/0.612	MH (I)	1.402	1.782	1.889	1.904	2.018	2.528	1.246	1.343	1.815	1.933	1.939	2.069	2.488	1.325
Sexual Fun	MLE	-0.074	0.821	1.042	1.040	1.277	2.066	0.678	-0.280	0.870	1.128	1.116	1.370	2.191	0.827
$n = 91$	WM (S)	-0.130	0.760	0.988	0.976	1.203	2.061	0.576	-0.296	0.816	1.045	1.050	1.295	2.153	0.708
4×4	WM (I)	-0.080	0.811	1.032	1.046	1.296	2.016	0.694	-0.258	0.863	1.130	1.128	1.390	2.454	0.865
	MH (S)	-0.177	0.633	0.855	0.857	1.078	2.030	0.429	-0.413	0.679	0.920	0.918	1.157	2.135	0.525
Fit: 0.841/0.754	MH (I)	-0.182	0.704	0.919	0.934	1.178	1.989	0.536	-0.386	0.740	1.012	0.999	1.254	2.263	0.657
Job	MLE	-0.373	0.263	0.445	0.458	0.650	1.504	0.151	-0.426	0.314	0.532	0.530	0.733	1.698	0.194
Satisfication	WM (S)	-0.399	0.255	0.426	0.445	0.635	1.478	0.140	-0.409	0.305	0.519	0.513	0.720	1.639	0.180
$n = 96$	WM (I)	-0.401	0.277	0.468	0.480	0.685	1.485	0.171	-0.410	0.354	0.591	0.588	0.809	1.720	0.250
4×4	MH (S)	-0.502	0.189	0.356	0.370	0.557	1.410	0.110	-0.488	0.217	0.429	0.423	0.627	1.494	0.136
Fit: 0.977/0.941	MH (I)	-0.502	0.200	0.382	0.395	0.600	1.409	0.133	-0.586	0.241	0.483	0.475	0.694	1.612	0.182
Welfare	MLE	-0.322	-0.091	-0.019	-0.019	0.060	0.365	0.011	-0.366	-0.099	-0.020	-0.023	0.053	0.277	0.013
Spending	WM (S)	-0.320	-0.089	-0.016	-0.019	0.060	0.365	0.011	-0.364	-0.098	-0.020	-0.022	0.053	0.275	0.013
$n = 955$	WM (I)	-0.339	-0.092	-0.021	-0.021	0.059	0.353	0.011	-0.387	-0.102	-0.023	-0.025	0.052	0.279	0.013
3×5	MH (S)	-0.342	-0.107	-0.035	-0.035	0.040	0.352	0.011	-0.373	-0.112	-0.038	-0.039	0.035	0.251	0.013
Fit: 0.369/0.369	MH (I)	-0.366	-0.109	-0.040	-0.038	0.040	0.344	0.012	-0.399	-0.119	-0.041	-0.042	0.033	0.257	0.014
Intellectual	MLE	1.004	1.074	1.095	1.095	1.116	1.192	0.840	0.987	1.073	1.095	1.095	1.115	1.215	0.840
Capacity (Girls)	WM (S)	1.012	1.080	1.101	1.102	1.122	1.199	0.853	0.995	1.080	1.102	1.101	1.122	1.219	0.853
$n = 8310$	WM (I)	1.039	1.140	1.164	1.164	1.188	1.274	0.972	1.044	1.140	1.164	1.164	1.188	1.279	0.972
6×7	MH (S)	0.989	1.063	1.086	1.085	1.107	1.182	0.823	0.979	1.063	1.086	1.085	1.107	1.206	0.823
Fit: 0.000/0.000	MH (I)	1.005	1.106	1.129	1.129	1.153	1.237	0.904	1.009	1.106	1.130	1.129	1.153	1.242	0.905

Table 4.13: Simulation study for the different estimation techniques for the **global** association parameter θ_0^G with $K = 1,000$ simulations under the **saturated** model for the sample data. Tables are generated conditional on 'no sampling zeros' (left) and using single cell correction with $c = 0.5$ (right). Summary statistics are based on the tables fitting the global uniform association model. Squared errors to the true model parameter and percentage of model fit are given. Remark that for the job satisfication data the saturated model under single cell correction has been used to prevent structural zeros. Since for the intellectual capacity (Girls) data, the uniform association model never holds, summary statistics are given without conditioning on model fit.

		Conditional on 'no sampling zeros'							using single cell correction ($c = 0.5$)						
Data source	Type	Summary Statistics Min	1Q	Med	Mean	3Q	Max	Sqr. Error	Summary Statistics Min	1Q	Med	Mean	3Q	Max	Sqr. Error
Streptococcus	MLE	-0.022	0.465	0.609	0.607	0.754	1.276	0.051	-0.143	0.454	0.591	0.599	0.743	1.384	0.051
$n = 1398$	WM (S)	-0.022	0.465	0.607	0.606	0.752	1.274	0.050	-0.139	0.454	0.589	0.597	0.742	1.383	0.051
2×3	WM (I)	-0.022	0.461	0.615	0.614	0.767	1.358	0.056	-0.133	0.449	0.586	0.603	0.744	1.435	0.057
	MH (S)	-0.022	0.455	0.600	0.598	0.742	1.271	0.050	-0.154	0.441	0.583	0.589	0.733	1.382	0.051
Fit: 0.950/0.948	MH (I)	-0.022	0.457	0.609	0.606	0.755	1.353	0.055	-0.148	0.439	0.579	0.595	0.740	1.433	0.056
Cannabis	MLE	1.315	1.766	1.865	1.871	1.976	2.361	0.027	1.432	1.749	1.857	1.858	1.966	2.493	0.025
$n = 1054$	WM (S)	1.305	1.762	1.860	1.866	1.973	2.355	0.027	1.427	1.745	1.850	1.852	1.960	2.488	0.025
4×3	WM (I)	1.205	1.741	1.884	1.888	2.027	2.616	0.042	1.329	1.744	1.874	1.876	1.996	2.673	0.037
	MH (S)	1.284	1.756	1.852	1.858	1.968	2.350	0.027	1.417	1.739	1.841	1.844	1.952	2.479	0.026
Fit: 0.958/0.937	MH (I)	1.188	1.727	1.870	1.873	2.012	2.544	0.040	1.317	1.731	1.859	1.859	1.982	2.606	0.035
Sexual Fun	MLE	-0.123	0.868	1.085	1.075	1.293	2.115	0.109	-0.064	0.877	1.094	1.093	1.328	2.205	0.118
$n = 91$	WM (S)	-0.111	0.821	1.030	1.026	1.233	1.962	0.106	-0.031	0.832	1.049	1.046	1.284	2.055	0.113
4×4	WM (I)	-0.142	0.867	1.091	1.093	1.316	2.101	0.116	-0.056	0.888	1.107	1.116	1.361	2.222	0.131
	MH (S)	-0.168	0.769	0.979	0.970	1.171	1.919	0.118	-0.058	0.778	0.991	0.991	1.230	2.010	0.126
Fit: 0.969/0.952	MH (I)	-0.215	0.803	1.032	1.020	1.241	2.048	0.119	-0.093	0.819	1.030	1.036	1.284	2.032	0.131
Job	MLE	-0.533	0.288	0.504	0.504	0.729	1.676	0.106	-0.617	0.305	0.524	0.527	0.753	1.582	0.113
Satisfaction	WM (S)	-0.507	0.283	0.481	0.487	0.705	1.660	0.102	-0.601	0.299	0.507	0.510	0.730	1.475	0.108
$n = 96$	WM (I)	-0.507	0.288	0.506	0.506	0.733	1.628	0.106	-0.589	0.316	0.526	0.533	0.761	1.890	0.119
4×4	MH (S)	-0.573	0.206	0.413	0.415	0.630	1.629	0.117	-0.906	0.217	0.426	0.434	0.654	1.445	0.123
Fit: 0.984/0.988	MH (I)	-0.569	0.214	0.429	0.428	0.653	1.505	0.118	-0.866	0.223	0.441	0.449	0.678	1.680	0.128
Welfare	MLE	-0.397	-0.088	-0.014	-0.014	0.062	0.331	0.012	-0.346	-0.093	-0.019	-0.017	0.060	0.286	0.012
Spending	WM (S)	-0.396	-0.087	-0.013	-0.014	0.062	0.328	0.012	-0.344	-0.092	-0.018	-0.017	0.060	0.286	0.012
$n = 955$	WM (I)	-0.390	-0.088	-0.013	-0.015	0.061	0.340	0.012	-0.345	-0.094	-0.019	-0.017	0.060	0.290	0.012
3×5	MH (S)	-0.409	-0.096	-0.020	-0.022	0.054	0.319	0.012	-0.352	-0.103	-0.026	-0.025	0.054	0.273	0.012
Fit: 0.938/0.943	MH (I)	-0.403	-0.098	-0.020	-0.023	0.053	0.330	0.012	-0.353	-0.101	-0.025	-0.025	0.054	0.267	0.012
Intellectual	MLE	0.980	1.075	1.096	1.096	1.120	1.231	0.001	0.997	1.072	1.096	1.096	1.118	1.196	0.001
Capacity (Girls)	WM (S)	0.980	1.074	1.095	1.095	1.119	1.229	0.001	0.996	1.072	1.096	1.095	1.118	1.196	0.001
$n = 8310$	WM (I)	0.978	1.072	1.098	1.097	1.122	1.258	0.001	0.995	1.072	1.097	1.097	1.122	1.223	0.001
6×7	MH (S)	0.979	1.073	1.094	1.094	1.118	1.226	0.001	0.994	1.071	1.095	1.094	1.116	1.196	0.001
Fit: 0.956/0.944	MH (I)	0.977	1.071	1.096	1.095	1.120	1.255	0.001	0.993	1.071	1.095	1.095	1.120	1.219	0.001

Table 4.14: Simulation study for the different estimation techniques for the **global** association parameter θ_0^G with $K = 1,000$ simulations under the **global uniform association** model for the sample data. Tables are generated conditional on 'no sampling zeros' (left) and using single cell correction with $c = 0.5$ (right). Summary statistics are based on the tables fitting the global uniform association model. Squared errors to the true model parameter and percentage of model fit are given.

where MH estimation undervalues the true value for the sexual fun and job satisfaction data. All closed-form estimators give more precise results (closer to MLE) when the simulation study is based on the uniform association model, compared to the study based on the saturated model. For the intellectual capacity (Girls) data the mean of the MH (S) estimate for using single cell correction ($c = 0.5$) is 0.193 (MLE: 0.178) for the study based on the saturated model, while it is 0.180 (MLE: 0.179), when the local uniform association model holds. This effect seems to be stronger for the global uniform association model, where all estimation techniques give very similar results when the global uniform association model holds.

Overall, the WM (I) estimation seems to be a good alternative, giving a good compromise between compatibility with sampling zeros and quality of the estimation. In addition, formulas for the weights $\boldsymbol{\omega} = \boldsymbol{\omega}(S)$ and $\boldsymbol{\omega} = \boldsymbol{\omega}(I)$ are available in the case of global uniform association (Clayton [1974]). Notice that the closed-form estimator should only be applied, if the uniform association model holds, which can be checked using the non-iterative goodness-of-fit tests which will be presented in Section 4.7. Otherwise the use of uniform association parameters as measures of association can be misleading in a similar way as shown in the motivation (Section 4.1).

4.5 Estimation of Row-Effect Association

As the motivation (Section 4.1) shows, a one-dimensional measure can not always approach the association structure in contingency tables. The previous results for the uniform association model based measures on more complex models, which were presented in Section 1.9.4, shall be generalised. The association structure of an $I \times J$ contingency table can be described using $(I-1)(J-1)$ parameters, for example using the local odds ratios (cf. Remark 1.9.1). The R and C models, (1.57) and (1.58), classify the association assuming constant local row ($\log \theta_{ij}^L = c_{1i}, i = 1, ..., I-1, j = 1, ..., J-1$) and column ($\log \theta_{ij}^L = c_{2j}, i = 1, ..., I-1, j = 1, ..., J-1$) odds ratio for some parameters $\{c_{1i}\}$ and $\{c_{2j}\}$ and therefore give a more complex, but still manageable association structure. Clearly, the C model becomes the R model, when the classification variables are changed. Thus, the study can be restricted on the R model.

Section 4.3 presented closed-form estimates introduced by Clayton [1974] for the generalised uniform association model parameter ($\theta_{ij}^h = c, i = 1, ..., I-1, j = 1, ..., J-1$), focusing on local and global association ($h = L, G$), making maximum likelihood estimation dispensable in these cases. Clayton's ideas can be spread to the R model and the generalised odds ratios. Let

$$\boldsymbol{\theta}_i^h = (\theta_{i1}^h, ..., \theta_{iJ-1}^h), \qquad i = 1, ..., I-1, \tag{4.15}$$

be the generalised odds ratios in row i. The generalised R-model based on $\log \boldsymbol{\theta}_i^h$ becomes

$$\log \theta_{ij}^h = c_{1i}, \qquad\qquad i = 1, ..., I-1 \tag{4.16}$$

Model (4.16) is a GLLM (1.59) and an iterative maximum likelihood estimation technique introduced by Lang [2004, 2005] is available with the R-function `mph.fit` (cf. Section 1.9.5), which is used in the R function `calc.MLE.R` to get MLEs for model (4.16) (see Appendix B.4).

4.5.1 R-Weighted Mean and R-Mantel Haenszel Estimators

Let $\mathbf{N} \sim \mathcal{M}(n, \boldsymbol{\pi})$ be multinomial distributed with sample size $n \in \mathbb{N}$ and $I \times J$ multinomial probability vector $\boldsymbol{\pi} \in \Delta_{I,J}$. Let $\mathbf{n}$ be a realisation of $\mathbf{N}$. Closed-form estimators for model parameters in (4.16) will be studied. Following the basic concept of the uniform weighted mean estimator, a R-weighted mean (R-WM) estimator can be defined by weighting the odds ratio estimates in each row. Let $\hat{\boldsymbol{\theta}}^{h,RWM}$ be the R-WM estimator defined as

$$\log \hat{\boldsymbol{\theta}}^{h,RWM} = \log(\hat{\theta}_1^{h,RWM}, ..., \hat{\theta}_{I-1}^{h,RWM})$$

with

$$\log \hat{\theta}_i^{h,RWM} = \left(\boldsymbol{\omega}^{(i)}\right)^T \log \hat{\boldsymbol{\theta}}_i^h, \qquad i = 1, ..., I-1,$$

where $\boldsymbol{\omega}^{(i)} = (\omega_1^{(i)}, ..., \omega_{J-1}^{(i)})$ are weights with $\sum_{j=1}^{J-1} \omega_j^{(i)} = 1, i = 1, ..., I-1$. Using the $(I-1) \times (I-1)(J-1)$ matrix

$$\mathbf{A} = \begin{pmatrix} \boldsymbol{\omega}^{(1)} & & \\ & \ddots & \\ & & \boldsymbol{\omega}^{(I-1)} \end{pmatrix}$$

the R-WM estimator becomes

$$\log \hat{\boldsymbol{\theta}}^{h,RWM} = \mathbf{A} \log \hat{\boldsymbol{\theta}}^h. \tag{4.17}$$

The weights can again be chosen variance minimal in the sense, that each $\mathbb{V}\text{ar}\left((\boldsymbol{\omega}^{(i)})^T \log \hat{\boldsymbol{\theta}}_i^h\right), i = 1, ..., I-1$ is minimal: Let $\boldsymbol{\Sigma}^h(\boldsymbol{\pi})$ be the variance-covariance matrix of $\log \hat{\boldsymbol{\theta}}^h$ given in (1.45), which is partitioned according to the row odds ratios $\log \hat{\boldsymbol{\theta}}_i^h$, e.g. $\boldsymbol{\Sigma}^h(\boldsymbol{\pi}) = (\boldsymbol{\Sigma}_{kl}^h(\boldsymbol{\pi}))$, where

$$\boldsymbol{\Sigma}_{kl}^h(\boldsymbol{\pi}) = \text{Cov}(\log \hat{\boldsymbol{\theta}}_k^h, \log \hat{\boldsymbol{\theta}}_l^h), \qquad k, l = 1, ..., (I-1). \tag{4.18}$$

Then $\boldsymbol{\Sigma}_{kk}^h(\boldsymbol{\pi})$ is the variance-covariance matrix of $\log \hat{\boldsymbol{\theta}}_k^h$ and the variance optimal weights are

$$\boldsymbol{\omega}^{(k)} = \frac{\left(\boldsymbol{\Sigma}_{kk}^h(\boldsymbol{\pi})\right)^{-1} \mathbf{1}}{\mathbf{1}^T \left(\boldsymbol{\Sigma}_{kk}^h(\boldsymbol{\pi})\right)^{-1} \mathbf{1}}, \qquad k = 1, ..., I-1, \tag{4.19}$$

by Lemma 4.3.2. Using $\log \hat{\boldsymbol{\theta}}^{h,RWM} = \mathbf{A} \log \hat{\boldsymbol{\theta}}^h$, the variance-covariance matrix for the R-WM estimator becomes

$$\mathbb{V}\text{ar } \log \hat{\boldsymbol{\theta}}^{h,RWM} = \mathbf{A}^T \boldsymbol{\Sigma}^h(\boldsymbol{\pi}) \mathbf{A}.$$

A MH type estimator can be constructed as follows: Let $a_{ij}^h, b_{ij}^h, c_{ij}^h, d_{ij}^h$ be the essential sums (1.41), making out the generalised odds ratios $\boldsymbol{\theta}^h, h = L, CO, C, G$. Let $\tilde{\boldsymbol{\omega}}^{(i)} = (\tilde{\omega}_1^{(i)}, ..., \tilde{\omega}_{J-1}^{(i)}), i = 1, ..., I-1$, be weights. Define the R-Mantel Haenszel (R-MH) estimator as

$$\hat{\boldsymbol{\theta}}^{h,RMH} = \hat{\boldsymbol{\theta}}^{h,RMH}(\mathbf{N}) = (\hat{\theta}_1^{h,RMH}, ..., \hat{\theta}_{I-1}^{h,RMH}) \tag{4.20}$$

where

$$\hat{\theta}_i^{h,RMH} = \frac{\sum_{j=1}^{J-1} \tilde{\omega}_j^{(i)} a_{ij}^h(\mathbf{N}) d_{ij}^h(\mathbf{N})}{\sum_{j=1}^{J-1} \tilde{\omega}_j^{(i)} b_{ij}^h(\mathbf{N}) c_{ij}^h(\mathbf{N})}, \qquad i = 1, ..., I-1.$$

The R-WM and R-MH estimates can be calculated using the R functions `WM.calc.R.model` and `MH.calc.R.model` (see Appendix B.4).

Lemma 4.5.1. The R-WM (4.17) and R-MH (4.20) estimators are consistent under the R-model.

Proof. Assume that the generalised R-model (4.16) holds with parameters $\{c_{1i}\}$. Following the proof of Lemma 4.3.1 it holds $\hat{\theta}_{ij} \xrightarrow{P} c_{1i}, i = 1, ..., I-1, j = 1, ..., J-1$. Thus, for the components $\theta_i^{h,RWM}$ and $\theta_i^{h,RMH}$, given in (4.17) and (4.20), it holds

$$\log \theta_i^{h,RWM} = (\boldsymbol{\omega}^{(i)})^T \log \hat{\boldsymbol{\theta}}_i^h = \sum_{j=1}^{J-1} \omega_j^{(i)} \log \hat{\theta}_{ij}^h \xrightarrow{P} \sum_{j=1}^{J-1} \omega_j^{(i)} c_{1i} = c_{1i} \sum_{j=1}^{J-1} \omega_j^{(i)} = c_{1i}$$

and

$$\begin{aligned} \log \theta_i^{h,RMH} &= \frac{\sum_{j=1}^{J-1} \tilde{\omega}_j^{(i)} a_{ij}^h(\mathbf{N}) d_{ij}^h(\mathbf{N})}{\sum_{j=1}^{J-1} \tilde{\omega}_j^{(i)} b_{ij}^h(\mathbf{N}) c_{ij}^h(\mathbf{N})} = \frac{\sum_{j=1}^{J-1} \tilde{\omega}_j^{(i)} \hat{\theta}_{ij}^h b_{ij}^h(\mathbf{N}) c_{ij}^h(\mathbf{N})}{\sum_{j=1}^{J-1} \tilde{\omega}_j^{(i)} b_{ij}^h(\mathbf{N}) c_{ij}^h(\mathbf{N})} \\ &\xrightarrow{P} \frac{\sum_{j=1}^{J-1} \tilde{\omega}_j^{(i)} c_{1i} b_{ij}^h(\mathbf{N}) c_{ij}^h(\mathbf{N})}{\sum_{j=1}^{J-1} \tilde{\omega}_j^{(i)} b_{ij}^h(\mathbf{N}) c_{ij}^h(\mathbf{N})} = c_{1i} \frac{\sum_{j=1}^{J-1} \tilde{\omega}_j^{(i)} b_{ij}^h(\mathbf{N}) c_{ij}^h(\mathbf{N})}{\sum_{j=1}^{J-1} \tilde{\omega}_j^{(i)} b_{ij}^h(\mathbf{N}) c_{ij}^h(\mathbf{N})} = c_{1i}, \end{aligned}$$

for $i = 1, ..., I-1$. Thus, the R-WM (4.17) and R-MH (4.20) estimators are consistent. □

Similar to Lemma 4.3.4 one can find variance-minimal weights for the R-MH:

Lemma 4.5.2. Let $a_{ij}^h, b_{ij}^h, c_{ij}^h, d_{ij}^h$ be the essential sums (1.41) that build the generalised odds ratios $\boldsymbol{\theta}^h$. Let $\boldsymbol{\omega}^{(i)} = (\omega_1^{(i)}, ..., \omega_{J-1}^{(i)}), i = 1, ..., I-1$, be the variance-minimal weights (4.19) for the R-WM estimator. Then the variance-minimal weights $\tilde{\boldsymbol{\omega}}^{(i)} = (\omega_1^{(i)}, ..., \omega_{J-1}^{(i)}), i = 1, ..., I-1$, for the R-MH estimator are given by

$$\tilde{\omega}_j^{(i)} = \frac{\omega_j^{(i)}}{a_{ij}^h(\mathbf{n}) d_{ij}^h(\mathbf{n})}, \qquad i = 1, ..., I-1, j = 1, ..., J-1.$$

In this case, the R-WM (4.17) and R-MH (4.20) estimators have the same asymptotic variance-covariance matrix.

Proof. Let $\mathbf{N} \sim \mathcal{M}(n, \boldsymbol{\pi})$. Let $\boldsymbol{\Sigma}^h(\boldsymbol{\pi}) = (\boldsymbol{\Sigma}_{kl}^h(\boldsymbol{\pi}))$ the asymptotic variance-covariance matrix of $\log \hat{\boldsymbol{\theta}}^h$ with partition given in (4.18), i.e. $\boldsymbol{\Sigma}_{kl}^h(\boldsymbol{\pi}) = \text{Cov}(\log \hat{\boldsymbol{\theta}}_k^h, \log \hat{\boldsymbol{\theta}}_l^h)$. Following the proof of Lemma 4.3.4, define $\tilde{\boldsymbol{\alpha}}^{(i)}(\mathbf{N}) = (\tilde{\omega}_j^{(i)} b_{ij}^h(\mathbf{N}) c_{ij}^h(\mathbf{N}))$ and $\tilde{\boldsymbol{\alpha}}^{(i)}(\boldsymbol{\pi}) = (\tilde{\omega}_j^{(i)} b_{ij}^h(\boldsymbol{\pi}) c_{ij}^h(\boldsymbol{\pi}))$. It holds

$$\hat{\theta}_i^{h,RMH} = \frac{\sum_{j=1}^{J-1} \tilde{\omega}_j^{(i)} a_{ij}^h(\mathbf{N}) d_{ij}^h(\mathbf{N})}{\sum_{j=1}^{J-1} \tilde{\omega}_j^{(i)} b_{ij}^h(\mathbf{N}) c_{ij}^h(\mathbf{N})} = \frac{\sum_{j=1}^{J-1} \tilde{\omega}_j^{(i)} \hat{\boldsymbol{\theta}}_i b_{ij}^h(\mathbf{N}) c_{ij}^h(\mathbf{N})}{\sum_{j=1}^{J-1} \tilde{\omega}_j^{(i)} b_{ij}^h(\mathbf{N}) c_{ij}^h(\mathbf{N})} = \frac{\left(\tilde{\boldsymbol{\alpha}}^{(i)}(\mathbf{N})\right)^T \hat{\boldsymbol{\theta}}_i / n^2}{\left(\tilde{\boldsymbol{\alpha}}^{(i)}(\mathbf{N})\right)^T \mathbf{1} / n^2}, \tag{4.21}$$

for $i = 1, ..., I-1$. Assuming $(\tilde{\boldsymbol{\alpha}}^{(i)}(\mathbf{N}))^T \mathbf{1}/n^2 = 1$ (see proof of Lemma 4.3.4), the asymptotic variance is given by

$$\mathbb{V}\text{ar } (\hat{\theta}_i^{h,RMH}) = \mathbb{V}\text{ar } (\left(\tilde{\boldsymbol{\alpha}}^{(i)}(\mathbf{N})\right)^T \hat{\boldsymbol{\theta}}_i/n^2) = \mathbb{V}\text{ar } (\left(\tilde{\boldsymbol{\alpha}}^{(i)}(\hat{\boldsymbol{\pi}})\right)^T \hat{\boldsymbol{\theta}}_i).$$

It holds

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_i^h - \boldsymbol{\theta}_i^h) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \text{diag}(\boldsymbol{\theta}_i^h)\boldsymbol{\Sigma}_{ii}^h(\boldsymbol{\pi})\text{diag}(\boldsymbol{\theta}_i^h)), \quad \text{for} \quad n \to \infty.$$

Since $\tilde{\boldsymbol{\alpha}}^{(i)}(\mathbf{N})/n^2 = \tilde{\boldsymbol{\alpha}}^{(i)}(\hat{\boldsymbol{\pi}}) \xrightarrow{P} \tilde{\boldsymbol{\alpha}}^{(i)}(\boldsymbol{\pi})$ and by Slutsky's theorem the asymptotic variance becomes

$$\mathbb{V}\text{ar } \left(\left(\tilde{\boldsymbol{\alpha}}^{(i)}(\hat{\boldsymbol{\pi}})\right)^T \hat{\boldsymbol{\theta}}_i^h\right) = (\tilde{\boldsymbol{\alpha}}^{(i)}(\boldsymbol{\pi}))^T \text{diag}(\boldsymbol{\theta}_i^h)\boldsymbol{\Sigma}_{ii}^h(\boldsymbol{\pi})\text{diag}(\boldsymbol{\theta}_i^h)\tilde{\boldsymbol{\alpha}}^{(i)}(\boldsymbol{\pi}),$$

for $i = 1, ..., I-1$, which gives the asymptotic variance

$$\mathbb{V}\text{ar } \left(\left(\log(\tilde{\boldsymbol{\alpha}}^{(i)}(\hat{\boldsymbol{\pi}})\right)^T \hat{\boldsymbol{\theta}}_i^h)\right) = \frac{(\tilde{\boldsymbol{\alpha}}^{(i)}(\boldsymbol{\pi}))^T \text{diag}(\boldsymbol{\theta}_i^h)}{\left(\tilde{\boldsymbol{\alpha}}^{(i)}(\hat{\boldsymbol{\pi}})\right)^T \boldsymbol{\theta}_i^h} \boldsymbol{\Sigma}_{ii}^h(\boldsymbol{\pi}) \frac{\text{diag}(\boldsymbol{\theta}_i^h)\tilde{\boldsymbol{\alpha}}^{(i)}(\boldsymbol{\pi})}{\left(\tilde{\boldsymbol{\alpha}}^{(i)}(\boldsymbol{\pi})\right)^T \boldsymbol{\theta}_i^h}.$$

Setting $k(\tilde{\boldsymbol{\omega}}^{(i)}) = \left(\tilde{\boldsymbol{\alpha}}^{(i)}(\boldsymbol{\pi})\right)^T \boldsymbol{\theta}_i^h$, the asymptotic variance is minimal if

$$\begin{aligned}
\frac{(\tilde{\boldsymbol{\alpha}}^{(i)}(\boldsymbol{\pi}))^T \text{diag}(\boldsymbol{\theta}_i^h)}{k(\tilde{\boldsymbol{\omega}}^{(i)})} = \boldsymbol{\omega}^{(i)} \quad &\Leftrightarrow \quad \tilde{\omega}_j^{(i)} b_{ij}^h(\boldsymbol{\pi}) c_{ij}^h(\boldsymbol{\pi}) \boldsymbol{\theta}_{ij}^h = k(\tilde{\boldsymbol{\omega}}^{(i)})\omega_j^{(i)}, \quad j = 1, ..., J-1 \\
&\Leftrightarrow \quad \tilde{\omega}_j^{(i)} a_{ij}^h(\boldsymbol{\pi}) d_{ij}^h(\boldsymbol{\pi}) = k(\tilde{\boldsymbol{\omega}}^{(i)})\omega_j^{(i)}, \quad j = 1, ..., J-1 \\
&\Leftrightarrow \quad \tilde{\omega}_j^{(i)} = \frac{k(\tilde{\boldsymbol{\omega}}^{(i)})\omega_j^{(i)}}{a_{ij}^h(\boldsymbol{\pi}) d_{ij}^h(\boldsymbol{\pi})}, \quad j = 1, ..., J-1,
\end{aligned}$$

for $i = 1, .., I-1$. The weights can be estimated by $\tilde{\omega}_j^{(i)} = \frac{\omega_j^{(i)}}{a_{ij}^h(\hat{\boldsymbol{\pi}}) d_{ij}^h(\hat{\boldsymbol{\pi}})} = \frac{k(\tilde{\boldsymbol{\omega}}^{(i)}) n^2 \omega_j^{(i)}}{a_{ij}^h(\mathbf{n}) d_{ij}^h(\mathbf{n})}$. Since by (4.21) the estimator is unaffected when the weights are multiplied by $k(\tilde{\boldsymbol{\omega}}^{(i)})\boldsymbol{\omega}_j^{(i)}$, the adequate choice for the weights is $\tilde{\omega}_j^{(i)} = \frac{\omega_j^{(i)}}{a_{ij}^h(\mathbf{n}) d_{ij}^h(\mathbf{n})}$. □

As in the uniform case, the choice of the weight estimation influences the outcome of the closed-form estimator (cf. Remark 4.3.5). Again, appropriate choices are based on the saturated ($\hat{\boldsymbol{\pi}}$) or the independence ($\hat{\boldsymbol{\pi}}^I = (\hat{\pi}_{i+}\hat{\pi}_{+j})$) model. For the weights $\boldsymbol{\omega} = (\boldsymbol{\omega}^{(1)}, ..., \boldsymbol{\omega}^{(I-1)})$, the notation $\boldsymbol{\omega} = \boldsymbol{\omega}(S)$ is used, when the variance-covariance is estimated by $\boldsymbol{\Sigma}^h(\hat{\boldsymbol{\pi}})$ and $\boldsymbol{\omega} = \boldsymbol{\omega}(I)$ is used for the independence based estimate $\boldsymbol{\Sigma}^h(\hat{\boldsymbol{\pi}}^I)$. In the case of the independence based weights, the following property holds:

Lemma 4.5.3. Let $\boldsymbol{\omega}(I) = (\boldsymbol{\omega}^{(1)}(I), ..., \boldsymbol{\omega}^{(I-1)}(I))$ be the weights for the local R-WM as in (4.17) for $h = L$ based on the estimate $\boldsymbol{\Sigma}^L(\hat{\boldsymbol{\pi}}^I)$. It holds:

$$\boldsymbol{\omega}^{(1)}(I) = ... = \boldsymbol{\omega}^{(I-1)}(I).$$

Proof. Let $\boldsymbol{\pi} \in \Delta_{I,J}(\boldsymbol{\pi}^X, \boldsymbol{\pi}^Y)$ be a multinomial probability vector with marginal probabilities $\boldsymbol{\pi}^X$ and $\boldsymbol{\pi}^Y$. Let $\boldsymbol{\pi}^I = (\boldsymbol{\pi}^X)^T \boldsymbol{\pi}^Y$ be the independence probability. Let $\boldsymbol{\Sigma}^h(\boldsymbol{\pi}) = (\boldsymbol{\Sigma}_{kl}^h(\boldsymbol{\pi}))$ the asymptotic variance-covariance matrix of $\log \hat{\boldsymbol{\theta}}^h$ with partition given in (4.18). For the local odds ratio, the variance-covariance is defined by the formulas given in Example 4.3.3. Under independence, these formulas do only differ by a multiplicative constant, when changing the row odds ratios ($\boldsymbol{\theta}_i^L \mapsto \boldsymbol{\theta}_{i+1}^L$), i.e. it holds $\boldsymbol{\Sigma}_{ii}^L(\boldsymbol{\pi}^I) = c(i)\boldsymbol{\Sigma}_{i+1i+1}^L(\boldsymbol{\pi}^I)$ for some function c only depending on i, which can be seen by the asymptotic variance-covariances:

$$\frac{\mathbb{V}\text{ar}\,\sqrt{n}\log\hat{\theta}_{ij}^L}{\mathbb{V}\text{ar}\,\sqrt{n}\log\hat{\theta}_{i+1j}^L} \overset{asymp.}{=} \frac{\frac{1}{\pi_{i+}\pi_{+j}} + \frac{1}{\pi_{i+}\pi_{+j+1}} + \frac{1}{\pi_{i+1+}\pi_{+j}} + \frac{1}{\pi_{i+1+}\pi_{+j+1}}}{\frac{1}{\pi_{i+1+}\pi_{+j}} + \frac{1}{\pi_{i+1+}\pi_{+j+1}} + \frac{1}{\pi_{i+2+}\pi_{+j}} + \frac{1}{\pi_{i+2+}\pi_{+j+1}}}$$
$$= \frac{\left(\frac{1}{\pi_{i+}} + \frac{1}{\pi_{i+1+}}\right)\left(\frac{1}{\pi_{+j}} + \frac{1}{\pi_{+j+1}}\right)}{\left(\frac{1}{\pi_{i+1+}} + \frac{1}{\pi_{i+2+}}\right)\left(\frac{1}{\pi_{+j}} + \frac{1}{\pi_{+j+1}}\right)} = \frac{\frac{1}{\pi_{i+}} + \frac{1}{\pi_{i+1+}}}{\frac{1}{\pi_{i+1+}} + \frac{1}{\pi_{i+2+}}} =: c(i).$$

Similar it holds

$$\frac{\text{Cov}\left(\sqrt{n}\log\hat{\theta}_{ij}^L, \sqrt{n}\log\hat{\theta}_{i+1j}^L\right)}{\text{Cov}\left(\sqrt{n}\log\hat{\theta}_{i+1j}^L, \sqrt{n}\log\hat{\theta}_{i+2j}^L\right)} \overset{asymp.}{=} \frac{\frac{1}{\pi_{i+}\pi_{+j}} + \frac{1}{\pi_{i+1+}\pi_{+j}}}{\frac{1}{\pi_{i+1+}\pi_{+j}} + \frac{1}{\pi_{i+2+}\pi_{+j}}}$$
$$= \frac{\frac{1}{\pi_{i+}} + \frac{1}{\pi_{i+1+}}}{\frac{1}{\pi_{i+1+}} + \frac{1}{\pi_{i+2+}}} = c(i).$$

and

$$\frac{\text{Cov}\left(\sqrt{n}\log\hat{\theta}_{ij}^L, \sqrt{n}\log\hat{\theta}_{ij+1}^L\right)}{\text{Cov}\left(\sqrt{n}\log\hat{\theta}_{i+1j}^L, \sqrt{n}\log\hat{\theta}_{i+1j+1}^L\right)} \overset{asymp.}{=} \frac{\frac{1}{\pi_{i+}\pi_{+j+1}} + \frac{1}{\pi_{i+1+}\pi_{+j+1}}}{\frac{1}{\pi_{i+1+}\pi_{+j+1}} + \frac{1}{\pi_{i+2+}\pi_{+j+1}}}$$
$$= \frac{\frac{1}{\pi_{i+}} + \frac{1}{\pi_{i+1+}}}{\frac{1}{\pi_{i+1+}} + \frac{1}{\pi_{i+2+}}} = c(i).$$

When $\hat{\theta}_{ij}^L$ and $\hat{\theta}_{kl}^L$ use mutually exclusive sets of cells, both asymptotic covariances are zero, giving the previously stated relation $\boldsymbol{\Sigma}_{ii}^L(\boldsymbol{\pi}^I) = c(i)\boldsymbol{\Sigma}_{i+1i+1}^L(\boldsymbol{\pi}^I)$. The variance-optimal weights become

$$\boldsymbol{\omega}^{(i)}(I) = \frac{\left(\boldsymbol{\Sigma}_{ii}^L(\boldsymbol{\pi}^I)\right)^{-1}\mathbf{1}}{\mathbf{1}^T\left(\boldsymbol{\Sigma}_{ii}^L(\boldsymbol{\pi}^I)\right)^{-1}\mathbf{1}} = \frac{\left(c(i)\boldsymbol{\Sigma}_{i+1i+1}^L(\boldsymbol{\pi}^I)\right)^{-1}\mathbf{1}}{\mathbf{1}^T\left(c(i)\boldsymbol{\Sigma}_{i+1i+1}^L(\boldsymbol{\pi}^I)\right)^{-1}\mathbf{1}}$$
$$= \frac{(c(i))^{-1}}{(c(i))^{-1}}\frac{\left(\boldsymbol{\Sigma}_{i+1i+1}^L(\boldsymbol{\pi}^I)\right)^{-1}\mathbf{1}}{\mathbf{1}^T\left(\boldsymbol{\Sigma}_{i+1i+1}^L(\boldsymbol{\pi}^I)\right)^{-1}\mathbf{1}} = \boldsymbol{\omega}^{(i+1)}(I),$$

for all $i = 1, ..., I-2$. □

4.5.2 The Role of the Scores in Association Models

Let $\{\mu_i\}$ and $\{\nu_j\}$ be scores for the row and column variables in a contingency table. The local R-model, $\log\theta_{ij}^L = c_{1i}, i = 1, ..., I-1, j = 1, ..., J-1$, can be reparameterized under known column scores $\{\nu_j\}$ and is often given in the following parameterization:

$$\log\theta_{ij}^L = \varphi(\mu_{i+1} - \mu_i)(\nu_{j+1} - \nu_j), \qquad i = 1, ..., I-1, j = 1, ..., J-1. \tag{4.22}$$

Assuming $\phi \geq 0$, $\sum_{i=1}^{I} \mu_i = 0$ and $\sum_{i=1}^{I} \mu_i^2 = 1$ (or the marginal weights constraints (1.56)), the parameters become unique and the model is identifiable. Notice that a negative association parameter φ can be encoded in the reordering of the scores. The scores are used to value the stochastic ordering of the underlying classification variables (cf. Agresti [2013, pp. 391]):

Lemma 4.5.4. Assume $\nu_{j+1} > \nu_j$ for all $j = 1, ..., J-1$. If $\mu_i > \mu_k$ in model (4.22), the conditional distribution $Y|X = i$ is stochastically higher than the distribution of $Y|X = k$, e.g. the probability of a higher Y response category is higher in row i than in row k. Especially, for $\mu_i = \mu_k$, row i and row k have the same conditional distribution.

Furthermore, if $\mu_i \leq \mu_{i+1}$ for all $i = 1, ..., I-1$ with $\mu_1 < \mu_I$, then $\theta_{ij}^L \geq 1$ for all $i = 1, ..., I-1, j = 1, ..., J-1$ with at least one strict inequality, which implies stochastic ordering of the conditional distribution in the rows.

Proof. Use the definition of likelihood ratio ordering (Definition 1.9.3) for two discrete variables and $\theta_{ij}^L = \exp(\varphi(\mu_{i+1} - \mu_i)(\nu_{j+1} - \nu_j)), i = 1, ..., I-1, j = 1, ..., J-1.$ □

This result was originally proved by Goodman [1981]. Another reference in the context of association models is Kateri [2014, Ch. 6, p. 165-166]. Remark the clear connection to Lemma 1.9.4, giving the relation between likelihood-ratio ordering and $\theta_{ij}^L \geq 1$ is used in the proof.

The classical R-Model (4.22) can be extended by using the generalised odds ratios

$$\log \theta_{ij}^h = \varphi(\mu_{i+1} - \mu_i)(\nu_{j+1} - \nu_j), \qquad i = 1, ..., I-1, j = 1, ..., J-1, \tag{4.23}$$

which is called *generalised row-effect model.* Following the ideas behind Lemma 4.5.4, the following extensions to $h = CO, C, G$ odds ratios can be made:

Lemma 4.5.5. Assume $\nu_{j+1} > \nu_j$ for all $j = 1, ..., J-1$. If $\mu_{i+1} > \mu_i$ ($\mu_{i+1} < \mu_i$) in model (4.23), it holds

(i) $\frac{\mathbf{P}(Y=j \mid X=i)}{\mathbf{P}(Y>j \mid X=i)}$ is nonincreasing (nondecreasing) when i is changed to $i+1$ for $h = CO$

(ii) $\mathbf{P}(Y \leq j \mid X = i)$ is nonincreasing (nondecreasing) when i is changed to $i+1$ for $h = C$

(iii) $\mathbf{P}(Y \leq j \mid X \leq i) \geq \mathbf{P}(Y \leq j \mid X > i)$ resp. $\mathbf{P}(Y \leq j \mid X \leq i) \leq \mathbf{P}(Y \leq j \mid X > i)$ for $h = G$.

Proof. Since by the choice of the scores $\theta_{ij}^h \geq 1$, the results follow by direct calculation:

(i)

$$1 \leq \theta_{ij}^{CO} = \frac{\mathbf{P}(Y = j \mid X = i)\mathbf{P}(Y > j \mid X = i+1)}{\mathbf{P}(Y = j \mid X = i+1)\mathbf{P}(Y > j \mid X = i)}$$
$$\Leftrightarrow \quad \frac{\mathbf{P}(Y = j \mid X = i)}{\mathbf{P}(Y > j \mid X = i)} \geq \frac{\mathbf{P}(Y = j \mid X = i+1)}{\mathbf{P}(Y > j \mid X = i+1)}$$

(ii)

$$
\begin{aligned}
& 1 \leq \theta_{ij}^{C} = \frac{\mathbf{P}(Y \leq j \mid X = i)}{1 - \mathbf{P}(Y \leq j \mid X = i)} \cdot \frac{1 - \mathbf{P}(Y \leq j \mid X = i+1)}{\mathbf{P}(Y \leq j \mid X = i+1)} \\
\Leftrightarrow \quad & \frac{\mathbf{P}(Y \leq j \mid X = i)}{1 - \mathbf{P}(Y \leq j \mid X = i)} \geq \frac{\mathbf{P}(Y \leq j \mid X = i+1)}{1 - \mathbf{P}(Y \leq j \mid X = i+1)} \\
\Leftrightarrow \quad & \mathbf{P}(Y \leq j \mid X = i) \geq \mathbf{P}(Y \leq j \mid X = i+1)
\end{aligned}
$$

(iii)

$$
\begin{aligned}
& 1 \leq \theta_{ij}^{G} = \frac{\mathbf{P}(Y \leq j \mid X \leq i)}{1 - \mathbf{P}(Y \leq j \mid X \leq i)} \cdot \frac{1 - \mathbf{P}(Y \leq j \mid X > i)}{\mathbf{P}(Y \leq j \mid X > i)} \\
\Leftrightarrow \quad & \frac{\mathbf{P}(Y \leq j \mid X \leq i)}{1 - \mathbf{P}(Y \leq j \mid X \leq i)} \geq \frac{\mathbf{P}(Y \leq j \mid X > i)}{1 - \mathbf{P}(Y \leq j \mid X > i)} \\
\Leftrightarrow \quad & \mathbf{P}(Y \leq j \mid X \leq i) \geq \mathbf{P}(Y \leq j \mid X > i)
\end{aligned}
$$

For decreasing row scores ($\mu_{i+1} > \mu_i, i = 1, ..., I-1$), it holds $\theta_{ij}^h \leq 1$ for all $i = 1, ..., I-1, j = 1, ..., J-1$, and the same calculations can be done by changing the inequalities. □

The interpretation of Lemma 4.5.5 is simple: For increasing scores and $h = CO$ or $h = C$, being in column category j (or below), a change from row category i to row category $i+1$ does not increase the probability to lie in a lower column category than j (or below), reflecting the positive association induced by monotone scores ($\mu_{i+1} > \mu_i$). The same holds for $h = G$ in the global interpretation: Changing the row category i or smaller to greater than i does not increase the probability to lie in a lower column category than j.

Model (4.23) is a reparametrization of $\log \theta_{ij}^h = c_{1i}, i = 1, ..., I-1, j = 1, ..., J-1$, which can be obtained by solving the LSE

$$
\tilde{\mu}_{i+1} - \tilde{\mu}_i = \frac{c_{1i}}{\nu_{j+1} - \nu_j}, \qquad i = 1, ..., I-1, \qquad \sum_{i=1}^{I} \tilde{\mu}_i = 0. \tag{4.24}
$$

Setting $\tilde{\boldsymbol{\mu}} = (\tilde{\mu}_i)$, $\varphi = \sqrt{\sum_{i=1}^{I} \tilde{\mu}_i^2}$ and choosing $\boldsymbol{\mu} = \frac{\tilde{\boldsymbol{\mu}}}{\varphi}$ gives the normalized row scores for any choice of column scores $\{\nu_j\}$ with $\nu_j \neq \nu_{j+1}, j = 1, ..., J-1$.

The scores can be used to interpret the R-model estimates. After calculating the R-model parameter $\{c_{1i}\}$, for example using fitting methods or the presented closed-form estimators, the normalized scores $\{\mu_j\}$ can be calculated.

4.6 Quality of R-model-based Estimates

The R-WM and R-MH estimates are studied for the local and global R-model using the previously selected data sets. Weight estimates are independence based ($\boldsymbol{\omega} = \boldsymbol{\omega}(I)$) and saturated based ($\boldsymbol{\omega} = \boldsymbol{\omega}(S)$). The results are given in Table 4.15. Notice that the R-model for the streptococcus data (Table 4.2) reduces to the uniform association model and is thus removed from the following study.

	Local					Global				
	MLE	WM (S)	WM (I)	MH (S)	MH (I)	MLE	WM (S)	WM (I)	MH (S)	MH (I)
Cannabis - Table 4.3										
$\hat{c}_{11}$	0.771	0.761	0.779	0.761	0.746	2.108	2.096	2.202	2.093	2.155
$\hat{c}_{12}$	0.750	0.731	0.741	0.730	0.722	1.828	1.823	1.874	1.819	1.858
$\hat{c}_{13}$	0.840	0.841	0.858	0.798	0.744	1.865	1.900	1.885	1.899	1.887
Sexual Fun - Table 4.4										
$\hat{c}_{11}$	0.667	0.597	0.565	0.513	0.787	1.614	1.717	1.564	1.679	1.594
$\hat{c}_{12}$	0.339	0.346	0.347	0.279	0.347	1.138	1.138	1.141	1.099	1.133
$\hat{c}_{13}$	-0.046	-0.048	-0.044	-0.103	-0.080	0.710	0.674	0.739	0.636	0.717
Job Satisfaction - Table 4.5										
$\hat{c}_{11}$	-0.071	-	-0.021	-0.117	-0.092	0.313	0.334	0.348	0.328	0.344
$\hat{c}_{12}$	0.178	-	0.154	0.029	0.211	0.472	0.503	0.510	0.497	0.508
$\hat{c}_{13}$	0.669	-	∞	-	0.702	0.895	-	∞	-	1.002
Welfare Spending - Table 4.6										
$\hat{c}_{11}$	0.114	0.110	0.108	0.081	0.111	0.156	0.163	0.151	0.148	0.155
$\hat{c}_{12}$	-0.126	-0.129	-0.129	-0.136	-0.127	-0.144	-0.138	-0.143	-0.148	-0.141
Intellectual Capacity Girls - Table 4.7										
$\hat{c}_{11}$	-0.043	-0.046	-0.067	-0.092	-0.010	0.794	0.820	0.833	0.817	0.824
$\hat{c}_{12}$	0.333	0.341	0.341	0.336	0.339	1.144	1.165	1.197	1.162	1.184
$\hat{c}_{13}$	-0.053	-0.054	-0.051	-0.055	-0.057	0.892	0.922	0.934	0.918	0.930
$\hat{c}_{14}$	0.321	0.316	0.319	0.312	0.316	1.194	1.241	1.273	1.235	1.261
$\hat{c}_{15}$	0.315	0.304	0.343	0.285	0.303	1.347	1.397	1.562	1.390	1.478

Table 4.15: Estimates for R-WM and R-MH for the local and global R-model on log-scale. Weights are calculated variance optimal based on the independence (I) or saturated (S) model.

The difference between maximum likelihood estimation and R-WM resp. R-MH is low for most data classes. Whereas estimates using $\boldsymbol{\omega}(S)$ fail for sampling zeros, and the weighted mean (WM) estimation can give infinite estimates, the MH (I) is always defined, but has a higher divergence compared to the other estimators. Analytical formulas are available for both weight estimation types (S and I), by just inserting $\hat{\boldsymbol{\pi}}$ resp. $\hat{\boldsymbol{\pi}}^I$ in the variance-covariance matrix. Overall, the closed-form estimators have a small divergence from the original MLE and are a good alternative regarding the mathematical effort of iterative fitting algorithms.

The newly defined estimators, R-WM and R-MH, can now be used to estimate the scores in the R-model by solving (4.24), which can be done using the R function `get.scores` in Appendix B.4. The results are given in Table 4.16. The estimates are robust regarding the estimation type and give similar results as the maximum likelihood estimation. Again, for sampling zeros, WM and MH (S) estimation fail as some R-model odds ratios are undefined or infinite. The scores for the local and global R-model are of similar value, but the association parameter φ is greater for the global odds ratio.

4.6.1 Simulation Study

A simulation study for the presented closed-form estimators has also been conducted, similar as in the case of the uniform association (Section 4.4). The results are not given here, but the closed-form R-association model estimates show the same effect as in the uniform case: The R-MH (S) estimator underestimates the true parameter for the job satisfaction and cannabis data. The WM (S) and WM (I) estimators lie quite close to the MLE. In the case of sampling zeros, weighted mean estimators with single cell correction

	Local					Global				
	MLE	WM (S)	WM (I)	MH (S)	MH (I)	MLE	WM (S)	WM (I)	MH (S)	MH (I)
	Cannabis - Table 4.3									
$\hat{\mu}_1$	-0.664	-0.663	-0.664	-0.668	-0.673	-0.688	-0.686	-0.692	-0.686	-0.689
$\hat{\mu}_2$	-0.224	-0.223	-0.221	-0.220	-0.219	-0.198	-0.200	-0.194	-0.200	-0.197
$\hat{\mu}_3$	0.204	0.200	0.199	0.209	0.220	0.227	0.223	0.230	0.222	0.228
$\hat{\mu}_4$	0.684	0.686	0.686	0.679	0.672	0.660	0.663	0.656	0.663	0.659
$\hat{\varphi}$	1.752	1.730	1.761	1.699	1.645	4.303	4.313	4.421	4.307	4.375
	Sexual Fun - Table 4.4									
$\hat{\mu}_1$	-0.819	-0.810	-0.806	-0.818	-0.830	-0.748	-0.758	-0.742	-0.760	-0.746
$\hat{\mu}_2$	0.011	-0.016	-0.029	0.024	0.048	-0.131	-0.115	-0.140	-0.112	-0.134
$\hat{\mu}_3$	0.433	0.445	0.448	0.482	0.435	0.304	0.310	0.299	0.313	0.302
$\hat{\mu}_4$	0.376	0.381	0.387	0.313	0.346	0.575	0.563	0.583	0.559	0.578
$\hat{\varphi}$	0.804	0.751	0.727	0.609	0.897	2.616	2.673	2.598	2.589	2.601
	Job Satisfaction - Table 4.5									
$\hat{\mu}_1$	-0.301	-	-	-	-0.295	-0.548	-	-	-	-0.545
$\hat{\mu}_2$	-0.407	-	-	-	-0.423	-0.301	-	-	-	-0.299
$\hat{\mu}_3$	-0.143	-	-	-	-0.130	0.071	-	-	-	0.064
$\hat{\mu}_4$	0.851	-	-	-	0.847	0.777	-	-	-	0.781
$\hat{\varphi}$	0.674	-	-	-	0.719	1.267	-	-	-	1.398
	Welfare Spending - Table 4.6									
$\hat{\mu}_1$	-0.350	-0.308	-0.298	-0.090	-0.323	-0.457	-0.504	-0.441	-0.407	-0.465
$\hat{\mu}_2$	0.814	0.809	0.807	0.748	0.811	0.814	0.808	0.816	0.816	0.814
$\hat{\mu}_3$	-0.464	-0.501	-0.509	-0.658	-0.488	-0.358	-0.304	-0.375	-0.409	-0.349
$\hat{\varphi}$	0.098	0.099	0.098	0.096	0.098	0.122	0.124	0.120	0.121	0.121
	Intellectual Capacity Girls - Table 4.7									
$\hat{\mu}_1$	-0.411	-0.413	-0.389	-0.372	-0.443	-0.555	-0.554	-0.548	-0.554	-0.551
$\hat{\mu}_2$	-0.466	-0.473	-0.474	-0.499	-0.455	-0.378	-0.377	-0.376	-0.377	-0.377
$\hat{\mu}_3$	-0.037	-0.030	-0.045	-0.037	-0.021	-0.124	-0.126	-0.127	-0.125	-0.127
$\hat{\mu}_4$	-0.106	-0.100	-0.108	-0.113	-0.094	0.075	0.073	0.066	0.074	0.070
$\hat{\mu}_5$	0.307	0.311	0.293	0.315	0.312	0.341	0.341	0.331	0.341	0.336
$\hat{\mu}_6$	0.713	0.706	0.724	0.707	0.700	0.641	0.643	0.654	0.642	0.648
$\hat{\varphi}$	0.777	0.770	0.796	0.728	0.780	4.488	4.633	4.822	4.614	4.734

Table 4.16: Row scores based on the R-WM and R-MH estimates for the local and global R-model. Weights for the estimators are calculated variance optimal based on the independence (I) or saturated (S) model. Column scores are chosen equidistant ($\nu_{j+1} - \nu_j = 1, j = 1, ..., J-1$).

seem to give a good alternative to maximum likelihood estimation.

4.7 Non-iterative Goodness-of-fit Tests

The presented closed-form estimators give in general similar results as the MLE. However, application of these model-based estimators is only useful, if the underlying model fits the data well. Otherwise, their application can be misleading. Since the iterative methods also give single cell frequency estimates under the model, their fit can be valued with the ordinary goodness-of-fit tests, like Pearson's X^2 (1.10), the likelihood ratio statistic G^2 (1.11) or the Cressie-Read statistic $CR_\lambda, \lambda \neq -1, 0$ (1.12) from Section 1.5. Since the aim of the closed-form estimators is to avoid iterative procedures, a method to calculate cell frequency estimates under the model parameter is required.

Let $\log \hat{\boldsymbol{\theta}}_0^h$ be the estimate for the generalised log-odds ratios ($h = L, CO, C, G$). Let $\hat{\boldsymbol{\pi}}^X$ and $\hat{\boldsymbol{\pi}}^Y$ be the sample estimates for the marginal probabilities. As known by the results of Section 5.3, knowledge of $\{\log \hat{\boldsymbol{\theta}}^h, \hat{\boldsymbol{\pi}}^X, \hat{\boldsymbol{\pi}}^Y\}$ is equivalent to the knowledge of $\hat{\boldsymbol{\pi}}_0$, the probability estimate under the model. An inversion function $F(\log \hat{\boldsymbol{\theta}}_0^G, \hat{\boldsymbol{\pi}}^X, \hat{\boldsymbol{\pi}}^Y) = \hat{\boldsymbol{\pi}}_0$ for the global log-odds ratios is available and given in (5.32). The other cases have to be solved numerically following the techniques presented in the Appendix A.1.

Table 4.17 shows the values of the likelihood ratio statistic together with the p-value for the local/global uniform/row-effect association model based on inversion of the closed-form model estimators, which can be calculated using the R function `get.global.U.fits` in Appendix B.4. In addition, the R functions `get.local.U.fits`, `get.global.R.fits` and `get.local.R.fits` can be used to calculate the goodness-of-fit for the local uniform association model and global/local row-effect association model, respectively. It turns out that the different estimation types do not influence the outcome of the goodness-of-fit tests substantially with the exception of the BDNI and MNI estimators in the case of the cannabis data, which do not induce a fitting local uniform association model and supports the results of Section 4.4, which do not suggest the use of BDNI or MNI estimators. Note also that the BDNI estimator is an estimator for the intrinsic association parameter in the correlation model (1.77), which does not hold for the cannabis data, which will be shown later (cf. Table 6.2).

4.8 Discussion

This chapter used association models to introduce new association measures that are linked to association models. Focus was on measures based on the uniform and row-effect association. The newly introduced measures can be regarded as a generalisation of the results of Clayton [1974]. Next to theory and examples, some extensive simulation studies were presented for the uniform association, which also were compared the non-iterative estimators of Beh and Farver [2009] to give adequate suggestions for application. These results have then been extended to row-effect association. Overall, weighted mean estimation using single cell correction with independence based scores ($\boldsymbol{\omega} = \boldsymbol{\omega}(I)$) is an adequate closed-form replacement for the MLE and iterative fitting processes, giving values close to the true parameters in local or global association based on the uniform or row-effect model. In a symmetrical manner, estimation for the C model can be developed.

	Strepto.		Cannabis		Sexual Fun		Job Satisf.		Welfare Spending		IC Girls	
	n=1398		n=1054		n=91		n=96		n=955		n=8310	
	2×3		4×3		4×4		4×4		3×5		6×7	
	Local uniform association											
	G^2	p-val	G^2	p-val	G^2	p-val	G^2	p-val	G^2	p-val	G^2	p-val
MLE	0.24	0.63	1.47	0.92	5.01	0.76	3.72	0.88	10.21	0.18	201.51	0.00
WM (S)	0.24	0.63	1.47	0.92	5.01	0.76	-	-	10.23	0.18	202.86	0.00
WM (I)	0.25	0.62	1.54	0.91	5.04	0.75	-	-	10.23	0.18	205.38	0.00
MH (S)	0.24	0.62	1.57	0.91	6.14	0.63	-	-	10.52	0.16	239.60	0.00
MH (I)	0.24	0.63	2.21	0.82	5.01	0.76	3.73	0.88	10.24	0.18	201.53	0.00
logNI	0.25	0.62	1.54	0.91	5.04	0.75	-	-	10.23	0.18	205.38	0.00
BDNI	0.24	0.62	10.96	0.05	5.05	0.75	3.72	0.88	10.21	0.18	213.27	0.00
MNI	0.26	0.61	8.40	0.14	5.11	0.75	3.85	0.87	10.23	0.18	204.37	0.00
	Global uniform association											
	G^2	p-val	G^2	p-val	G^2	p-val	G^2	p-val	G^2	p-val	G^2	p-val
MLE	0.30	0.58	6.04	0.30	5.85	0.66	3.93	0.86	10.34	0.17	232.59	0.00
WM (S)	0.30	0.58	6.04	0.30	5.85	0.66	-	-	10.34	0.17	232.69	0.00
WM (I)	0.32	0.57	6.36	0.27	5.85	0.66	-	-	10.34	0.17	237.69	0.00
MH (S)	0.30	0.58	6.04	0.30	6.11	0.64	-	-	10.38	0.17	232.58	0.00
MH (I)	0.32	0.57	6.20	0.29	5.95	0.65	3.94	0.86	10.40	0.17	234.10	0.00
	Local row-effect association											
	G^2	p-val	G^2	p-val	G^2	p-val	G^2	p-val	G^2	p-val	G^2	p-val
MLE	0.24	0.63	1.30	0.73	2.27	0.89	1.77	0.94	5.57	0.47	28.65	0.28
WM (S)	0.24	0.63	1.31	0.73	2.33	0.89	-	-	5.58	0.47	29.04	0.26
WM (I)	0.25	0.62	1.32	0.72	2.39	0.88	-	-	5.59	0.47	30.21	0.22
MH (S)	0.24	0.62	1.47	0.69	3.10	0.80	-	-	5.99	0.42	33.30	0.12
MH (I)	0.24	0.63	2.14	0.54	2.44	0.88	1.80	0.94	5.58	0.47	30.21	0.22
	Global row-effect association											
	G^2	p-val	G^2	p-val	G^2	p-val	G^2	p-val	G^2	p-val	G^2	p-val
MLE	0.30	0.58	5.26	0.15	2.62	0.85	2.76	0.84	5.53	0.48	90.42	0.00
WM (S)	0.30	0.58	5.31	0.15	2.72	0.84	-	-	5.53	0.48	92.08	0.00
WM (I)	0.32	0.57	5.33	0.15	2.65	0.85	-	-	5.53	0.48	106.82	0.00
MH (S)	0.30	0.58	5.31	0.15	2.73	0.84	-	-	5.53	0.48	91.68	0.00
MH (I)	0.31	0.58	5.29	0.15	2.63	0.85	2.81	0.83	5.53	0.48	96.91	0.00

Table 4.17: Values of the likelihood ratio test statistic G^2 with p-values for the closed-form estimators of the uniform and row-effect association model parameter. Model fit has been calculated using the inversion algorithm presented in Appendix A.1 for the local or using the formula (5.32) for the global association, respectively.

Chapter 5

ϕ-Divergence based Generalized Odds Ratios

Association in contingency tables has already been debated in Section 1.6. As pointed out, describing exactly the association structure in an $I \times J$ table requires $(I-1)(J-1)$ parameters. The generalised odds ratios (Section 1.9) are such a multidimensional parameter set. As seen in Chapter 4, the generalised log-odds ratios are the basis of new mutlidimensional measures of association, which are based on generalised association models. In this Chapter, the information-theoretic approach of Chapter 3 is extended to define generalised ϕ-scaled odds ratios, which are directly linked to the extension of the local ϕ-association model of Kateri and Papaioannou [1995].

The generalised odds ratios on ϕ-scale is extended in Section 5.1, leading to generalised ϕ-scaled odds ratios. The ϕ-scaling inherits most of the properties of the generalised log-odds ratios. Independence between both classification variables in a two-way tables holds, if and only if all generalised log-odds ratios are zero (cf. (1.39)). This property is generalised to ϕ-scale in Section 5.3. Association between two classification variables is not encoded within the marginal distribution since the marginals do not give additional information for the connection between one variable to the other. Therefore knowledge of the marginal distribution and a parameter set of size $(I-1)(J-1)$, that encodes the association, should be equivalent to knowledge of the joint distribution, which is the case for the local odds ratios (Remark 1.9.1). Such parameter sets are called *object of association* (Osius [2004]). This property extends - under certain conditions - to ϕ-scale. The asymptotic distribution of the generalised ϕ-scaled odds ratios is presented in Section 5.4. The generalised odds ratios are closely related to dependence concepts (Section 1.9.2). For example, the local odds ratios correspond with likelihood ratio ordering. The generalised ϕ-scaled odds ratios loose this property, which is discussion in Section 5.5. Examples are given in Section 5.6, while the results of this chapter are summarized in Section 5.7.

5.1 The Generalized ϕ-scaled Odds Ratios

The generalised odds ratios (Section 1.9) are a very basic and well-known multidimensional measure of association in categorical data analysis. They occur naturally as parameters in loglinear models and are used for interpretation (Section 1.9.4).

The ϕ-divergence between two discrete finite bivariate distributions $\mathbf{p} = (p_{ij})$ and $\mathbf{q} = (q_{ij})$ it is defined Definition 1.68.

In chapter 3 the log-odds ratio for 2×2 tables was extended using the ϕ-divergence. In the following, the scale change approach will be studied for the generalised odds ratios in $I \times J$ tables.

Let $\boldsymbol{\pi} = (\pi_{ij}) \in \Delta_{I,J}$ be a multinomial probability vector with marginal distribution $\boldsymbol{\pi}^X \in \Delta_I$ and $\boldsymbol{\pi}^Y \in \Delta_J$ for the two classification variables X and Y, respectively. Let $\boldsymbol{\pi}^I = \boldsymbol{\pi}^X(\boldsymbol{\pi}^Y)^T = (\pi_{i+}\pi_{+j})$ be the probability vector under independence (1.2) and let $a_{ij}^h, b_{ij}^h, c_{ij}^h$ and d_{ij}^h be the essential sums defined in (1.41). There are two ways to motivate the definition of generalised ϕ-scaled odds ratios. First it holds

$$\begin{aligned}
\log\theta_{ij}^h(\boldsymbol{\pi}) &= \log\theta_{ij}^h(\boldsymbol{\pi}) - \log\theta_{ij}^h(\boldsymbol{\pi}^I) \\
&= \log\left(\frac{a_{ij}^h(\boldsymbol{\pi})d_{ij}^h(\boldsymbol{\pi})}{b_{ij}^h(\boldsymbol{\pi})c_{ij}^h(\boldsymbol{\pi})}\right) - \log\left(\frac{a_{ij}^h(\boldsymbol{\pi}^I)d_{ij}^h(\boldsymbol{\pi}^I)}{b_{ij}^h(\boldsymbol{\pi}^I)c_{ij}^h(\boldsymbol{\pi}^I)}\right) \\
&= \log\left(\frac{a_{ij}^h(\boldsymbol{\pi})}{a_{ij}^h(\boldsymbol{\pi}^I)}\right) - \log\left(\frac{b_{ij}^h(\boldsymbol{\pi})}{b_{ij}^h(\boldsymbol{\pi}^I)}\right) - \log\left(\frac{c_{ij}^h(\boldsymbol{\pi})}{c_{ij}^h(\boldsymbol{\pi}^I)}\right) + \log\left(\frac{d_{ij}^h(\boldsymbol{\pi})}{d_{ij}^h(\boldsymbol{\pi}^I)}\right), \quad (5.1)
\end{aligned}$$

for $h = L, CO, C, G$. Therefore, the generalised log-odds ratio measure the divergence of the essential sums (1.41) from their hypothetical values under independence. The terms

$$\frac{a_{ij}^h(\boldsymbol{\pi})}{a_{ij}^h(\boldsymbol{\pi}^I)}, \qquad \frac{b_{ij}^h(\boldsymbol{\pi})}{b_{ij}^h(\boldsymbol{\pi}^I)}, \qquad \frac{c_{ij}^h(\boldsymbol{\pi})}{c_{ij}^h(\boldsymbol{\pi}^I)} \qquad \text{and} \qquad \frac{d_{ij}^h(\boldsymbol{\pi})}{d_{ij}^h(\boldsymbol{\pi}^I)} \qquad (5.2)$$

can be regarded as *generalised association factors*, which for the 2×2 case were already used by Good [1956]. The generalised log-odds ratio weight the generalised association factors on log-scale. Using the Kullback-Leibler divergence (1.70), the generalised log-odds ratio (5.1) can also be written as

$$\log\theta_{ij}^h(\boldsymbol{\pi}) = \phi'_{KL}\left(\frac{a_{ij}^h(\boldsymbol{\pi})}{a_{ij}^h(\boldsymbol{\pi}^I)}\right) - \phi'_{KL}\left(\frac{b_{ij}^h(\boldsymbol{\pi})}{b_{ij}^h(\boldsymbol{\pi}^I)}\right) - \phi'_{KL}\left(\frac{c_{ij}^h(\boldsymbol{\pi})}{c_{ij}^h(\boldsymbol{\pi}^I)}\right) + \phi'_{KL}\left(\frac{d_{ij}^h(\boldsymbol{\pi})}{d_{ij}^h(\boldsymbol{\pi}^I)}\right),$$

Replacing the Kullback-Leibler divergence by a general $\phi \in \Phi$, the ϕ-scaled odds ratios arise:

Definition 5.1.1. Let $\phi \in \Phi$ be differentiable and $\boldsymbol{\pi} = (\pi_{ij})$ a non-degenerate $I \times J$ multinomial probability vector, i.e. $\pi_{ij} > 0$ $(i = 1, ..., I, j = 1, ..., J)$. The *generalised ϕ-scaled odds ratio* $(h = L, CO, C, G)$ is defined as

$$\boldsymbol{\theta}^{\phi,h}(\boldsymbol{\pi}) = (\theta_{11}^{\phi,h}(\boldsymbol{\pi}), ..., \theta_{I-1,J-1}^{\phi,h}(\boldsymbol{\pi})), \qquad \text{where} \qquad (5.3)$$

$$\theta_{ij}^{\phi,h}(\boldsymbol{\pi}) = \phi'\left(\frac{a_{ij}^h(\boldsymbol{\pi})}{a_{ij}^h(\boldsymbol{\pi}^I)}\right) - \phi'\left(\frac{b_{ij}^h(\boldsymbol{\pi})}{b_{ij}^h(\boldsymbol{\pi}^I)}\right) - \phi'\left(\frac{c_{ij}^h(\boldsymbol{\pi})}{c_{ij}^h(\boldsymbol{\pi}^I)}\right) + \phi'\left(\frac{d_{ij}^h(\boldsymbol{\pi})}{d_{ij}^h(\boldsymbol{\pi}^I)}\right), \qquad (5.4)$$

for $i = 1, ..., I-1, j = 1, ..., J-1$. $\boldsymbol{\theta}^{\phi,h}(\boldsymbol{\pi})$ is called local ϕ-scaled odds ratio $(h = L)$, continuation ϕ-scaled odds ratio $(h = CO)$, cumulative ϕ-scaled odds ratio $(h = C)$ or global ϕ-scaled odds ratio $(h = G)$.

As in the 2×2 case, the generalised ϕ-scaled odds ratio measure the divergence from independence on the scale ϕ by weighting the generalised association factors in terms of ϕ'. As already seen in the motivation, they coincide with the generalised log-odds ratio in the case of the Kullback-Leibler divergence.

Another motivation for the introduction of generalised ϕ-scaled odds ratios goes back to ϕ-association models (1.80), introduced in Section 1.11. In the same way the local log-odds ratio occur when dealing with log-scaled association models, the local ϕ-scaled odds ratio occur when dealing with ϕ-association models as shown in (1.81). Thus, there is a clear connection between ϕ-scaled odds ratio and ϕ-association models. Using the generalised essential sums ($h = C, CO, G$) the generalised ϕ-scaled odds ratios are derived.

The generalised ϕ-scaled odds ratio estimates are calculated using the R functions

- `local.por.phi` for the local ϕ-scaled odds ratio ($h = L$),
- `continuation.por.phi` for the continuation ϕ-scaled odds ratio ($h = L$),
- `cumulative.por.phi` for the cumulative ϕ-scaled odds ratio ($h = L$) and
- `global.por.phi` for the global ϕ-scaled odds ratio ($h = L$)

in Appendix B.5.

5.1.1 Properties

The generalised log-odds ratio are very compatible in terms of table transformation (rotation, change of rows or columns, etc.). Their behaviour under transformation on ϕ-scale will be derived:

Proposition 5.1.2. Let $\phi \in \Phi$ be differentiable and $\boldsymbol{\pi}$ be an $I \times J$ multinomial probability. The transformation properties for the generalised odds ratios of Proposition 1.9.2 extend to ϕ-scale.

The estimator for the generalised ϕ-scaled odds ratios $\boldsymbol{\theta}^{\phi,h}(\boldsymbol{\pi})$ is given by $\boldsymbol{\theta}^{\phi,h}(\hat{\boldsymbol{\pi}})$. When sampling zeros occur, it is possible that some of the association factors (5.2) are estimated as zero. When $\phi'(0) = -\infty$, $\boldsymbol{\theta}^{\phi,h}(\hat{\boldsymbol{\pi}})$ is either undefined or has infinite components. On the other hand, the ϕ-scaled odds ratio is well-defined, even when sampling zeros occur, for ϕ-divergence functions with $\phi'(0) > -\infty$.

For the 2×2 ϕ-scaled odds ratio (Chapter 3) it already turned out that infinite estimates can be prevented using an adequate scale change, leading to bounded ϕ-scaled odds ratios (Section 3.3). This result can be extended to $\boldsymbol{\theta}^{\phi,h}$:

Lemma 5.1.3. Let $\phi \in \Phi$ be differentiable. Then $\boldsymbol{\theta}^{\phi,h}$ is bounded on the subspace of fixed marginals if and only if

$$\phi'(0) = \lim_{t \searrow 0} \phi'(t) > -\infty.$$

Proof. Let $\Delta := \Delta_{I,J}(\boldsymbol{\pi}^X, \boldsymbol{\pi}^Y)$ be the set of $I \times J$ multinomial probability vectors with marginal distribution $\boldsymbol{\pi}^X$ for X and $\boldsymbol{\pi}^Y$ for Y with $\pi_{ij} > 0$. Let $\overline{\Delta}$ be the closure of Δ allowing $\pi_{ij} = 0$.

Let $\boldsymbol{\pi}^0 = (\pi_{ij}^0) \in \overline{\Delta}$ be such that $\pi_{11}^0 = 0$ and $\pi_{ij} > 0$ for $(i,j) \neq (1,1)$. Since $\overline{\Delta}$ is the closure of Δ, there exists a sequence $(\boldsymbol{\pi}^{(n)})_n \subseteq \Delta$ such that $\boldsymbol{\pi}^{(n)} \to \boldsymbol{\pi}^0$ for $n \to \infty$. Especially, $(a_{11}^h(\boldsymbol{\pi}^{(n)}), b_{11}^h(\boldsymbol{\pi}^{(n)}), c_{11}^h(\boldsymbol{\pi}^{(n)}), d_{11}^h(\boldsymbol{\pi}^{(n)}))_n \to (0, a_1, a_2, a_3)$ for some $a_1, a_2, a_3 > 0$. Then $\theta_{11}^{\phi,h}(\boldsymbol{\pi}^{(n)}) \to \infty$ and $\boldsymbol{\theta}^{\phi,h}$ is unbounded.

Now let $\phi'(0) > -\infty$. Under fixed marginals the essential sums $a_{ij}^h(\boldsymbol{\pi}^I), b_{ij}^h(\boldsymbol{\pi}^I), c_{ij}^h(\boldsymbol{\pi}^I)$ and $d_{ij}^h(\boldsymbol{\pi}^I)$ (cf. (1.41)) are constant in $\overline{\Delta}$. Set $m := \max_{i,j}(a_{ij}^h(\boldsymbol{\pi}^I)^{-1})$. It holds

$$0 \leq \frac{a_{ij}^h(\boldsymbol{\pi})}{a_{ij}^h(\boldsymbol{\pi}^I)} \leq \frac{1}{a_{ij}^h(\boldsymbol{\pi}^I)} \leq m < \infty, \qquad \text{for all } \boldsymbol{\pi} \in \overline{\Delta}.$$

Using the convexity of ϕ, the derivative ϕ' is monotonic increasing:

$$-\infty < \phi'(0) \leq \phi'\left(\frac{a_{ij}^h(\boldsymbol{\pi})}{a_{ij}^h(\boldsymbol{\pi}^I)}\right) \leq \phi'\left(\frac{1}{a_{ij}^h(\boldsymbol{\pi}^I)}\right) \leq \phi'(m) < \infty$$

for $i = 1, ..., I, j = 1, ..., J, \boldsymbol{\pi} \in \overline{\Delta}$. The same holds for the other generalised association factors. Thus $\boldsymbol{\theta}^{\phi,h}(\boldsymbol{\pi})$ is finite for all $\boldsymbol{\pi} \in \overline{\Delta}$ as a sum of finite elements. Since $\overline{\Delta}$ is closed, $\boldsymbol{\theta}^{\phi,h}$ is bounded. □

5.2 ϕ-based Association Models

Let $\boldsymbol{\mu} = (\mu_i)$ and $\boldsymbol{\nu} = (\nu_j)$ be the scores of the row and column classification variables in an $I \times J$ table. The classical association model (1.76) and the correlation model (1.77) have already been introduced in Section 1.11. These models can be generalised by a scale change using the ϕ-divergence, which leads to the ϕ-association model (1.80). This model reduces to the ϕ-scaled uniform association model (U_ϕ), when both sets of scores are known, and it reduces to the ϕ-scaled row-effect model R_ϕ (column-effect model C_ϕ) model, when $\boldsymbol{\mu} = (\mu_i)$ ($\boldsymbol{\nu} = (\nu_j)$) are known. When both sets of scores are unknown, the RC_ϕ model is obtained. The classical association models use the log-scale, which can be applied by using the Kullback-Leibler divergence (1.70).

These models can further be extended. The classic RC model leaves $(I-2)(J-2)$ degrees of freedom in the model and will possibly not cover more complex association structures within an $I \times J$ table. More interaction terms can be added to the model, leading - on log-scale - to the RC(K) model:

$$\log \pi_{ij} = \lambda + \lambda_i^X + \lambda_j^Y + \sum_{k=1}^{K} \varphi_k \mu_{ik} \nu_{jk}, \quad i = 1, ..., I-1, j = 1, ..., J-1, \tag{5.5}$$

which can be constructed using a singular value decomposition of the interaction terms $\boldsymbol{\Lambda} = (\lambda_{ij}^{XY})$ in model (1.46) (cf. Kateri [2014, Section 6.5]). The $\boldsymbol{\varphi} = (\varphi_k)$ result as the eigenvalues of the decomposition, and $\boldsymbol{\mu}_k = (\mu_{1k}, ..., \mu_{Ik})$ and $\boldsymbol{\nu}_k = (\nu_{1k}, ..., \nu_{Jk})$ are the eigenvectors, $k = 1, ..., K$. Since the maximum rank of $\boldsymbol{\Lambda}$ is $K^* = \min(I, J) - 1$, model (5.5) can be defined for $0 \leq K \leq K^*$, where RC(K*) is the saturated model. The RC(K) model has $(I-K-1)(J-K-1)$ degrees of freedom.

The local log-odds ratio can be derived from (5.5) as

$$\log \theta_{ij}^{L} = \sum_{k=1}^{K} \varphi_k(\mu_{ik} - \mu_{i+1k})(\nu_{jk} - \nu_{j+1k}), \tag{5.6}$$

for $i = 1, ..., I-1, j = 1, ..., J-1$ and is used for physical interpretation of the model. Replacing the local log-odds ratio in model (5.6) by the generalised ϕ-scaled odds ratio, the generalised ϕ-scaled RC(K) model is derived:

$$\theta_{ij}^{\phi,h} = \sum_{k=1}^{K} \varphi_k(\mu_{ik} - \mu_{i+1k})(\nu_{jk} - \nu_{j+1k}), \qquad h = L, CO, C, G, \tag{5.7}$$

for $i = 1, ..., I-1, j = 1, ..., J-1$, which is denoted as $\mathrm{RC}_{\phi}^{h}(\mathrm{K})$. In the case of $h = L$, model $\mathrm{RC}_{\phi}^{L}(\mathrm{K})$ has already been mentioned in Kateri and Papaioannou [1995], which can be expressed in terms of the multinomial probabilities:

$$\pi_{ij} = \pi_{i+}\pi_{+j}(\phi')^{-1}\left(\alpha_i + \beta_j + \sum_{k=1}^{K} \varphi_k \mu_{ik}\nu_{jk}\right), \qquad i = 1, ..., I, j = 1, ..., J. \tag{5.8}$$

5.3 Association Measurement

The generalised log-odds ratio (Section 1.9) are the golden standard in association analysis of $I \times J$ contingency tables and in the case of local odds ratios are directly linked to classical association models (Section 1.9.4). The newly introduced generalised ϕ-scaled odds ratios $\boldsymbol{\theta}^{\phi,h}$ are based on weighting the generalised association factors (5.2) using a ϕ-divergence function $\phi \in \Phi$. Therefore, association measurement is directly linked to the measurement of divergence from independence.

Let $\boldsymbol{\pi} \in \Delta_{I,J}$ be a multinomial probability vector with marginal distributions $\boldsymbol{\pi}^X$ for the row classification variable X and $\boldsymbol{\pi}^Y$ for the column classification variable Y. The generalised ϕ-scaled odds ratios should inherit property (1.39) from the generalised log-odds ratios, i.e.

$$\boldsymbol{\theta}^{\phi,h}(\boldsymbol{\pi}) = \mathbf{0} \qquad \Leftrightarrow \qquad X \text{ and } Y \text{ are independent, i.e. } \boldsymbol{\pi} = (\boldsymbol{\pi}^X)^T\boldsymbol{\pi}^Y. \tag{5.9}$$

Property (5.9) is a direct consequence of the following result.

Theorem 5.3.1. Let $\boldsymbol{\pi}$ and $\tilde{\boldsymbol{\pi}}$ be two $I \times J$ probability tables with positive entries, having the same marginal probabilities $\boldsymbol{\pi}^X = \tilde{\boldsymbol{\pi}}^X$ and $\boldsymbol{\pi}^Y = \tilde{\boldsymbol{\pi}}^Y$ for the row and column classification variable, respectively. Then for any strictly convex $\phi \in \Phi$ it holds

$$\boldsymbol{\theta}^{\phi,L}(\boldsymbol{\pi}) = \boldsymbol{\theta}^{\phi,L}(\tilde{\boldsymbol{\pi}}) \qquad \Leftrightarrow \qquad \boldsymbol{\pi} = \tilde{\boldsymbol{\pi}}.$$

Proof. The part $\boldsymbol{\pi} = \tilde{\boldsymbol{\pi}} \Rightarrow \boldsymbol{\theta}^{\phi,h}(\boldsymbol{\pi}) = \boldsymbol{\theta}^{\phi,h}(\tilde{\boldsymbol{\pi}})$ is obvious. Start with the case $h = L$: Set $\boldsymbol{\theta}^{\phi,L} = \boldsymbol{\theta}^{\phi,L}(\boldsymbol{\pi})$ and $\tilde{\boldsymbol{\theta}}^{\phi,L} = \boldsymbol{\theta}^{\phi,L}(\tilde{\boldsymbol{\pi}})$. The multinomial probabilities can be expressed using the saturated model $\mathrm{RC}_{\phi}^{L}(\mathrm{K}^*)$, $\mathrm{K}^* = \min(I, J) - 1$ given in (5.8). Then it holds

$$\phi'\left(\frac{\pi_{ij}}{\pi_{i+}\pi_{+j}}\right) = \alpha_i + \beta_j + \gamma_{ij}, \quad \text{where} \quad \gamma_{ij} = \sum_{k=1}^{K^*} \varphi_k \mu_{ik}\nu_{jk},$$

with the scores satisfying

$$\sum_{i=1}^{I}\pi_{i+}\mu_{ik} = \sum_{j=1}^{J}\pi_{+j}\nu_{jk} = 0, \qquad k = 1, ..., K^*$$

$$\sum_{i=1}^{I}\pi_{i+}\mu_{ik}\mu_{ik'} = \sum_{j=1}^{J}\pi_{+j}\nu_{jk}\nu_{jk'} = \mathbf{1}_{\{k=k'\}}, \qquad k, k' = 1, ..., K^*,$$

and

$$\theta_{ij}^{\phi,L} = \gamma_{ij} - \gamma_{ij+1} - \gamma_{i+1j} + \gamma_{i+1j+1}, \qquad i = 1, ..., I-1, j = 1, ..., J-1. \tag{5.10}$$

Similarly it holds

$$\tilde{\theta}_{ij}^{\phi,L} = \tilde{\gamma}_{ij} - \tilde{\gamma}_{ij+1} - \tilde{\gamma}_{i+1j} + \tilde{\gamma}_{i+1j+1}, \qquad i = 1, ..., I-1, j = 1, ..., J-1.$$

Hence, $\boldsymbol{\theta}^{\phi,L}$ is a function only of $\boldsymbol{\varphi} = (\varphi_k), \boldsymbol{\mu} = (\mu_{ik})$ and $\boldsymbol{\nu} = (\nu_{jk})$. It holds $\boldsymbol{\theta}^{\phi,L} = \mathbf{C}^L\boldsymbol{\gamma}$, where $\boldsymbol{\gamma} = (\gamma_{ij})$ is a $IJ \times 1$ vector with entries γ_{ij}, expanded by rows, and $\mathbf{C}^L$ is the $(I-1)(J-1) \times IJ$ design matrix for producing $\log \boldsymbol{\theta}^L$ as given in (1.43). Hence

$$\boldsymbol{\theta}^{\phi,L} = \tilde{\boldsymbol{\theta}}^{\phi,L} \Rightarrow \mathbf{C}^L\boldsymbol{\gamma} = \mathbf{C}^L\tilde{\boldsymbol{\gamma}} \Rightarrow \mathbf{C}^L(\boldsymbol{\gamma} - \tilde{\boldsymbol{\gamma}}) = \mathbf{0}. \tag{5.11}$$

It can be proved that $\text{rank}(\mathbf{C}^h) = I + J - 1$ and the kernel of $\mathbf{C}^L$ is

$$\mathbf{K}_{\mathbf{C}^L} = \{\mathbf{k}_1, \mathbf{k}_2, ..., \mathbf{k}_{I+J-1}\}$$

with

$$\mathbf{k}_1 = (\underbrace{k_{11}, ..., k_{11}}_{(I-1)\text{-times}}, k_{12}), \quad \text{where} \quad k_{11} = (\underbrace{-1, ..., -1}_{(J-1)\text{-times}}, 0), \quad k_{12} = (\underbrace{0, ..., 0}_{(J-1)\text{-times}}, 1),$$

$$\mathbf{k}_j = (\underbrace{k_{j1}, ..., k_{j1}}_{I\text{-times}}), \quad j = 2, ..., J, \quad \text{with} \quad k_{j1} = (k_{j1}^r, ..., k_{jJ}^r), \quad k_{jl}^r = \mathbf{1}_{\{l=I-j+1\}},$$

$$\mathbf{k}_j = (\underbrace{k_{1j}, ..., k_{Ij}}_{I\text{-times}}), \quad j = J+1, ..., J+1-1,$$

$$\text{with} \quad k_{lj} = (\underbrace{k_{lj}^c, ..., k_{lj}^c}_{J\text{-times}}), \quad k_{lj}^c = \mathbf{1}_{\{l=I+J-j\}}.$$

Hence, from (5.11) it follows $\boldsymbol{\gamma} - \tilde{\boldsymbol{\gamma}} = \mathbf{k}_p, p = 1, ..., I+J-1$ or $\boldsymbol{\gamma} = \tilde{\boldsymbol{\gamma}}$. Recall that

$$\pi_{ij} = \pi_{i+}\pi_{+j}(\phi')^{-1}(\alpha_i + \beta_j + \gamma_{ij}) \tag{5.12}$$

$$\tilde{\pi}_{ij} = \pi_{i+}\pi_{+j}(\phi')^{-1}\left(\tilde{\alpha}_i + \tilde{\beta}_j + \tilde{\gamma}_{ij}\right), \tag{5.13}$$

for $i = 1, ..., I, j = 1, ..., J$. In the case of $\boldsymbol{\gamma} - \tilde{\boldsymbol{\gamma}} = \mathbf{k}_p, p = 2, ..., J$, note that $\gamma_{ij} = \tilde{\gamma}_{ij}, i = 1, ..., I, j \neq J - p + 1 =: j_p$ and $\gamma_{ij_p} = \tilde{\gamma}_{ij_p} + 1$. Thus for the solutions $\mathbf{k}_p, p = 2, ..., J$, it holds:

$$\gamma_{ij} = \tilde{\gamma}_{ij} + \mathbf{1}_{\{j=J-p+1\}},$$

and it holds

$$\beta_j + \gamma_{ij} = \underbrace{\beta_j + \mathbf{1}_{\{j=J-p+1\}}}_{=:\beta_j^*} + \tilde{\gamma}_{ij}, \qquad i = 1, ..., I, j = 1, ..., J-1.$$

Similarly for the solutions $\mathbf{k}_p, p = J+1, ..., J+I-1$:

$$\gamma_{ij} = \tilde{\gamma}_{ij} + \mathbf{1}_{\{i=I+J-p\}}, \qquad i = 1, ..., I, j = 1, ..., J-1$$

and it holds

$$a_i + \gamma_{ij} = \underbrace{a_i + \mathbf{1}_{\{i=I+J-p\}}}_{=:a_i^*}, \qquad i = 1, ..., I-1, j = 1, ..., J.$$

Setting further $\beta_J^* = \beta_J$ and $\alpha_I^* = \alpha_I$ and considering overall

$$\begin{aligned} \boldsymbol{\alpha} &= (\alpha_1^*, ..., \alpha_I^*), && \text{for solutions } p = 2, ..., J \\ \boldsymbol{\beta} &= (\beta_1^*, ..., \beta_J^*), && \text{for solutions } p = J+1, ..., I+J-1 \end{aligned}$$

and $\boldsymbol{\alpha}^* = \boldsymbol{\alpha}, \boldsymbol{\beta}^* = \boldsymbol{\beta}$ in the case $\boldsymbol{\gamma} = \tilde{\boldsymbol{\gamma}}$, it follows from (5.12) that

$$\pi_{ij} = \pi_{i+}\pi_{+j}(\phi')^{-1}\left(\alpha_i^* + \beta_j^* + \tilde{\gamma}_{ij}\right). \tag{5.14}$$

For models (5.12) and (5.13) the following constraints hold:

$$\sum_{j=1}^{J} \pi_{+j}(\phi')^{-1}\left(\alpha_i^* + \beta_j^* + \tilde{\gamma}_{ij}\right) = \sum_{j=1}^{J} \pi_{+j}(\phi')^{-1}\left(\tilde{\alpha}_i + \tilde{\beta}_j + \tilde{\gamma}_{ij}\right) = 1, \quad i = 1, ..., I.$$

such that

$$\sum_{j=1}^{J} \pi_{+j}[(\phi')^{-1}\left(\alpha_i^* + \beta_j^* + \tilde{\gamma}_{ij}\right) - (\phi')^{-1}\left(\tilde{\alpha}_i + \tilde{\beta}_j + \tilde{\gamma}_{ij}\right)] = 0, \quad i = 1, ..., I. \tag{5.15}$$

Similarly, one can conclude

$$\sum_{i=1}^{I} \pi_{i+}\underbrace{[(\phi')^{-1}\left(\alpha_i^* + \beta_j^* + \tilde{\gamma}_{ij}\right) - (\phi')^{-1}\left(\tilde{\alpha}_i + \tilde{\beta}_j + \tilde{\gamma}_{ij}\right)]}_{=:\Delta_{ij}} = 0, \quad j = 1, ..., J. \tag{5.16}$$

In order of (5.15) and (5.16) to hold, the difference Δ_{ij} must be either $\Delta_{ij} = 0, i = 1, ..., I, j = 1, ..., J$, or change sign at least once over j (for (5.15)) and over i (for (5.16)). Changing sign for (5.15) means that for all $i = 1, ..., I$, there exist $j_{1(i)}$ and $j_{2(i)}$ such that

$$(\phi')^{-1}\left(\alpha_i^* + \beta_{j_{1(i)}}^* + \tilde{\gamma}_{ij_{1(i)}}\right) < (\phi')^{-1}\left(\tilde{\alpha}_i^* + \tilde{\beta}_{j_{1(i)}}^* + \tilde{\gamma}_{ij_{1(i)}}\right)$$

and

$$(\phi')^{-1}\left(\alpha_i^* + \beta_{j_{2(i)}}^* + \tilde{\gamma}_{ij_{2(i)}}\right) > (\phi')^{-1}\left(\tilde{\alpha}_i^* + \tilde{\beta}_{j_{2(i)}}^* + \tilde{\gamma}_{ij_{2(i)}}\right),$$

such that by the strictly monotony of ϕ' (ϕ is strictly convex) it holds

$$\beta_{j_{1(i)}}^* < \tilde{\beta}_{j_{1(i)}} + (\tilde{\alpha}_i - \alpha_i^*) \qquad \text{and} \qquad \beta_{j_{2(i)}}^* > \tilde{\beta}_{j_{2(i)}} + (\tilde{\alpha}_i - \alpha_i^*).$$

It follows for all $i = 1, ..., I$,

$$\beta_{j_{1(i)}}^* - \tilde{\beta}_{j_{1(i)}} < \tilde{\alpha}_i - \alpha_i^* < \beta_{j_{2(i)}}^* - \tilde{\beta}_{j_{2(i)}}.$$

Let $j_1, j_2 \in \{1, ..., J\}$ be the columns with

$$\beta^*_{j_1} - \tilde{\beta}_{j_1} = \min_{i=1,\dots,I} (\beta^*_{j_{1(i)}} - \tilde{\beta}_{j_{1(i)}}) \quad \text{and} \quad \beta^*_{j_2} - \tilde{\beta}_{j_2} = \max_{i=1,\dots,I} (\beta^*_{j_{2(i)}} - \tilde{\beta}_{j_{2(i)}}).$$

Then for $i = 1, ..., I$ it holds

$$\beta^*_{j_1} - \tilde{\beta}_{j_1} < \tilde{\alpha}_i - \alpha^*_i < \beta^*_{j_2} - \tilde{\beta}_{j_2}. \tag{5.17}$$

Similarly, based on (5.16), for all $j = 1, ..., J$, there exist $i_{1(j)}$ and $i_{2(j)}$ such that

$$(\phi')^{-1}\left(\alpha^*_{i_{1(j)}} + \beta^*_j + \tilde{\gamma}_{i_{1(j)}j}\right) < (\phi')^{-1}\left(\tilde{\alpha}_{i_{1(j)}} + \tilde{\beta}_j + \tilde{\gamma}_{i_{1(j)}j}\right)$$

and

$$(\phi')^{-1}\left(\alpha^*_{i_{2(j)}} + \beta^*_j + \tilde{\gamma}_{i_{2(j)}j}\right) > (\phi')^{-1}\left(\tilde{\alpha}_{i_{2(j)}} + \tilde{\beta}_j + \tilde{\gamma}_{i_{2(j)}j}\right),$$

such that

$$\tilde{\alpha}_{i_{2(j)}} - \alpha^*_{i_2} < \beta^*_j - \tilde{\beta}_j < \tilde{\alpha}_{i_1} - \alpha^*_{i_1}, \qquad j = 1, ..., J. \tag{5.18}$$

Let $i_1, i_2 \in \{1, ..., I\}$ be the rows such that

$$\tilde{\alpha}_{i_2} - \alpha^*_{i_2} = \min_{j=1,\dots,J}\left(\tilde{\alpha}_{i_{2(j)}} - \alpha_{i^*_{2(j)}}\right) \quad \text{and} \quad \tilde{\alpha}_{i_1} - \alpha^*_{i_1} = \max_{j=1,\dots,J}\left(\tilde{\alpha}_{i_{1(j)}} - \alpha_{i^*_{1(j)}}\right).$$

Then from (5.18) it follows for $j = j_2$:

$$\begin{aligned} & \tilde{\alpha}_{i_1} - \alpha^*_{i_1} > \beta^*_{j_2} - \tilde{\beta}_{j_2} \\ \Rightarrow \quad & \tilde{\alpha}_{i_1} + \tilde{\beta}_{j_2} + \tilde{\gamma}_{i_1j_2} > \alpha^*_{i_1} + \beta^*_{j_2} + \tilde{\gamma}_{i_1j_2} \\ \Rightarrow \quad & \tilde{\pi}_{i_1j_2} > \pi_{i_1j_2} \end{aligned}$$

But as a consequence of (5.17) it holds:

$$\begin{aligned} & \tilde{\alpha}_{i_1} - \alpha^*_{i_1} < \beta^*_{j_2} - \tilde{\beta}_{j_2} \\ \Rightarrow \quad & \tilde{\alpha}_{i_1} + \tilde{\beta}_{j_2} + \tilde{\gamma}_{i_1j_2} < \alpha^*_{i_1} + \beta^*_{j_2} + \tilde{\gamma}_{i_1j_2} \\ \Rightarrow \quad & \tilde{\pi}_{i_1j_2} < \pi_{i_1j_2}, \end{aligned}$$

which is a contradiction. Hence Δ_{ij} cannot change sign and it holds $\Delta_{ij} = 0$. Thus

$$(\phi')^{-1}\left(\alpha^*_i + \beta^*_j + \tilde{\gamma}_{ij}\right) = (\phi')^{-1}\left(\tilde{\alpha}_i + \tilde{\beta}_j + \tilde{\gamma}_{ij}\right) \Rightarrow \pi_{ij} = \tilde{\pi}_{ij},\ i = 1, ..., I, j = 1, ..., J,$$

when $\boldsymbol{\gamma} = \tilde{\boldsymbol{\gamma}}$ or $\boldsymbol{\gamma} - \tilde{\boldsymbol{\gamma}} = \mathbf{k}_p, p = 2, ..., I + J - 1$.

In the case $\boldsymbol{\gamma} - \tilde{\boldsymbol{\gamma}} = \mathbf{k}_1$, it holds

$$\begin{aligned} \gamma_{ij} &= \tilde{\gamma}_{ij} - 1, && i = 1, ..., I-1, j = 1, ..., J-1, \\ \gamma_{Ij} &= \tilde{\gamma}_{Ij}, && j = 1, ..., J-1, \qquad \text{and} \\ \gamma_{IJ} &= \tilde{\gamma}_{IJ} + 1. \end{aligned}$$

Hence setting $\alpha^*_i = \alpha_i - 1, i = 1, ..., I$ and $\beta^*_j = \beta_j, j = 1, ..., J-1$ and $\beta^*_J = \beta_J + 1$, it follows

$$\pi_{ij} = \pi_{i+}\pi_{+j}(\phi')^{-1}\left(\alpha^*_i + \beta^*_j + \tilde{\gamma}_{ij} + \mathbf{1}_{\{i=I\}}\mathbf{1}_{\{j=J\}}\right), \quad i = 1, ..., I, j = 1, ..., J.$$

For $i = 1, ..., I-1$ it has exactly the same form as previously. Hence for the column j_3 defined so that $\beta^*_{j_3} - \tilde{\beta}_{j_3} = \max_{i=1,...,I-1}\left(\beta^*_{j_{2(i)}} - \tilde{\beta}_{j_{2(i)}}\right)$, it holds from (5.17) using again the monotony of ϕ' for $i = 1, ..., I-1$:

$$\begin{aligned} & \tilde{\alpha}_i - \alpha^*_i < \beta^*_{j_3} - \tilde{\beta}_{j_3} \\ \Rightarrow\quad & (\phi')^{-1}\left(\tilde{\alpha}_i + \tilde{\beta}_{j_3} + \tilde{\gamma}_{ij_3}\right) < (\phi')^{-1}\left(\alpha^*_i + \beta^*_{j_3} + \tilde{\gamma}_{ij_3}\right) \\ \Rightarrow\quad & \tilde{\alpha}_i + \tilde{\beta}_{j_3} + \tilde{\gamma}_{ij_3} < \alpha^*_i + \beta^*_{j_3} + \tilde{\gamma}_{ij_3} \\ \Rightarrow\quad & \tilde{\pi}_{ij_3} < \pi_{ij_3}. \end{aligned}$$

Consider j_4 to be the column for which $\beta^*_{j_4} - \tilde{\beta}_{j_4} = \min_{i=1,...,I-1}\left(\beta^*_{j_{1(i)}} - \tilde{\beta}_{j_{1(i)}}\right)$. Then for all $i = 1, ..., I-1$ it holds

$$\begin{aligned} & \beta^*_{j_4} - \tilde{\beta}_{j_4} < \tilde{\alpha}_i - \alpha^*_i \\ \Rightarrow\quad & \alpha^*_i + \beta^*_{j_4} < \tilde{\alpha}_i + \tilde{\beta}_{j_4} \\ \Rightarrow\quad & \pi_{ij_4} < \tilde{\pi}_{ij_4}. \end{aligned}$$

Thus

$$\tilde{\pi}_{ij_3} < \pi_{ij_3} \quad\Rightarrow\quad \phi'\left(\frac{\tilde{\pi}_{ij_3}}{\pi_{i+}\pi_{+j_3}}\right) < \phi'\left(\frac{\pi_{ij_3}}{\pi_{i+}\pi_{+j_3}}\right) \tag{5.19}$$

$$\pi_{ij_4} < \tilde{\pi}_{ij_4} \quad\Rightarrow\quad \phi'\left(\frac{\tilde{\pi}_{ij_4}}{\pi_{i+}\pi_{+j_4}}\right) > \phi'\left(\frac{\pi_{ij_4}}{\pi_{i+}\pi_{+j_4}}\right), \tag{5.20}$$

while

$$\begin{aligned} \tilde{\pi}_{ij_3} < \pi_{ij_3}, \quad i = 1, ..., I-1 \quad & \Rightarrow \sum_{i=1}^{I-1}\tilde{\pi}_{ij_3} < \sum_{i=1}^{I-1}\pi_{ij_3} \\ & \overset{\pi_{i+}=\tilde{\pi}_{i+}}{\Rightarrow} \tilde{\pi}_{Ij_3} > \pi_{Ij_3} \\ & \Rightarrow \phi'\left(\frac{\tilde{\pi}_{Ij_3}}{\pi_{I+}\pi_{+j_3}}\right) > \phi'\left(\frac{\pi_{Ij_3}}{\pi_{I+}\pi_{+j_3}}\right) \end{aligned} \tag{5.21}$$

and

$$\begin{aligned} \pi_{ij_4} < \tilde{\pi}_{ij_4}, \quad i = 1, ..., I-1 \quad & \Rightarrow \sum_{i=1}^{I-1}\pi_{ij_4} > \sum_{i=1}^{I-1}\tilde{\pi}_{ij_4} \\ & \overset{\pi_{i+}=\tilde{\pi}_{i+}}{\Rightarrow} \pi_{Ij_4} > \tilde{\pi}_{Ij_4} \\ & \Rightarrow \phi'\left(\frac{\tilde{\pi}_{Ij_4}}{\pi_{I+}\pi_{+j_4}}\right) < \phi'\left(\frac{\pi_{Ij_4}}{\pi_{I+}\pi_{+j_4}}\right). \end{aligned} \tag{5.22}$$

Consider the ϕ-scaled odds ratios θ^ϕ_i and $\tilde{\theta}^\phi_i, i = 1, ..., I-1$, formed by rows i and I, and columns j_3 and j_4: Then

$$(5.19) \quad\Rightarrow\quad \phi'\left(\frac{\tilde{\pi}_{ij_3}}{\pi_{i+}\pi_{+j_3}}\right) < \phi'\left(\frac{\pi_{ij_3}}{\pi_{i+}\pi_{+j_3}}\right)$$

$$\begin{aligned}
(5.22) \quad &\Rightarrow \quad \phi'\left(\frac{\tilde{\pi}_{Ij_4}}{\pi_{I+}\pi_{+j_4}}\right) < \phi'\left(\frac{\pi_{Ij_4}}{\pi_{I+}\pi_{+j_4}}\right)\\
(5.20) \quad &\Rightarrow \quad -\phi'\left(\frac{\tilde{\pi}_{ij_4}}{\pi_{i+}\pi_{+j_4}}\right) < -\phi'\left(\frac{\pi_{ij_4}}{\pi_{i+}\pi_{+j_4}}\right)\\
(5.21) \quad &\Rightarrow \quad -\phi'\left(\frac{\tilde{\pi}_{Ij_3}}{\pi_{I+}\pi_{+j_3}}\right) < -\phi'\left(\frac{\pi_{Ij_3}}{\pi_{I+}\pi_{+j_3}}\right)
\end{aligned}$$

Summing the previous inequalities gives $\tilde{\theta}_i^{\phi} < \theta_i^{\phi}$, which contradicts $\tilde{\boldsymbol{\theta}}^{\phi,L} = \boldsymbol{\theta}^{\phi,L}$. Hence also for solution $\mathbf{k}_1$, it holds $\Delta_{ij} = 0, i = 1, ..., I, j = 1, ..., J$ and thus $\tilde{\boldsymbol{\pi}} = \boldsymbol{\pi}$, which completes the proof in the case $h = L$. □

Remark 5.3.2. Theorem 5.3.1 can be extended to the global ϕ-scaled odds ratios as well (see Corollary 5.3.3) based on a property of the global ϕ-scaled odds ratios, exposed in the following subsection.

5.3.1 Global ϕ-scaled Odds Ratios and Their Reduction Property

The global ϕ-scaled odds ratio ($h = G$) can be reduced to the ordinary ϕ-scaled odds ratio θ^{ϕ} (Chapter 3) in 2×2 tables. For $\boldsymbol{\pi} = (\pi_{ij}) \in \Delta_{I,J}$, let $\mathbb{T}(i,j)$ be the 2×2 matrix

$$\mathbb{T}(i,j) = \begin{pmatrix} \mathbb{T}_{11}(i,j) & \mathbb{T}_{12}(i,j) \\ \mathbb{T}_{21}(i,j) & \mathbb{T}_{22}(i,j) \end{pmatrix} = \begin{pmatrix} a_{ij}^G(\boldsymbol{\pi}) & b_{ij}^G(\boldsymbol{\pi}) \\ c_{ij}^G(\boldsymbol{\pi}) & d_{ij}^G(\boldsymbol{\pi}) \end{pmatrix}, \tag{5.23}$$

where $a_{ij}^h, b_{ij}^h, c_{ij}^h$ and d_{ij}^h are the essential sums (1.41). Since

$$a_{ij}^G(\boldsymbol{\pi}) + b_{ij}^G(\boldsymbol{\pi}) + c_{ij}^G(\boldsymbol{\pi}) + d_{ij}^G(\boldsymbol{\pi}) = \sum_{a,b=1}^{I,J} \pi_{ab} = 1, \qquad i = 1, ..., I-1, j = 1, ..., J-1,$$

and $\mathbb{T}_{kl}(i,j) > 0$ for $k, l = 1, 2$, it holds $\mathbb{T}(i,j) = (\mathbb{T}_{kl}(i,j)) \in \Delta_{2\times 2}$, which is a 2×2 multinomial probability matrix. Let $\mathbb{T}_{k+}(i,j)$ and $\mathbb{T}_{+l}(i,j)$, be the marginal distributions of $\mathbb{T}(i,j)$. It holds

$$\begin{aligned}
\mathbb{T}_{1+}(i,j)\mathbb{T}_{+1}(i,j) &= (a_{ij}^G(\boldsymbol{\pi}) + b_{ij}^G(\boldsymbol{\pi}))(a_{ij}^G(\boldsymbol{\pi}) + c_{ij}^G(\boldsymbol{\pi}))\\
&= \left(\sum_{k\le i}\sum_{l\le j}\pi_{kl} + \sum_{k\le i}\sum_{l>j}\pi_{kl}\right)\left(\sum_{k\le i}\sum_{l\le j}\pi_{kl} + \sum_{k>i}\sum_{l\le j}\pi_{kl}\right) \qquad (5.24)\\
&= \left(\sum_{k\le i}\pi_{k+}\right)\left(\sum_{l\le j}\pi_{+l}\right) = \sum_{k\le i}\sum_{l\le j}\pi_{k+}\pi_{+l} = a_{ij}^G(\boldsymbol{\pi}^I).
\end{aligned}$$

Similar it holds

$$b_{ij}^G(\boldsymbol{\pi}^I) = \mathbb{T}_{1+}(i,j)\mathbb{T}_{+2}(i,j), \quad c_{ij}^G(\boldsymbol{\pi}^I) = \mathbb{T}_{2+}(i,j)\mathbb{T}_{+1}(i,j), \quad d_{ij}^G(\boldsymbol{\pi}^I) = \mathbb{T}_{2+}(i,j)\mathbb{T}_{+2}(i,j), \tag{5.25}$$

for $i = 1, ..., I-1, j = 1, ..., J-1$. Using the 2×2 ϕ-scaled odds ratio θ^{ϕ} from (3.1) the following identity, called *reduction property of the global odds ratio*, holds:

$$\theta_{ij}^{\phi,G}(\boldsymbol{\pi}) = \theta^{\phi}(\mathbb{T}(i,j)), \qquad i = 1, ..., I-1, j = 1, ..., J-1 \tag{5.26}$$

since

$$\begin{aligned}\theta_{ij}^{\phi,G}(\boldsymbol{\pi}) &= \phi'\left(\frac{a_{ij}^h(\boldsymbol{\pi})}{a_{ij}^h(\boldsymbol{\pi}^I)}\right) - \phi'\left(\frac{b_{ij}^h(\boldsymbol{\pi})}{b_{ij}^h(\boldsymbol{\pi}^I)}\right) - \phi'\left(\frac{c_{ij}^h(\boldsymbol{\pi})}{c_{ij}^h(\boldsymbol{\pi}^I)}\right) + \phi'\left(\frac{d_{ij}^h(\boldsymbol{\pi})}{d_{ij}^h(\boldsymbol{\pi}^I)}\right) \qquad (5.27)\\ &= \phi'\left(\frac{\mathbb{T}_{11}(i,j)}{\mathbb{T}_{1+}(i,j)\mathbb{T}_{+1}(i,j)}\right) - \phi'\left(\frac{\mathbb{T}_{12}(i,j)}{\mathbb{T}_{1+}(i,j)\mathbb{T}_{+2}(i,j)}\right)\\ &\quad - \phi'\left(\frac{\mathbb{T}_{21}(i,j)}{\mathbb{T}_{2+}(i,j)\mathbb{T}_{+1}(i,j)}\right) + \phi'\left(\frac{\mathbb{T}_{22}(i,j)}{\mathbb{T}_{2+}(i,j)\mathbb{T}_{+2}(i,j)}\right)\\ &= \theta^{\phi}(\mathbb{T}(i,j)).\end{aligned}$$

Theorem 5.3.1 can be extended to the global ϕ-scaled odds ratio as a direct consequence of the reduction property:

Corollary 5.3.3. Let $\boldsymbol{\pi}$ and $\tilde{\boldsymbol{\pi}}$ be two $I \times J$ probability tables with positive entries, having the same marginal probabilities $\boldsymbol{\pi}^X = \tilde{\boldsymbol{\pi}}^X$ and $\boldsymbol{\pi}^Y = \tilde{\boldsymbol{\pi}}^Y$ for the row and column classification variable, respectively. Then for any strictly convex $\phi \in \Phi$ it holds

$$\boldsymbol{\theta}^{\phi,G}(\boldsymbol{\pi}) = \boldsymbol{\theta}^{\phi,G}(\tilde{\boldsymbol{\pi}}) \qquad \Leftrightarrow \qquad \boldsymbol{\pi} = \tilde{\boldsymbol{\pi}}.$$

Proof. Set $\boldsymbol{\theta}_0^{\phi,G} = (\theta_{0,ij}^{\phi,G}) = \boldsymbol{\theta}^{\phi,G}(\tilde{\boldsymbol{\pi}})$. It is enough to show that the equation

$$\boldsymbol{\theta}^{\phi,G}(\boldsymbol{\pi}) = \boldsymbol{\theta}_0^{\phi,G} \qquad (5.28)$$

has a unique solution if the marginal distributions $\boldsymbol{\pi}^X$ and $\boldsymbol{\pi}^Y$ are known. Let $\mathbb{T}(i,j), i = 1,...,I-1, j = 1,...,J-1$ be the 2×2 collapsed subtables based on $\boldsymbol{\pi}$ as in Section 5.3.1. For fixed marginal distribution $\boldsymbol{\pi}^X$ and $\boldsymbol{\pi}^Y$, the marginals $\mathbb{T}_{k+}(i,j)$ and $\mathbb{T}_{+l}(i,j), k,l = 1,2$ are also fixed. Applying (5.26) on (5.28) gives for fixed (i,j):

$$\theta_{0,ij}^{\phi,G} = \theta_{ij}^{\phi,G}(\boldsymbol{\pi}) = \theta^{\phi}(\mathbb{T}(i,j)).$$

Application of Lemma 3.4.1 (invertibility of θ^{ϕ}) gives a unique set of solutions

$$\mathbb{T}^{(0)}(i,j), \qquad i = 1,...,I-1, j = 1,...,J-1, \qquad (5.29)$$

which induce a multinomial probability vector $\boldsymbol{\pi}^{(0)} \in \Delta_{I,J}(\boldsymbol{\pi}^X, \boldsymbol{\pi}^Y)$ by setting

$$\begin{aligned}\pi_{11}^{(0)} &= \mathbb{T}_{11}^{(0)}(1,1)\\ \pi_{i1}^{(0)} &= \mathbb{T}_{11}^{(0)}(i,1) - \mathbb{T}_{11}^{(0)}(i-1,1), \qquad i = 2,...,I-1\\ \pi_{1j}^{(0)} &= \mathbb{T}_{11}^{(0)}(1,j) - \mathbb{T}_{11}^{(0)}(1,j-1), \qquad j = 2,...,J-1 \qquad (5.30)\\ \pi_{ij}^{(0)} &= \mathbb{T}_{11}^{(0)}(i,j) - \mathbb{T}_{11}^{(0)}(i-1,j) - \mathbb{T}_{11}^{(0)}(i,j-1) + \mathbb{T}_{11}^{(0)}(i-1,j-1),\\ &\qquad i = 2,...,I-1, j = 2,...,J-1,\end{aligned}$$

and

$$\pi_{iJ}^{(0)} = \pi_{i+} - \sum_{j=1}^{J-1} \pi_{ij}^{(0)}, \quad i = 1,...,I, \qquad \pi_{Ij}^{(0)} = \pi_{+j} - \sum_{i=1}^{I-1} \pi_{ij}^{(0)}, \quad j = 1,...,J-1. \quad (5.31)$$

Then $\boldsymbol{\pi}^{(0)}$ constructed by (5.30) and (5.31) fulfils (5.28). Since this solution is unique, it follows $\boldsymbol{\pi}^{(0)} = \boldsymbol{\pi} = \tilde{\boldsymbol{\pi}}$. □

Remark 5.3.4. The solution of (5.28) for the global log-odds ratio $\log \boldsymbol{\theta}_0^G = (\log \theta_{0,ij}^G)$ can be calculated analytically. Using $F_{ij} = \mathbf{P}(X \leq i, Y \leq j)$, Dale [1986] already solved the equation (5.28) for the Kullback-Leibler (ϕ_{KL}, Section 1.70) divergence:

$$F_{ij} = F_{ij}(\theta_{0,ij}^G, \eta_i, \xi_j) = \begin{cases} A_{ij} - B_{ij} & \text{for } \exp(\theta_{0,ij}^G) \neq 0 \\ \eta_i \xi_j & \text{for } \exp(\theta_{0,ij}^G) = 0. \end{cases}, \tag{5.32}$$

where $\eta_i = \mathbf{P}(X \leq i), i = 1, ..., I$ and $\eta_j = \mathbf{P}(Y \leq j), j = 1, ..., J$ are the cumulative marginal probabilities of $\boldsymbol{\pi}^X$ and $\boldsymbol{\pi}^Y$, respectively, and

$$A_{ij} = \frac{1 + (\eta_i + \xi_j)(\exp(\theta_{0,ij}^G) - 1)}{2(\exp(\theta_{0,ij}^G) - 1)}$$

$$B_{ij} = \sqrt{\left(\frac{1 + (\eta_i + \xi_j)(\exp(\theta_{0,ij}^G) - 1)}{2(\exp(\theta_{0,ij}^G) - 1)}\right)^2 - \frac{\eta_i \xi_j \exp(\theta_{0,ij}^G)}{1 - \exp(\theta_{0,ij}^G)}}.$$

Notice that the solution (5.32) for the global log-odds ratios is calculated using the R function `invert.global.log.odds` in Appendix B.4.

The solution (5.29) for general $\phi \in \Phi$ is calculated using the R function `get.lambda.ij`, while the $\boldsymbol{\pi}$ as given in (5.30) is then obtained by using `get.pi`. The overall algorithm is summarized in the R function `inv.global.phi` (see Appendix B.5).

5.3.2 Association Objects

Relation (5.9) presented in Section 5.3 characterizes independence in terms of the generalised ϕ-scaled odds ratios. There is a more general characterization that association measure should fulfil, which goes back to Osius [2004], who formally defined association objects $\mathcal{A} : \Delta_{I,J} \to \mathbb{R}^{(I-1)(J-1)}$.

Let $\boldsymbol{\pi}^X \in \Delta_I$ and $\boldsymbol{\pi}^Y \in \Delta_J$ be marginal distributions. The set $\Delta_{I,J}(\boldsymbol{\pi}^X, \boldsymbol{\pi}^Y)$ is defined as the elements $\boldsymbol{\pi} \in \Delta_{I,J}$ with marginal distribution $\boldsymbol{\pi}^X$ for row classification variable X and $\boldsymbol{\pi}^Y$ for column classification variable Y. Clearly, $\Delta_{I,J}(\boldsymbol{\pi}^X, \boldsymbol{\pi}^Y)$ is not empty, since the probability of independence $\boldsymbol{\pi}^I = (\boldsymbol{\pi}^X)^T \boldsymbol{\pi}^Y$ is always included. The basic idea behind Osius' concept is, that the association in a contingency table is not encoded within the marginal distributions. Thus, $\mathcal{A}(\boldsymbol{\pi})$ includes the association structure of the table $\boldsymbol{\pi}$. Formally, knowledge of the triple $(\mathcal{A}(\boldsymbol{\pi}), \boldsymbol{\pi}^X, \boldsymbol{\pi}^Y)$ is equivalent to knowledge of the joint distribution $\boldsymbol{\pi}$.

The object of association property can now formally be introduced:

Definition 5.3.5. The function $\mathcal{A} : \Delta_{I,J} \to \mathbb{R}^{(I-1)(J-1)}$ is an *object of association*, if its restriction $\mathcal{A} : \Delta_{I,J}(\boldsymbol{\pi}^X, \boldsymbol{\pi}^Y) \to \mathbb{R}^{(I-1)(J-1)}$ is invertible on its image for every $\boldsymbol{\pi}^X \in \Delta_I$ and $\boldsymbol{\pi}^Y \in \Delta_J$.

Notice that the condition (5.9) is weaker than the object of association property.

Osius [2004] proved, that the local log-odds ratio are objects of association. The result was already known and proofed by using a sequence $(\boldsymbol{\pi}_n)_n$ of probabilities with given local odds ratio $\boldsymbol{\theta}_0^L$, which converges to the $\boldsymbol{\pi}$ with the desired property (Sinkhorn [1967]) and in an alternative way using a result of Haberman [1974a, Theorem 2.6], who obtains $\boldsymbol{\pi}$ by maximizing a strictly concave function which is linked to the log-likelihood function of Poisson distributions. The object of association property could already been obtained for the ϕ-scaled odds ratios in 2×2 tables using basic analysis (Proposition 3.1.3).

That the generalised ϕ-scaled odds ratios are association objects is a direct consequence of Theorem 5.3.1, since it proved that the function $\mathcal{A}$ is injective and therefore invertible on its image. In the case $h = G$, the reduction property (5.26) can be used to construct an inversion algorithm (cf. Remark 5.3.4).

5.4 Asymptotic Behaviour

The asymptotic normality for the generalised log-odds ratio has already been proven in Theorem 1.9.6. It turns out, that this result can be extended to ϕ-scaled by using the delta method (Theorem 1.4.3):

Theorem 5.4.1. Let $\mathbf{N} \sim \mathcal{M}(n, \boldsymbol{\pi})$ be multinomial distributed with sample size $n \in \mathbb{N}$ and probability vector $\boldsymbol{\pi} = (\pi_{ij}) \in \Delta_{I,J}$. Let $\hat{\boldsymbol{\pi}} = \mathbf{N}/n$ be the MLE of $\boldsymbol{\pi}$ and let $\boldsymbol{\theta}^{\phi,h}(\boldsymbol{\pi})$ be the generalised ϕ-scaled odds ratio (Definition 5.1.1) with MLE $\boldsymbol{\theta}^{\phi,h}(\hat{\boldsymbol{\pi}})$. Assume that $\mathbf{D}^h = \partial\boldsymbol{\theta}^{\phi,h}(\boldsymbol{\pi})/\partial\boldsymbol{\pi} \neq \mathbf{0}$. Then $\boldsymbol{\theta}^{\phi,h}(\hat{\boldsymbol{\pi}})$ is asymptotic normal distributed with variance-covariance matrix

$$\boldsymbol{\Sigma}^{\phi,h}(\boldsymbol{\pi}) = (\mathbf{D}^h)^T \boldsymbol{\Sigma}(\boldsymbol{\pi}) \mathbf{D}^h \qquad \text{for} \qquad h = L, CO, C, G, \tag{5.33}$$

where $\boldsymbol{\Sigma}(\boldsymbol{\pi}) = \operatorname{diag}(\boldsymbol{\pi}) - \boldsymbol{\pi}^T\boldsymbol{\pi}$. In other words, it holds

$$\sqrt{n}(\boldsymbol{\theta}^{\phi,h}(\hat{\boldsymbol{\pi}}) - \boldsymbol{\theta}^{\phi,h}(\boldsymbol{\pi})) \xrightarrow{d} \mathcal{N}(0, \boldsymbol{\Sigma}^{\phi,h}(\boldsymbol{\pi})), \qquad \text{for } n \to \infty.$$

Proof. By Lemma 1.4.2, it holds

$$\sqrt{n}(\hat{\boldsymbol{\pi}} - \boldsymbol{\pi}) \xrightarrow{d} \mathcal{N}(0, \boldsymbol{\Sigma}(\boldsymbol{\pi})), \qquad \text{for } n \to \infty.$$

Since by assumption

$$\mathbf{D}^h = \partial\boldsymbol{\theta}^{\phi,h}(\boldsymbol{\pi})/\partial\boldsymbol{\pi} \neq \mathbf{0}, \tag{5.34}$$

the delta method (Theorem 1.4.3) can be applied. It follows

$$\sqrt{n}(\boldsymbol{\theta}^{\phi,h}(\hat{\boldsymbol{\pi}}) - \boldsymbol{\theta}^{\phi,h}(\boldsymbol{\pi})) \xrightarrow{\mathcal{D}} \mathcal{N}(\mathbf{0}, (\mathbf{D}^h)^T \boldsymbol{\Sigma}(\boldsymbol{\pi}) \mathbf{D}^h) \qquad \text{for } n \to \infty. \tag{5.35}$$

The asymptotic variance-covariance matrix of $\boldsymbol{\theta}^{\phi,h}(\hat{\boldsymbol{\pi}})$ is then $\boldsymbol{\Sigma}^{\phi,h}(\boldsymbol{\pi}) = (\mathbf{D}^h)^T \boldsymbol{\Sigma}(\boldsymbol{\pi}) \mathbf{D}^h$. □

Example 5.4.2. Let $\phi \in \Phi$ be strictly convex. In the case of the global ϕ-scaled odds ratios it can be proven that the Jacobian matrix $\mathbf{D}^G = \partial\boldsymbol{\theta}^{\phi,G}(\boldsymbol{\pi})/\partial\boldsymbol{\pi}$ is invertible for all $\boldsymbol{\pi} \in \Delta_{I,J}$, since $\det(\mathbf{D}^G) = \prod_{i,j=1}^{I-1,J-1} \frac{\partial\theta^\phi(\mathbb{T}(i,j))}{\partial\mathbb{T}_{11}(i,j)} > 0$ for strictly convex $\phi \in \Phi$. The derivation of $\mathbf{D}^G$ in Theorem 5.4.1 can be drawn back to the case of 2×2 tables using the reduction property (5.26), $\theta_{ij}^{\phi,G}(\boldsymbol{\pi}) = \theta^{\phi}(\mathbb{T}(i,j))$. Write $\mathbb{T}(i,j) = (\mathbb{T}_{kl}(i,j))$ as vector. It holds

$$\frac{\partial \mathbb{T}(i_1,j_1)}{\partial \pi_{i_2j_2}} = \begin{cases} (1,0,0,0) & \text{for } i_2 \leq i_1, j_2 \leq j_1 \\ (0,1,0,0) & \text{for } i_2 \leq i_1, j_2 > j_1 \\ (0,0,1,0) & \text{for } i_2 > i_1, j_2 \leq j_1 \\ (0,0,0,1) & \text{for } i_2 > i_1, j_2 > j_1 \end{cases}$$

and

$$\frac{\partial \theta^{\phi}(\mathbb{T}(i,j))}{\partial \mathbb{T}(i,j)} = \left(\frac{\partial \theta^{\phi}(\mathbb{T}(i,j))}{\partial \mathbb{T}_{11}(i,j)}, \frac{\partial \theta^{\phi}(\mathbb{T}(i,j))}{\partial \mathbb{T}_{12}(i,j)}, \frac{\partial \theta^{\phi}(\mathbb{T}(i,j))}{\partial \mathbb{T}_{21}(i,j)}, \frac{\partial \theta^{\phi}(\mathbb{T}(i,j))}{\partial \mathbb{T}_{22}(i,j)} \right).$$

Applying the multidimensional chain rule gives

$$\frac{\partial \theta_{i_1j_1}^{\phi,G}(\boldsymbol{\pi})}{\partial \pi_{i_2j_2}} = \frac{\partial \theta^{\phi}(\mathbb{T}(i_1,j_1))}{\partial \mathbb{T}(i_1,j_1)} \cdot \frac{\partial \mathbb{T}(i_1,j_1)}{\partial \pi_{i_2j_2}} = \begin{cases} \frac{\partial \theta^{\phi}(\mathbb{T}(i_1,j_1))}{\partial \mathbb{T}_{11}(i_1,j_1)} & \text{for } i_2 \leq i_1, j_2 \leq j_1 \\ \frac{\partial \theta^{\phi}(\mathbb{T}(i_1,j_1))}{\partial \mathbb{T}_{12}(i_1,j_1)} & \text{for } i_2 \leq i_1, j_2 > j_1 \\ \frac{\partial \theta^{\phi}(\mathbb{T}(i_1,j_1))}{\partial \mathbb{T}_{21}(i_1,j_1)} & \text{for } i_2 > i_1, j_2 \leq j_1 \\ \frac{\partial \theta^{\phi}(\mathbb{T}(i_1,j_1))}{\partial \mathbb{T}_{22}(i_1,j_1)} & \text{for } i_2 > i_1, j_2 > j_1, \end{cases}$$

where the derivatives (3.5), (3.6), (3.7) and (3.8) for the 2×2 ϕ-scaled odds ratio can be used (cf. prove of Theorem 3.2.1).

Remark 5.4.3. The derivations for $h = L, CO, C$ are available but not given here due to their high technicality. The R functions are given in Appendix B.5 to calculate the partial derivatives $\frac{\partial \theta_{i_1j_1}^{\phi,h}(\boldsymbol{\pi})}{\partial \pi_{i_2j_2}}$ and the Jacobian matrix $\mathbf{D}^h$ from (5.34). These are

- `D.div.local` and `local.derivation.matrix` for $h = L$,
- `D.div.continuation` and `continuation.derivation.matrix` for $h = CO$,
- `D.div.cumulative` and `cumulative.derivation.matrix` for $h = C$ and
- `D.div.global`, and `global.derivation.matrix` for $h = G$.

The variance-covariance estimate $\mathbf{\Sigma}^{\phi,h}(\hat{\boldsymbol{\pi}})$ (cf. 5.33)) is then calculated using the R functions `cov.matrix.local`, `cov.matrix.continuation`, `cov.matrix.cumulative` and `cov.matrix.global` for $h = L, CO, C$ and G, respectively.

Under independence, the asymptotic variance-covariance matrix $\mathbf{\Sigma}^{\phi,h}$ has the special property that a scale change does only effect the value of the variance-covariance matrix by a multiplicative constant but does not influence the structure of the matrix:

Lemma 5.4.4. Let $\boldsymbol{\pi} \in \Delta_{I,J}$ be a multinomial probability vector with marginal probability $\boldsymbol{\pi}^X$ for the row classification variable X and $\boldsymbol{\pi}^Y$ for the column classification variable Y. Let $\boldsymbol{\pi}^I = (\boldsymbol{\pi}^X)^T \boldsymbol{\pi}^Y$ be the probability under independence. Let $\mathbf{\Sigma}^{\phi,h}(\boldsymbol{\pi})$ be the variance-covariance matrix of $\boldsymbol{\theta}^{\phi,h}(\hat{\boldsymbol{\pi}})$ as given in (5.33). It holds for all $\phi_1, \phi_2 \in \Phi$:

$$\mathbf{\Sigma}^{\phi_1,h}(\boldsymbol{\pi}^I) = \frac{\phi_1''(1)^2}{\phi_2''(1)^2} \mathbf{\Sigma}^{\phi_2,h}(\boldsymbol{\pi}^I).$$

Especially, $\mathbf{\Sigma}^{\phi_1,h}(\boldsymbol{\pi}^I) = \mathbf{\Sigma}^{\phi_2,h}(\boldsymbol{\pi}^I)$ if $\phi_1''(1) = \phi_2''(1)$.

Proof. Each element in $\mathbf{\Sigma}^{\phi,h}(\boldsymbol{\pi})$ is a sum of elements of the form

$$\boldsymbol{\omega}_{ij}(\boldsymbol{\pi})\boldsymbol{\omega}_{kl}(\boldsymbol{\pi})\phi''\left(\frac{s^h_{ij}(\boldsymbol{\pi})}{s^h_{ij}(\boldsymbol{\pi}^I)}\right)\phi''\left(\frac{s^h_{kl}(\boldsymbol{\pi})}{s^h_{kl}(\boldsymbol{\pi}^I)}\right),$$

where $s^h_{ij} \in \{a^h_{ij}, b^h_{ij}, c^h_{ij}, d^h_{ij}\}$ are the essential sums (1.41), that make out the generalised ϕ-scaled odds ratio and $\boldsymbol{\omega}_{ij}$ and $\boldsymbol{\omega}_{kl}$ are some function of $\boldsymbol{\pi}$, not depending on ϕ. Under independence it holds $s^h_{ij}(\boldsymbol{\pi}) = s^h_{ij}(\boldsymbol{\pi}^I)$ and thus $\phi''\left(\frac{s^h_{ij}(\boldsymbol{\pi})}{s^h_{ij}(\boldsymbol{\pi}^I)}\right) = \phi''\left(\frac{s^h_{ij}(\boldsymbol{\pi}^I)}{s^h_{ij}(\boldsymbol{\pi}^I)}\right) = \phi''(1)$. Let $\mathbf{\Sigma}$ be the matrix $\mathbf{\Sigma}^{\phi,h}(\boldsymbol{\pi}^I)$, where the ϕ'' terms are replaced by ones. Then $\mathbf{\Sigma}^{\phi_1,h}(\boldsymbol{\pi}^I) = \phi_1''(1)^2\mathbf{\Sigma}$ and $\mathbf{\Sigma}^{\phi_2,h}(\boldsymbol{\pi}^I) = \phi_2''(1)^2\mathbf{\Sigma}$. Thus

$$\mathbf{\Sigma}^{\phi_1,h}(\boldsymbol{\pi}^I) = \phi_1''(1)^2\mathbf{\Sigma} = \frac{\phi_2''(1)^2}{\phi_2''(1)^2}\phi_1''(1)^2\mathbf{\Sigma} = \frac{\phi_1''(1)^2}{\phi_2''(1)^2}\phi_2''(1)^2\mathbf{\Sigma} = \frac{\phi_1''(1)^2}{\phi_2''(1)^2}\mathbf{\Sigma}^{\phi_2,h}(\boldsymbol{\pi}^I).$$

□

5.5 Dependence Concepts

The generalised log-odds ratios are linked to different dependence concepts like likelihood-ratio dependence (cf. Section 1.9.2). But extension to generalised ϕ-scaled odds ratio does not inherit these properties, i.e. the relations in Lemma 1.9.4 does not hold in general for generalised ϕ-scaled odds ratios. Take for example the two fictional 3×3 data sets:

$$\mathbf{n}_1 = \begin{pmatrix} 761 & 460 & 5 \\ 560 & 800 & 19 \\ 177 & 521 & 270 \end{pmatrix} \qquad \mathbf{n}_2 = \begin{pmatrix} 912 & 989 & 122 \\ 472 & 541 & 870 \\ 62 & 194 & 338 \end{pmatrix}$$

It holds in the power divergence family (ϕ_λ as in (1.73)):

$$\boldsymbol{\theta}^{\phi_\lambda,CO}(\mathbf{n}_1) = \begin{pmatrix} 0.87 & 0.78 \\ 1.12 & 3.08 \end{pmatrix} > \mathbf{0} \qquad \text{for } \lambda = 0$$

$$\boldsymbol{\theta}^{\phi_\lambda,CO}(\mathbf{n}_1) = \begin{pmatrix} 0.88 & -0.30 \\ 0.92 & 3.30 \end{pmatrix} \not> \mathbf{0} \qquad \text{for } \lambda = 1.$$

$$\boldsymbol{\theta}^{\phi_\lambda,L}(\mathbf{n}_2) = \begin{pmatrix} 0.06 & 2.57 \\ 1.00 & 0.08 \end{pmatrix} > \mathbf{0} \qquad \text{for } \lambda = 0$$

$$\boldsymbol{\theta}^{\phi_\lambda,L}(\mathbf{n}_2) = \begin{pmatrix} 0.02 & 4.81 \\ 1.96 & -0.04 \end{pmatrix} \not> \mathbf{0} \qquad \text{for } \lambda = -1.$$

Thus, while $\mathbf{n}_1$ obtains a positive hazard rate ordering, the Pearson divergence (ϕ_P, $\lambda = 1$) gives negative ϕ_P-scaled continuation odds ratios. The same effect is shown for $\mathbf{n}_2$ for the local odds ratio and $\lambda = -1$. Other examples are known, but not given here. Thus, Lemma 1.9.4 cannot be generalised to ϕ-scale. Hence, the dependence concepts in contingency tables set-up are strictly linked to the Kullback-Leibler based generalised odds ratios, i.e. the generalised log-odds ratios (Section 1.9). While this, at first, is a drawback regarding the ϕ-scaled generalisation, it also reflects the greater flexibility of the generalised ϕ-scaled odds ratio, when measuring association.

5.6 Examples

The sample estimates of the local and global ϕ-scaled odds ratios based on the sexual fun data (Table 4.4) are given in Table 5.1 and the asymptotic variance-covariance matrix estimates are given in Table 5.2 (local) and Table 5.3 (global).

λ	$\boldsymbol{\theta}^{\phi_\lambda,L}(\hat{\boldsymbol{\pi}})$			$\boldsymbol{\theta}^{\phi_\lambda,G}(\hat{\boldsymbol{\pi}})$		
0	1.39	0.27	0.44	2.06	1.60	1.34
	0.22	0.76	-0.04	1.59	1.28	0.83
	-0.22	0.34	-0.37	1.17	0.90	0.41
1/3	1.58	0.22	0.39	2.17	1.56	1.25
	0.10	0.75	-0.01	1.54	1.27	0.82
	-0.18	0.35	-0.39	1.09	0.89	0.41
1	2.14	0.12	0.30	2.61	1.53	1.12
	-0.09	0.75	0.04	1.51	1.25	0.80
	-0.13	0.38	-0.45	0.98	0.87	0.42

Table 5.1: Sample estimates for the local and global ϕ-scaled odds ratio in the power divergence (Example 1.10.4 with $\lambda = 0, 1/3, 1$) based on the sexual fun data (Table 4.4).

The change of $\phi \in \Phi$ can lead to a change of association direction. Assume that the sample proportion for the sexual fun data (Table 4.4) is the true underlying multinomial probability, $\boldsymbol{\pi} = \hat{\boldsymbol{\pi}}$, then $\theta_{21}^{\phi_{KL},L} = 0.22 > 0$ but $\theta_{21}^{\phi_P,L} = -0.09 < 0$ for the Kullback-Leibler (ϕ_{KL}, $\lambda = 0$) and the Pearson (ϕ_P, $\lambda = 1$) divergence, respectively. Global odds ratio are more stable, i.e. they produce less spread values. While all global odds ratios are positive, the local odds ratio estimates tend to have a higher variation. This "smoothing effect" is a consequence of the fact that global odds ratios are based on all the cells of the contingency table while local odds ratio are based only on specific four cells of the table. The continuation and cumulative odds ratio have a similar smoothing but with a lower effect size as they do not use all cells in each generalised association factor (5.2).

The asymptotic variance-covariance matrices reflect the connections between the single odds ratios and the number of cells used. The more cells the generalised odds ratios use, the more the covariance values start to spread between the different single odds ratios, i.e. while the local odds covariance matrix has many very low covariances and only few high covariance values when odds ratios are adjacent, the global odds covariance matrix has smaller values but now values close to zero. This effect seems to hold for all kinds of generalised ϕ-scaled odds ratios. The small covariance on non-adjacent odds ratios can easily be explained by the small impact a change in a single variable has in the marginals, which then effects the generalised association factors (5.2) and therefore the overall odds ratio.

Covariance matrices are not always well-defined for contingency tables that include sampling zeros. Sampling zeros are non problematic if $\phi'(0) > -\infty$ or if each $s_{ij}^h(\boldsymbol{\pi}) > 0$ for all essential sums $s_{ij}^h \in \{a_{ij}^h, b_{ij}^h, c_{ij}^h, d_{ij}^h\}$, given by (1.41) for all $i = 1, ..., I-1, j = 1, ..., J-1$.

λ		$\theta_{11}^{\phi_\lambda,L}$	$\theta_{12}^{\phi_\lambda,L}$	$\theta_{13}^{\phi_\lambda,L}$	$\theta_{21}^{\phi_\lambda,L}$	$\theta_{22}^{\phi_\lambda,L}$	$\theta_{23}^{\phi_\lambda,L}$	$\theta_{31}^{\phi_\lambda,L}$	$\theta_{32}^{\phi_\lambda,L}$	$\theta_{33}^{\phi_\lambda,L}$
0	$\theta_{11}^{\phi_\lambda,L}$	**0.91073**	-0.26786	0.00000	-0.62500	0.12500	0.00000	0.00000	0.00000	0.00000
	$\theta_{12}^{\phi_\lambda,L}$	-0.26786	**1.10119**	-0.83333	0.12500	-0.45833	0.33333	0.00000	0.00000	-0.00000
	$\theta_{13}^{\phi_\lambda,L}$	0.00000	-0.83333	**1.30952**	-0.00000	0.33333	-0.47619	0.00000	0.00000	-0.00000
	$\theta_{21}^{\phi_\lambda,L}$	-0.62500	0.12500	0.00000	**1.82500**	-0.32500	0.00000	-1.20000	0.20000	0.00000
	$\theta_{22}^{\phi_\lambda,L}$	0.12500	-0.45833	0.33333	-0.32500	**0.90833**	-0.58333	0.20000	-0.45000	0.25000
	$\theta_{23}^{\phi_\lambda,L}$	0.00000	0.33333	-0.47619	0.00000	-0.58333	**0.83730**	-0.00000	0.25000	-0.36111
	$\theta_{31}^{\phi_\lambda,L}$	0.00000	0.00000	0.00000	-1.20000	0.20000	0.00000	**1.82500**	-0.32500	0.00000
	$\theta_{32}^{\phi_\lambda,L}$	0.00000	0.00000	0.00000	0.20000	-0.45000	0.25000	-0.32500	**0.68611**	-0.36111
	$\theta_{33}^{\phi_\lambda,L}$	0.00000	-0.00000	-0.00000	-0.00000	0.25000	-0.36111	0.00000	-0.36111	**0.54365**
1/3	$\theta_{11}^{\phi_\lambda,L}$	**0.03540**	-0.01717	-0.00334	-0.02900	0.00085	0.00013	0.00163	-0.00303	0.00339
	$\theta_{12}^{\phi_\lambda,L}$	-0.01717	**0.04472**	-0.03036	0.00721	-0.02188	0.01371	0.00020	-0.00043	0.00049
	$\theta_{13}^{\phi_\lambda,L}$	-0.00334	-0.03036	**0.04536**	-0.00016	0.01213	-0.02062	0.00037	-0.00071	0.00081
	$\theta_{21}^{\phi_\lambda,L}$	-0.02900	0.00721	-0.00016	**0.06320**	-0.01661	-0.00007	-0.03546	0.00876	0.00025
	$\theta_{22}^{\phi_\lambda,L}$	0.00085	-0.02188	0.01213	-0.01661	**0.03963**	-0.02648	0.00970	-0.02324	0.01453
	$\theta_{23}^{\phi_\lambda,L}$	0.00013	0.01371	-0.02062	-0.00007	-0.02648	**0.03996**	0.00002	0.01294	-0.01948
	$\theta_{31}^{\phi_\lambda,L}$	0.00163	0.00020	0.00037	-0.03546	0.00970	0.00002	**0.05535**	-0.01390	-0.00042
	$\theta_{32}^{\phi_\lambda,L}$	-0.00303	-0.00043	-0.00071	0.00876	-0.02324	0.01294	-0.01390	**0.03310**	-0.01891
	$\theta_{33}^{\phi_\lambda,L}$	0.00339	0.00049	0.00081	0.00025	0.01453	-0.01948	-0.00042	-0.01891	**0.02929**
1	$\theta_{11}^{\phi_\lambda,L}$	**-0.00036**	-0.00008	-0.00008	-0.00004	-0.00017	-0.00001	0.00003	-0.00010	0.00012
	$\theta_{12}^{\phi_\lambda,L}$	-0.00008	**0.00008**	-0.00004	0.00003	-0.00006	0.00002	0.00000	-0.00001	0.00001
	$\theta_{13}^{\phi_\lambda,L}$	-0.00008	-0.00004	**0.00005**	0.00000	-0.00000	-0.00004	0.00000	-0.00001	0.00002
	$\theta_{21}^{\phi_\lambda,L}$	-0.00004	0.00003	0.00000	**0.00009**	-0.00003	-0.00000	-0.00004	0.00002	-0.00000
	$\theta_{22}^{\phi_\lambda,L}$	-0.00017	-0.00006	-0.00000	-0.00003	**0.00003**	-0.00006	0.00003	-0.00009	0.00008
	$\theta_{23}^{\phi_\lambda,L}$	-0.00001	0.00002	-0.00004	-0.00000	-0.00006	**0.00009**	0.00000	0.00003	-0.00006
	$\theta_{31}^{\phi_\lambda,L}$	0.00003	0.00000	0.00000	-0.00004	0.00003	0.00000	**0.00006**	-0.00002	-0.00001
	$\theta_{32}^{\phi_\lambda,L}$	-0.00010	-0.00001	-0.00001	0.00002	-0.00009	0.00003	-0.00002	**0.00007**	-0.00004
	$\theta_{33}^{\phi_\lambda,L}$	0.00012	0.00001	0.00002	-0.00000	0.00008	-0.00006	-0.00001	-0.00004	**0.00007**

Table 5.2: Sample estimates for asymptotic variance-covariance matrix $\mathbf{\Sigma}^{\phi,L}(\hat{\boldsymbol{\pi}})$ for the **local** ϕ-scaled odds ratio in the power divergence ($\lambda = 0, 1/3, 1$) based on the sexual fun data (Table 4.4). Rounded on five digits.

5.7 Discussion

The generalisation ϕ-scaled odds ratios are introduced, which for the local odds ratio occur naturally as interpretation parameter in ϕ-association models (1.80). Since they are extended from the generalised log-odds ratio by a ϕ-divergence induced scale change, they inherit - under special conditions - most of their properties like transformation properties, the characterization of independence (cf. Section 5.3) or the object of association property from Osius [2004]. Thus, knowledge of the triple $\{\boldsymbol{\theta}^{h,\phi}, \boldsymbol{\pi}^X, \boldsymbol{\pi}^Y\}$ is equivalent to knowledge of the joint distribution $\boldsymbol{\pi}$. In addition, the asymptotic normality of generlised ϕ-scaled odds ratios has been proven and their asymptotic variance can be calculated using R functions, which can be used for inferential proposes.

Therefore, the basis of further association analysis based on generalised ϕ-scaled odds ratios is laid and the results of this chapter are used in the following.

λ		$\theta_{11}^{\phi_\lambda,G}$	$\theta_{12}^{\phi_\lambda,G}$	$\theta_{13}^{\phi_\lambda,G}$	$\theta_{21}^{\phi_\lambda,G}$	$\theta_{22}^{\phi_\lambda,G}$	$\theta_{23}^{\phi_\lambda,G}$	$\theta_{31}^{\phi_\lambda,G}$	$\theta_{32}^{\phi_\lambda,G}$	$\theta_{33}^{\phi_\lambda,G}$
0	$\theta_{11}^{\phi_\lambda,G}$	**0.44**	0.15	0.12	0.31	0.06	0.04	0.26	0.05	0.02
	$\theta_{12}^{\phi_\lambda,G}$	0.15	**0.33**	0.27	0.11	0.14	0.09	0.09	0.09	0.05
	$\theta_{13}^{\phi_\lambda,G}$	0.12	0.27	**0.45**	0.09	0.11	0.16	0.07	0.07	0.09
	$\theta_{21}^{\phi_\lambda,G}$	0.31	0.11	0.09	**0.50**	0.12	0.08	0.43	0.09	0.05
	$\theta_{22}^{\phi_\lambda,G}$	0.06	0.14	0.11	0.12	**0.20**	0.14	0.11	0.13	0.08
	$\theta_{23}^{\phi_\lambda,G}$	0.04	0.09	0.16	0.08	0.14	**0.21**	0.07	0.09	0.13
	$\theta_{31}^{\phi_\lambda,G}$	0.26	0.09	0.07	0.43	0.11	0.07	**0.65**	0.15	0.09
	$\theta_{32}^{\phi_\lambda,G}$	0.05	0.09	0.07	0.09	0.13	0.09	0.15	**0.21**	0.13
	$\theta_{33}^{\phi_\lambda,G}$	0.02	0.05	0.09	0.05	0.08	0.13	0.09	0.13	**0.20**
1/3	$\theta_{11}^{\phi_\lambda,G}$	**-0.05**	-0.23	-0.19	-0.09	-0.24	-0.16	-0.04	-0.17	-0.08
	$\theta_{12}^{\phi_\lambda,G}$	-0.23	**0.00**	-0.01	-0.17	-0.10	-0.07	-0.11	-0.07	-0.03
	$\theta_{13}^{\phi_\lambda,G}$	-0.19	-0.01	**0.15**	-0.15	-0.08	0.02	-0.10	-0.07	0.02
	$\theta_{21}^{\phi_\lambda,G}$	-0.09	-0.17	-0.15	**0.13**	-0.11	-0.07	0.14	-0.08	-0.03
	$\theta_{22}^{\phi_\lambda,G}$	-0.24	-0.10	-0.08	-0.11	**0.01**	0.01	-0.07	-0.00	0.02
	$\theta_{23}^{\phi_\lambda,G}$	-0.16	-0.07	0.02	-0.07	0.01	**0.12**	-0.04	0.00	0.09
	$\theta_{31}^{\phi_\lambda,G}$	-0.04	-0.11	-0.10	0.14	-0.07	-0.04	**0.33**	0.01	0.02
	$\theta_{32}^{\phi_\lambda,G}$	-0.17	-0.07	-0.07	-0.08	-0.00	0.00	0.01	**0.11**	0.08
	$\theta_{33}^{\phi_\lambda,G}$	-0.08	-0.03	0.02	-0.03	0.02	0.09	0.02	0.08	**0.18**
1	$\theta_{11}^{\phi_\lambda,G}$	**-5.93**	-3.77	-2.79	-3.59	-3.16	-2.05	-2.31	-2.21	-1.07
	$\theta_{12}^{\phi_\lambda,G}$	-3.77	**-2.08**	-1.55	-2.20	-1.78	-1.15	-1.42	-1.24	-0.59
	$\theta_{13}^{\phi_\lambda,G}$	-2.79	-1.55	**-1.05**	-1.63	-1.32	-0.80	-1.05	-0.92	-0.40
	$\theta_{21}^{\phi_\lambda,G}$	-3.59	-2.20	-1.63	**-1.94**	-1.78	-1.14	-1.25	-1.24	-0.59
	$\theta_{22}^{\phi_\lambda,G}$	-3.16	-1.78	-1.32	-1.78	**-1.38**	-0.88	-1.14	-0.96	-0.44
	$\theta_{23}^{\phi_\lambda,G}$	-2.05	-1.15	-0.80	-1.14	-0.88	**-0.46**	-0.74	-0.62	-0.21
	$\theta_{31}^{\phi_\lambda,G}$	-2.31	-1.42	-1.05	-1.25	-1.14	-0.74	**-0.68**	-0.75	-0.35
	$\theta_{32}^{\phi_\lambda,G}$	-2.21	-1.24	-0.92	-1.24	-0.96	-0.62	-0.75	**-0.57**	-0.24
	$\theta_{33}^{\phi_\lambda,G}$	-1.07	-0.59	-0.40	-0.59	-0.44	-0.21	-0.35	-0.24	**0.04**

Table 5.3: Sample estimates for asymptotic variance-covariance matrix $\boldsymbol{\Sigma}^{\phi,G}(\hat{\boldsymbol{\pi}})$ for the **global** ϕ-scaled odds ratio in the power divergence ($\lambda = 0, 1/3, 1$) based on the sexual fun data (Table 4.4).

Chapter 6

Measures based on Generalized ϕ-linear Models

Association models like the uniform association (1.52) or row-effect (1.57) model are able to capture complex association structures without reducing the manageability or interpretability of the analysis. But as already presented in Chapter 4, these model require fitting procedures, which can be disadvantageous and closed-form estimators for the model parameters are a good alternative. The previous introduced model-based measures are based on a unified theory for closed-form estimation of generalised association model parameter. Association models can be extended using the ϕ-divergence (Section 1.11). For example, the (local) ϕ-association model of Kateri and Papaioannou [1995] is directly linked to the local ϕ-scaled odds ratios introduced Chapter 5. In the power divergence family (Example 1.10.4), the local ϕ_λ-association model gives a connection between classic association models ($\lambda = 0$) and correlation models ($\lambda = 1$). The other generalised ϕ-scaled odds ratio can be used to define generalised ϕ-association models ($\theta_{ij}^{\phi,h} = \varphi(\mu_{i+1} - \mu_i)(\nu_{j+1} - \nu_j), i = 1, ..., I-1, j = 1, ..., J-1$). As already shown in Chapter 3 and Espendiller and Kateri [2016], such a scale change can improve the properties of association measures, which motivates the approach to generalise the closed-form estimation using the ϕ-divergence. The generalisation to ϕ-scale is therefore useful to define a unified theory.

Since the generalised ϕ-association models cannot be captured by ordinary loglinear models, the homogeneous linear predictor (HLP) model by Lang [2004, 2005] was introduced in Section 1.9.5 and applied in the context of ϕ-scaled odds ratios leading to generalised ϕ-linear models in Section 6.1. Examples are given in Section 6.2 together with a simulation study to show the effect a scale change can have on the model fit itself. The ϕ-weighted mean closed-form estimators for the generalised ϕ-uniform association and ϕ-row effect model parameters are given in Section 6.3 and 6.4, respectively. The ϕ-weighted mean estimators for the row-effect model parameters are then used to define an ordering test for the scores in Section 6.5. Next to theory and examples, a general simulation study is conducted. Results of this chapter are then summarized in Section 6.6.

6.1 Generalised ϕ-linear Models

The generalised log-linear model (GLLM)

$$\mathbf{C}\log(\mathbf{A}\boldsymbol{\pi}) = \mathbf{X}\boldsymbol{\beta}, \tag{6.1}$$

introduced in (1.59) gives the ability to define contingency table models for two categorical classification variables directly in terms of the generalised log-odds ratios (Section 1.9) by using the matrices $\mathbf{A} = \mathbf{A}^h$ and $\mathbf{C} = \mathbf{C}^h, h = L, CO, C, G$, as given in (1.43).

A generalisation approach to ϕ-scale can be made by replacing the generalised log-odds ratio on the left side of (6.1) by the generalised ϕ-scaled odds ratios (Section 5.1), i.e. by defining a model

$$\boldsymbol{\theta}^{\phi,h}(\boldsymbol{\pi}) = \mathbf{X}\boldsymbol{\beta}, \tag{6.2}$$

which is called *generalised ϕ-linear model* (GϕLM). Such a model is already available for the local ϕ-scaled odds ratio and can be described in terms of the multinomial probabilities $\boldsymbol{\pi}$ directly by using the ϕ-association model (1.80). Since such a formulation on $\boldsymbol{\pi}$-level is not available for the other generalised odds ratio, an alternative approach is to model

$$\mathbf{C}^h\phi'\left(\frac{\mathbf{A}^h\boldsymbol{\pi}}{\mathbf{A}^h\boldsymbol{\pi}^I}\right) = \mathbf{X}\boldsymbol{\beta}, \qquad h = L, CO, C, G.$$

But such a model is not covered by the GLLMs (6.1), since the marginal distributions $\mathbf{A}^h\boldsymbol{\pi}^I$ occur in the divisor. Thus, advanced models are required. Lang [2004, 2005] introduced HLP models, which take the form

$$\mathbf{L}(\boldsymbol{\pi}) = \mathbf{X}\boldsymbol{\beta}, \tag{6.3}$$

for a function $\mathbf{L}$ fulfilling special requirements mentioned in Appendix A.3 (see Definition A.3.3). The model given by setting

$$\mathbf{L}(\boldsymbol{\pi}) = \boldsymbol{\theta}^{\phi,h}(\boldsymbol{\pi})$$

in (6.3) is studied. It is checked under which conditions the GϕLM fulfils the requirements to become an HLP model:

Lemma 6.1.1. Let $\phi \in \Phi$. Let $(\mathbf{Z} = \mathbf{1}_{IJ}, \mathbf{Z}_F = \mathbf{Z}, \mathbf{n} = n)$ be the sampling plan of the multinomial sampling scheme as given in the Appendix (A.7). Let $a^h_{ij}, b^h_{ij}, c^h_{ij}, d^h_{ij}$ be the essential sums given by (1.41). Let $\boldsymbol{\mu} = (\mu_{ij}) \in \mathbb{R}^{I\times J}$. Define the generalised ϕ-scaled odds ratios in terms of the expected cell frequencies $\boldsymbol{\mu} = \mathbb{E}\mathbf{N}$ as $\boldsymbol{\theta}^{\phi,h}(\boldsymbol{\mu}) = (\theta^{\phi,h}_{11}(\boldsymbol{\mu}), ..., \theta^{\phi,h}_{I-1J-1}(\boldsymbol{\mu}))$, where

$$\begin{aligned}\theta^{\phi,h}_{ij}(\boldsymbol{\mu}) = \phi'\left(\frac{a^h_{ij}(\boldsymbol{\mu}/|\boldsymbol{\mu}|)}{a^h_{ij}(\boldsymbol{\mu}^I/|\boldsymbol{\mu}^I|)}\right) - \phi'\left(\frac{b^h_{ij}(\boldsymbol{\mu}/|\boldsymbol{\mu}|)}{b^h_{ij}(\boldsymbol{\mu}^I/|\boldsymbol{\mu}^I|)}\right)& \\ - \phi'\left(\frac{c^h_{ij}(\boldsymbol{\mu}/|\boldsymbol{\mu}|)}{c^h_{ij}(\boldsymbol{\mu}^I/|\boldsymbol{\mu}^I|)}\right) + \phi'\left(\frac{d^h_{ij}(\boldsymbol{\mu}/|\boldsymbol{\mu}|)}{d^h_{ij}(\boldsymbol{\mu}^I/|\boldsymbol{\mu}^I|)}\right)&\end{aligned} \tag{6.4}$$

for all $i = 1, ..., I-1, j = 1, ..., J-1$ and $\boldsymbol{\mu}^I$ is the expected cell frequency under independence and $|\boldsymbol{\mu}| = \sum_{i,j=1}^{I,J}\mu_{ij}$. The function $\mathbf{L}(\boldsymbol{\mu}) = \boldsymbol{\theta}^{\phi,h}(\boldsymbol{\mu})$ is $\mathbf{Z} = \mathbf{1}_{IJ}$-homogeneous for $h = L, CO, C, G$ (cf. Definition A.3.1).

Proof. Let $\delta > 0$ such that for $\boldsymbol{\mu} = (\mu_{11}, ..., \mu_{IJ})$ it holds

$$\mathrm{diag}(\mathbf{Z}\delta)\boldsymbol{\mu} = \delta\boldsymbol{\mu} = (\delta\mu_{11}, ..., \delta\mu_{IJ}).$$

In addition, it holds $|\delta\boldsymbol{\mu}| = \delta|\boldsymbol{\mu}|$ and $|(\delta\boldsymbol{\mu})^I| = \delta^2|\boldsymbol{\mu}^I|$ such that

$$\frac{\delta\boldsymbol{\mu}}{|\delta\boldsymbol{\mu}|} = \frac{\delta\boldsymbol{\mu}}{\delta|\boldsymbol{\mu}|} = \frac{\boldsymbol{\mu}}{|\boldsymbol{\mu}|} \qquad \text{and} \qquad \frac{(\delta\boldsymbol{\mu})^I}{|(\delta\boldsymbol{\mu})^I|} = \frac{\delta^2\boldsymbol{\mu}^I}{\delta^2|\boldsymbol{\mu}^I|} = \frac{\boldsymbol{\mu}^I}{|\boldsymbol{\mu}^I|}.$$

Inserting this in result in (6.4), a direct consequence is

$$\boldsymbol{\theta}^{\phi,h}(\delta\boldsymbol{\mu}) = \boldsymbol{\theta}^{\phi,h}(\boldsymbol{\mu}) \tag{6.5}$$

for $h = L, CO, C, G$, so that $\boldsymbol{\theta}^{\phi,h}$ is $\mathbf{Z}$-homogeneous. □

As a consequence of Lemma 6.1.1, all generalised ϕ-scaled odds ratios are $\mathbf{Z}$-homogeneous for the multinomial sampling scheme.

Definition 6.1.2. Let $\phi \in \Phi$ be strictly convex. Let $\mathbf{N} \sim \mathcal{M}(n, \boldsymbol{\pi})$. The generalised ϕ-linear model (GϕLM) is defined by

$$\boldsymbol{\theta}^{\phi,h}(\boldsymbol{\pi}) = \mathbf{X}\boldsymbol{\beta} \tag{6.6}$$

for $h = L, CO, C, G$.

By (6.5) the GϕLM in the previous definition can also be defined in terms of the expected cell frequencies $\boldsymbol{\mu} = \mathbb{E}\mathbf{N}$.

Lemma 6.1.3. Let $h = L, CO, C$ or G. Let $\phi \in \Phi$ be three times continuous differentiable such that $\frac{\partial\boldsymbol{\theta}^{\phi,h}(\boldsymbol{\mu})}{\partial\boldsymbol{\mu}}$ has full rank $(I-1)(J-1)$ on $\boldsymbol{\mu} \in \mathbb{R}_+^{IJ}$. Then the G$\phi$LM is an HLP model.

Proof. Let $\mathbf{N} \sim \mathcal{M}(n, \boldsymbol{\pi})$ be a multinomial distributed variable and let $\boldsymbol{\mu} = \mathbb{E}\mathbf{N} = n\boldsymbol{\pi}$ the expected cell frequencies. Let $\mathbf{L}(\boldsymbol{\mu}) = \boldsymbol{\theta}^{\phi,h}(\boldsymbol{\mu})$ as in (6.4) and let $\gamma = \mathbb{E}|\mathbf{N}| = n$ for the multinomial sampling scheme. It holds,

$$\mathbf{L}(\boldsymbol{\mu}) = \boldsymbol{\theta}^{\phi,h}(\boldsymbol{\mu}) = \boldsymbol{\theta}^{\phi,h}(\boldsymbol{\pi}) = \mathbf{L}(\boldsymbol{\pi})$$

by (6.5) so that **HLP1** in Definition A.3.3 holds with $\mathbf{a} = 0$. For **HLP2** all conditions on $\mathbf{L}$ are checked:

(i) It holds $\mathbf{L}(\boldsymbol{\mu}) = \mathbf{0}$ if $\boldsymbol{\mu}$ is the independence probability $\boldsymbol{\mu} = \boldsymbol{\pi}^I = (\pi_{i+}\pi_{+j})$ for any $\boldsymbol{\pi} \in \Delta_{I,J}$ so that

$$\omega(\mathbf{L} \mid 0) = \{\boldsymbol{\mu} \mid \boldsymbol{\mu} > 0, \mathbf{L}(\boldsymbol{\mu}) = \mathbf{0}\} \neq \emptyset.$$

(ii) For $\phi \in \Phi$ three times continuous differentiable, $\mathbf{L}$ is two times continuous differentiable by construction.

(iii) Holds by requirement.

(iv) Let $\mathbf{U}$ be a full rank matrix, an orthogonal complement of $\mathbf{X}$. For $\mathbf{h}(\boldsymbol{\mu}) = \mathbf{U}^{\mathbf{T}}\mathbf{L}(\boldsymbol{\mu})$ it holds

$$\mathbf{h}(\mathrm{diag}(\mathbf{Z}\boldsymbol{\delta})\boldsymbol{\mu}) = \mathbf{U}^T\mathbf{L}(\mathrm{diag}(\mathbf{Z}\boldsymbol{\delta})\boldsymbol{\mu}) = \mathbf{U}^T\mathbf{L}(\boldsymbol{\mu}) = \mathbf{h}(\boldsymbol{\mu})$$

by Lemma 6.1.1, so that $\mathbf{h}$ is $\mathbf{Z}$-homogeneous.

It follows that $\mathbf{L}(\boldsymbol{\mu}) = \boldsymbol{\theta}^{\phi,h}(\boldsymbol{\mu})$ defines an HLP model. □

Every GLLM is a GϕLM by using the Kullback-Leibler divergence as scale function. Dale [1986] defined global cross-ratio models (GCR) for bivariate, discrete, ordered responses using the global log-odds ratios, which is a special case of a GLLM with $h = G$. To define completely a parametric GCR model, one can additionally assume a model for the marginal distribution, which lead to a model for $F_{ij} = \mathbf{P}(X \le i, Y \le j)$ using the formula (5.32). Kateri and Papaioannou [1995] studied ϕ-association models, which corresponds to the local ϕ-scaled odds ratio, $\theta_{ij}^{\phi,L} = \varphi(\mu_{i+1} - \mu_i)(\nu_{j+1} - \nu_j), i = 1, ..., I-1, j = 1, ..., J-1$, for scores $\{\mu_i\}$, $\{\nu_j\}$ and association parameter φ (cf. Section 5.7). This model is a GϕLM model by Lemma 6.1.3.

Remark 6.1.4. As GϕLMs are HLP models, one can use the R-program `mph.fit` by Lang - which is available on request - to fit the models. His routine is for example used in the R functions `calc.MLE.R` (see Appendix B.4) or `fit.phi.U.model` (see Appendix B.6). Next to MLEs, `mph.fit` gives estimates for the asymptotic variance-covariance matrix. The asymptotic variance-covariance matrices estimate $\boldsymbol{\Sigma}^{\phi,h}(\hat{\boldsymbol{\pi}})$ (cf. (5.33)) for $h = L, CO, C, G$ were already calculated. Using the saturated model in (6.6), $\boldsymbol{\theta}^{\phi,h}(\boldsymbol{\pi}) = \text{diag}(\mathbf{1}_{IJ})\boldsymbol{\beta}$, i.e. $\theta_{ij}^{\phi,h} = \beta_{ij}, i = 1, ..., I-1, j = 1, ..., J-1$, it holds $\boldsymbol{\Sigma}^{\phi,h}(\hat{\boldsymbol{\pi}}) = \mathbb{V}\text{ar } \hat{\boldsymbol{\beta}}$, asymptotically.

6.2 Example and Scale Effects

The model fitting procedure shall be demonstrated for the popular cannabis data set (Table 4.3). The fits of the local ϕ-scaled uniform association model ($\theta_{ij}^{\phi,h} = c, i = 1, ..., I-1, j = 1, ..., J-1$) for different λ in the power divergence family (Example 1.10.4) are given in Table 6.1.

Alcohol consumption	I tried cannabis ... Never	Once or twice	More often
At mostonce/moth	204 (204.4/205.0/200.3)	6 (5.7/6.1/11.7)	1 (0.9/0.2/0.4)
Twice/moth	211 (211.4/209.3/197.6)	13 (13.1/14.7/18.3)	5 (4.5/5.3/13.7)
Twice/week	357 (352.8/345.5/341.2)	44 (48.8/48.8/45.7)	38 (37.4/44.7/51.2)
More often	92 (95.3/102.9/121.1)	34 (29.4/27.9/22.5)	49 (50.3/43.7/30.3)

Table 6.1: Students' survey about cannabis use at the University of Ioannia, Greece (1995). The expected cell frequencies are given (in parentheses) under the uniform association model $\theta_{ij}^{\phi,L} = c, i = 1, ..., I-1, j = 1, ..., J-1$, in the power divergence family, $\lambda = 0$ (left), $\lambda = 1/3$ (middle) and $\lambda = 1$ (right).

The cannabis data has already been studied for local uniform association on log scale (Kullback-Leibler) in Kateri [2014, pp. 156]. The independence model ($\boldsymbol{\theta}^{\phi,L} = \mathbf{0}$ for any $\phi \in \Phi$) and the uniform local association model

$$\theta_{ij}^{\phi,L} = c, \qquad i = 1, ..., I-1, j = 1, ..., J-1, \tag{6.7}$$

were fitted using `mph.fit` for selected members of the power divergence family. The R-function for fitting the ϕ-scaled uniform association model `fit.phi.U.model` is found in

Table 6.2: Model fits of the cannabis data based on different local ϕ-scaled uniform association models ($\theta_{ij}^{\phi,L} = c, i = 1, ..., I-1, j = 1, ..., J-1,$) in the power divergence family. The degree of freedom for the χ^2 distributed test statistics of the uniform association model is $df = 5$.

λ	-1/3	0	1/3	2/3	1
$G^2(I)$			152.793		
p-value			0		
$G^2(U)$	3.00	1.47	7.19	25.21	40.76
p-value	0.70	0.92	0.21	0.00	0.00
$G^2(I) - G^2(U)$	149.79	151.32	145.60	127.59	112.03
p-value	1.00	1.00	1.00	1.00	1.00

the Appendix B.6. Expected frequencies of the fitted models are given in Table 6.1 and values of the G^2-statistic and p-values are given in Table 6.2.

The independence model does not fit the data, while the additional parameter in the uniform association clearly improves the model fit as $G^2(U) << G^2(I)$ for all scales $\phi \in \Phi$. But while on an $\alpha = 5\%$-level the local uniform association model holds for the power divergence members $\lambda = -1/3, 0$ and $1/3$, it is rejected for $\lambda = 2/3$ and 1.

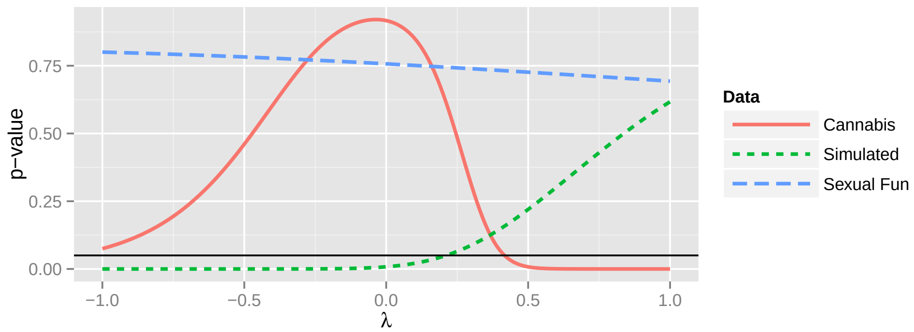


Figure 6.1: P-values for the fit of the uniform association model for ϕ-scales in the power divergence family ($\lambda \in (-1, 1)$).

As the cannabis data example shows, the scale change effects the outcome of the uniform association model. Figure 6.1 shows the p-values of the uniform association model fits dependent on $\lambda \in (-1, 1)$ in the power divergence family for selected data sets. While the fit of the cannabis data is best around $\lambda = 0$ and decreases for increasing $|\lambda|$, the sexual fun data is nearly not influenced by a scale change. A simulated data set

$$\begin{pmatrix} 462 & 142 & 9 \\ 154 & 137 & 37 \\ 11 & 40 & 8 \end{pmatrix}$$

has been chosen to have an improving fit with increasing λ to show that this effect can have any direction (improvement, stagnation or deterioration). Denote

$$U^\phi := \bigcup_{c\in\mathbb{R}} \{\boldsymbol{\pi} \in \Delta_{I,J} | \theta_{ij}^{\phi,L}(\boldsymbol{\pi}) = c, i = 1, ..., I-1, j = 1, ..., J-1\}, \qquad \phi \in \Phi$$

as the set of multinomial probability vectors fulfilling the local uniform association model on ϕ-scale. The effect of a scale change on U^{ϕ} can be shown geometrically using a simulation study for 2×3 contingency tables. Let $\boldsymbol{\pi}_1, ..., \boldsymbol{\pi}_{K_1}$ be randomly generated probabilities such that $\log \theta^L_{1j} \in [-4, 4], j = 1, 2$, where the marginal distributions are randomly generated using the algorithm presented in Appendix A.5. The corresponding R-function `get.model.prob.U` is found in B.6. In addition, for calculation of $\boldsymbol{\pi}$ the algorithm presented in Appendix A.1 was used. For each $\boldsymbol{\pi}_i$ tables $\mathbf{n}^{(i)}_1, ..., \mathbf{n}^{(i)}_{K_2}$ from a $\mathcal{M}(n, \boldsymbol{\pi}_i)$ distribution were generated and the estimates $\log \hat{\theta}^L_{1j}(\mathbf{n}^{(i)}_m), j = 1, 2, m = 1, ..., K_2$, are calculated. Using a significance level of $\alpha = 5\%$, the uniform association model on ϕ-scale has been fitted and tested. Using the log-odds ratio estimates as axis and the rejection (yes/no) as value, Figure 6.2 shows the rejection regions for the different members of the power divergence family. Notice that uniform association on log-scale is the straight line through $(0, 0)$ with slope 1. As expected, samples with sample odds ratio $\log \hat{\theta}^L_{1i}$ close to the line are not rejected and an increase of sample sizes decreases the acceptance region of the uniform association model hypothesis. An increase in λ for the power divergence family leads to more rejections.

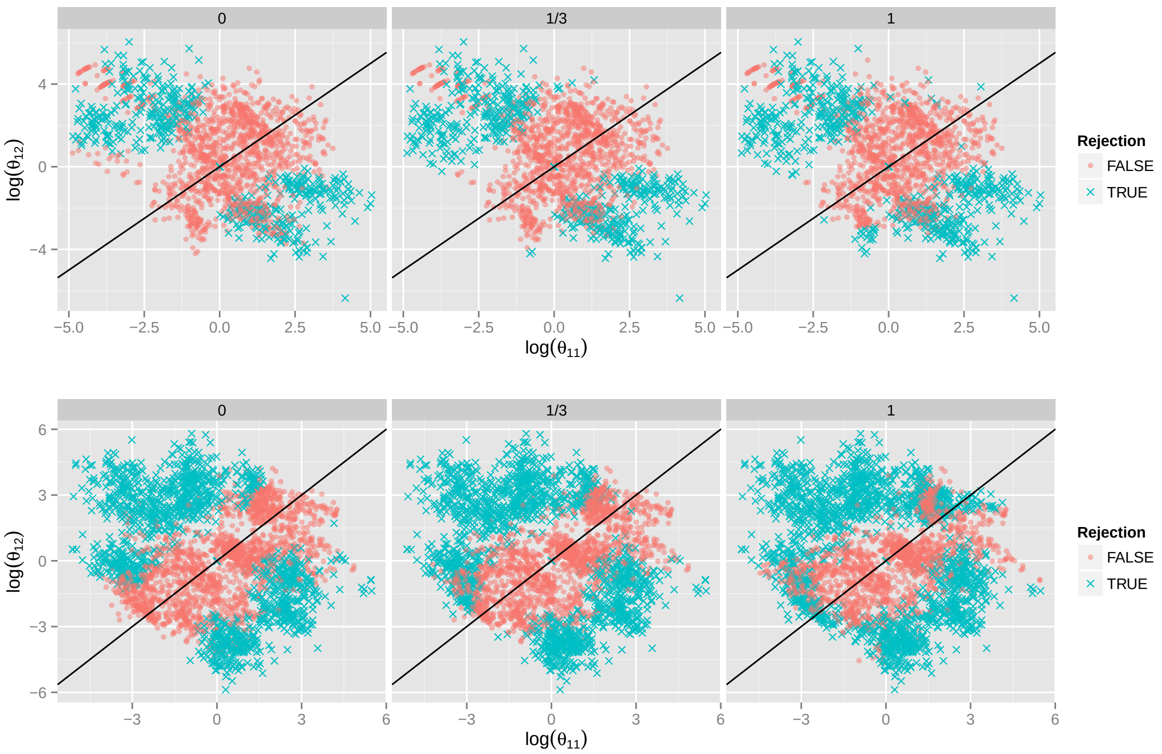


Figure 6.2: Rejection region for the local uniform association model on ϕ-scale for different members of the power divergence family ($\lambda = 0, 1/3, 1$) using $K_1 = 500, K_2 = 100$ and for sample sizes $n = 100$ (top) and $n = 250$ (bottom).

6.3 Estimating ϕ-Uniform Association

Iterative estimation procedures can have disadvantages as it has already been pointed out in Chapter 4. Therefore, closed-form estimators of the parameters of association models in contingency table analysis are introduced, which later can be used as measure of association. In Section 4.3, next to some other, weighted mean estimators were introduced for the uniform association parameter on log-scale. This approach can be generalised to ϕ-scale to get a closed-form weighted mean estimator for the ϕ-scaled uniform association model parameter ($\theta_{ij}^{\phi,h} = c, i = 1, ..., I-1, j = 1, ..., J-1$).

Let $\boldsymbol{\omega} = (\omega_{ij})$ be weights with $\sum_{i,j=1}^{I-1,J-1} \omega_{ij} = 1$. For $\mathbf{N} \sim \mathcal{M}(n, \boldsymbol{\pi})$ let $\hat{\boldsymbol{\theta}}^{\phi,h} = \boldsymbol{\theta}^{\phi,h}(\hat{\boldsymbol{\pi}}), \hat{\boldsymbol{\pi}} = \mathbf{N}/n$, be the MLE of the generalised ϕ-scaled odds ratio $\boldsymbol{\theta}^{\phi,h}$ (5.3). Define the *ϕ-weighted mean* (ϕ-WM) estimator for the ϕ-scaled uniform association model parameter as

$$\hat{\theta}^{\phi,h,WM} = \boldsymbol{\omega}^T \hat{\boldsymbol{\theta}}^{\phi,h}. \tag{6.8}$$

Notice that there is no straightforward generalisation of the Mantel Haenszel estimator to ϕ-scale, since all obvious generalisations of (4.10) cannot cover the structure of the ϕ-scaling by weighting the association factors with ϕ'.

Lemma 6.3.1. Let $\phi \in \Phi$. The ϕ-WM estimator (6.8) is consistent under the uniform association model ($\theta_{ij}^{\phi,h} = c$).

Proof. Since $\hat{\boldsymbol{\theta}}^{\phi,h}$ is the MLE of $\boldsymbol{\theta}^{\phi,h}$ it holds $\hat{\boldsymbol{\theta}}^{\phi,h} \xrightarrow{P} \boldsymbol{\theta}^{\phi,h}$ for $n \to \infty$ by the invariance property for MLEs (Casella and Berger [2002, Theorem 7.2.10, p. 320]). Thus under $\theta_{ij}^{\phi,h} = c, i = 1, ..., I-1, j = 1, ..., J-1$, it can be concluded that

$$\hat{\theta}^{\phi,h,WM} = \sum_{i,j=1}^{I,J} \omega_{ij} \hat{\theta}_{ij}^{\phi,h} \xrightarrow{P} \sum_{i,j=1}^{I,J} \omega_{ij} c = c \qquad \text{for} \qquad n \to \infty$$

and the ϕ-WM estimator is consistent. □

The weights should - again - be chosen variance minimal such that $\min_{\boldsymbol{\omega}} \mathbb{V}\text{ar}\ \boldsymbol{\omega}^T \hat{\boldsymbol{\theta}}^{\phi,h}$ is obtained:

Lemma 6.3.2. Let $\phi \in \Phi$ and let $\boldsymbol{\pi} \in \Delta_{I,J}$. Let $\boldsymbol{\Sigma}^{\phi,h}(\boldsymbol{\pi})$ be the variance-covariance matrix of $\hat{\boldsymbol{\theta}}^{\phi,h}$ as given in (5.33). The variance minimal weights for the ϕ-WM estimator (6.8) are:

$$\boldsymbol{\omega} = \frac{\left(\boldsymbol{\Sigma}^{\phi,h}(\boldsymbol{\pi})\right)^{-1} \mathbf{1}}{\mathbf{1}^T \left(\boldsymbol{\Sigma}^{\phi,h}(\boldsymbol{\pi})\right)^{-1} \mathbf{1}}. \tag{6.9}$$

Proof. Using the asymptotic normality of the ϕ-scaled odds ratio, $\sqrt{n}(\hat{\boldsymbol{\theta}}^{\phi,h} - \boldsymbol{\theta}^{\phi,h}) \xrightarrow{d} \mathcal{N}(0, \boldsymbol{\Sigma}^{\phi,h}(\boldsymbol{\pi}))$ for $n \to \infty$ (Theorem 5.4.1), the result is a direct consequence of Lemma 4.3.2. □

Weights on log-scale can be calculated by inserting the model probability estimate into the variance-covariance matrix (see Remark 4.3.5). This approach can be generalised to ϕ-scale and appropriate choices for the asymptotic variance-covariance matrix in (6.9) are again based on the saturated model ($\boldsymbol{\Sigma}^{\phi,h}(\hat{\boldsymbol{\pi}})$) or the independence model ($\boldsymbol{\Sigma}^{\phi,h}(\hat{\boldsymbol{\pi}}^I)$, $\hat{\boldsymbol{\pi}}^I = (\hat{\pi}_{i+}\hat{\pi}_{+j})$) as they do not require iterative procedures. To denote their origin the weights are denoted by $\boldsymbol{\omega} = \boldsymbol{\omega}(S)$ and $\boldsymbol{\omega} = \boldsymbol{\omega}(I)$ for the weights based on the saturated or independence model, respectively.

The ϕ-WM estimator (6.8) using the variance-minimal weights (6.9) is calculated using the R-function `calc.WM.U.phi` (see Appendix B.6).

Lemma 6.3.3. The weights (6.9) for the ϕ-WM estimator (6.8) based on the variance-covariance estimate under independence ($\boldsymbol{\omega} = \boldsymbol{\omega}(I)$) are independent of ϕ.

Proof. For $\phi \in \Phi$, denote by $\boldsymbol{\omega}_\phi$ the variance-minimal weights given in (6.9) on ϕ-scale. By Lemma 5.4.4 it holds for any $\phi_1, \phi_2 \in \Phi$:

$$\boldsymbol{\Sigma}^{\phi_1,h}(\boldsymbol{\pi}^I) = \frac{\phi_1''(1)^2}{\phi_2''(1)^2}\boldsymbol{\Sigma}^{\phi_2,h}(\boldsymbol{\pi}^I).$$

Let $\phi_1 = \phi$ and $\phi_2 = \phi_{KL}$ (Kullback-Leibler divergence, Example 1.10.2) with $\phi''_{KL}(1) = 1$, then

$$\begin{aligned}\boldsymbol{\omega}_\phi(I) &= \frac{\left(\boldsymbol{\Sigma}^{\phi,L}(\boldsymbol{\pi}^I)\right)^{-1}\mathbf{1}}{\mathbf{1}^T\left(\boldsymbol{\Sigma}^{\phi,L}(\boldsymbol{\pi}^I)\right)^{-1}\mathbf{1}} = \frac{\left((\phi''(1))^2\boldsymbol{\Sigma}^{\phi_{KL},L}(\boldsymbol{\pi}^I)\right)^{-1}\mathbf{1}}{\mathbf{1}^T\left((\phi''(1))^2\boldsymbol{\Sigma}^{\phi_{KL},L}(\boldsymbol{\pi}^I)\right)^{-1}\mathbf{1}} \\ &= \frac{\left(\boldsymbol{\Sigma}^{\phi_{KL},L}(\boldsymbol{\pi}^I)\right)^{-1}\mathbf{1}}{\mathbf{1}^T\left(\boldsymbol{\Sigma}^{\phi_{KL},L}(\boldsymbol{\pi}^I)\right)^{-1}\mathbf{1}} = \boldsymbol{\omega}_{\phi_{KL}}(I),\end{aligned}$$

which completes the proof. □

Notice that generalised ϕ-scaled odds ratios can be less sensitive for sampling zeros, dependent on the scale $\phi \in \Phi$. They produce finite estimates if $\phi'(0) := \lim_{t\searrow 0}\phi'(0) > -\infty$ and variance-covariance estimates exists if $\phi''(0) := \lim_{t\searrow 0}\phi''(0)$ is finite. In the case of the power divergence family (ϕ_λ, Example 1.10.4), estimates are finite for $\lambda > 0$ and additionally variance-covariance estimates based on the saturated model exist even when sampling zeros occur for $\lambda \geq 1$. In the other cases, the variance minimal weights (6.9) do not exist when based on the saturated estimate $\boldsymbol{\Sigma}^{\phi,h}(\hat{\boldsymbol{\pi}})$ and sampling zeros leads to a zero estimate in one of the essential sums $a_{ij}^h, b_{ij}^h, c_{ij}^h, d_{ij}^h$ given by (1.41). Clearly, $\boldsymbol{\Sigma}^{\phi,h}(\hat{\boldsymbol{\pi}}^I)$ does always exist and is independent of ϕ by Lemma 6.3.3, such that the weighted mean estimator with $\boldsymbol{\omega} = \boldsymbol{\omega}(I)$ can always be used for any $\phi \in \Phi$.

6.4 Estimating Row-Effect Association

Section 4.5 introduced a weighted mean estimator for the generalised row-effect model parameter (4.16) on log-scale. This closed-form estimator can be extended to ϕ-scale.

The *ϕ-row-effect weighted mean estimator* (ϕ-R-WM) can be defined based on the ϕ-scaled row-effect model

$$\theta_{ij}^{\phi,h} = c_{1i}, \qquad i = 1, ..., I-1, j = 1, ..., J-1. \tag{6.10}$$

Let

$$\boldsymbol{\theta}_i^{\phi,h} = \left(\theta_{i1}^{\phi,h}, ..., \theta_{iJ-1}^{\phi,h}\right) \tag{6.11}$$

be the ϕ-scaled odds ratio in row i and let $\boldsymbol{\omega}^{(i)} = (\omega_1^{(i)}, ..., \omega_{J-1}^{(i)})$ be weights with $\sum_{j=1}^{J-1} \omega_j^{(i)} = 1, i = 1, ..., I-1$. The ϕ-R-WM is defined by

$$\hat{\boldsymbol{\theta}}^{\phi,h,RWM} = \left(\hat{\theta}_1^{\phi,h,RWM}, ..., \hat{\theta}_{J-1}^{\phi,h,RWM}\right) \qquad \text{with} \qquad \hat{\theta}_i^{\phi,h,RWM} = \left(\boldsymbol{\omega}^{(i)}\right)^T \hat{\boldsymbol{\theta}}_i^{\phi,h}, \tag{6.12}$$

for $i = 1, ..., I-1$, where $\hat{\boldsymbol{\theta}}_i^{\phi,h} = \boldsymbol{\theta}_i^{\phi,h}(\hat{\boldsymbol{\pi}})$ are the sample estimators of the generalised ϕ-scaled odds ratios. Using the $(I-1) \times (I-1)(J-1)$ matrix

$$\mathbf{A} = \begin{pmatrix} \boldsymbol{\omega}^{(1)} & & \\ & \ddots & \\ & & \boldsymbol{\omega}^{(I-1)} \end{pmatrix}$$

the ϕ-R-WM estimator becomes

$$\hat{\boldsymbol{\theta}}^{\phi,h,RWM} = \mathbf{A}\hat{\boldsymbol{\theta}}^{\phi,h}. \tag{6.13}$$

Lemma 6.4.1. Let $\mathbf{N} \sim \mathcal{M}(n, \boldsymbol{\pi})$ and let $\boldsymbol{\Sigma}^{\phi,h}(\boldsymbol{\pi})$ be the variance-covariance matrix of $\hat{\boldsymbol{\theta}}^{\phi,h}$ as given in (5.33). In the situation above, the ϕ-R-WM estimator $\hat{\boldsymbol{\theta}}^{\phi,h,RWM} = \mathbf{A}\hat{\boldsymbol{\theta}}^{\phi,h}$ is asymptotic normal distributed with mean $\mathbf{A}\boldsymbol{\theta}^{\phi,h}$ and asymptotic variance-covariance matrix

$$\mathbf{A}^T\boldsymbol{\Sigma}^{\phi,h}(\boldsymbol{\pi})\mathbf{A}, \tag{6.14}$$

i.e.

$$\sqrt{n}\left(\hat{\boldsymbol{\theta}}^{\phi,h,RWM} - \mathbf{A}\boldsymbol{\theta}^{\phi,h}\right) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbf{A}^T\boldsymbol{\Sigma}^{\phi,h}(\boldsymbol{\pi})\mathbf{A}) \qquad \text{for} \qquad n \to \infty. \tag{6.15}$$

Proof. This is a direct consequence of the asymptotic normality of $\hat{\boldsymbol{\theta}}^{\phi,h}$ in equation (5.35) and the properties of multivariate normal distributions under matrix multiplication. □

Notice that the asymptotic variance-covariance matrix $\mathbf{A}^T\boldsymbol{\Sigma}^{\phi,h}(\boldsymbol{\pi})\mathbf{A}$ can either be estimated under the saturated ($\hat{\boldsymbol{\pi}}$) or independence ($\hat{\boldsymbol{\pi}}^I$) model. Following the technique in Section 4.5.1, the weights are again chosen variance-minimal in the sense that each $\min_{\boldsymbol{\omega}^{(i)}} \mathbb{V}\text{ar} \left(\boldsymbol{\omega}^{(i)}\right)^T \hat{\boldsymbol{\theta}}_i^{\phi,h}, i = 1, ..., I-1$, is obtained. To calculate the weights, separate the variance-covariance matrix $\boldsymbol{\Sigma}^{\phi,h}(\boldsymbol{\pi})$ according to the row odds ratios (6.11), e.g. $\boldsymbol{\Sigma}^{\phi,h}(\boldsymbol{\pi}) = \left(\boldsymbol{\Sigma}_{kl}^{\phi,h}(\boldsymbol{\pi})\right)$, with

$$\boldsymbol{\Sigma}_{kl}^{\phi,h}(\boldsymbol{\pi}) = \text{Cov}(\hat{\boldsymbol{\theta}}_k^{\phi,h}, \hat{\boldsymbol{\theta}}_l^{\phi,h}), \qquad k, l = 1, ..., I-1. \tag{6.16}$$

The variance-covariance matrix $\boldsymbol{\Sigma}_{kk}^{\phi,h}(\boldsymbol{\pi})$ of $\hat{\boldsymbol{\theta}}_k^{\phi,h}$ can then be used to calculate the variance minimal weights by applying Lemma 4.3.2:

$$\boldsymbol{\omega}^{(k)} = \frac{\left(\boldsymbol{\Sigma}_{kk}^{\phi,h}(\boldsymbol{\pi})\mathbf{1}\right)^{-1}}{\mathbf{1}^T\boldsymbol{\Sigma}_{kk}^{\phi,h}(\boldsymbol{\pi})\mathbf{1}}, \qquad k = 1, ..., I-1. \tag{6.17}$$

Clearly, the variance-covariance matrix is in general unknown, such that it must be estimated. Appropriate choices are the estimates based on the saturated model ($\boldsymbol{\Sigma}^{\phi,h}(\hat{\boldsymbol{\pi}})$) or the independence model ($\boldsymbol{\Sigma}^{\phi,h}(\hat{\boldsymbol{\pi}}^I)$) as they are available in closed-form. The ϕ-R-WM estimator (6.13) using the variance-minimal weights (6.17) is calculated using the R-function `WM.calc.R.model.phi` and the model fit of the ϕ-scaled row-effect association model (6.10) is obtained by `calc.MLE.R.phi`, which both can be found in the Appendix B.6.

Lemma 4.5.3 and 6.3.3 can be generalised to the ϕ-R-WM estimators:

Lemma 6.4.2. For $\phi \in \Phi$, let $\boldsymbol{\omega}_\phi = (\boldsymbol{\omega}_\phi^{(1)}, ..., \boldsymbol{\omega}_\phi^{(I-1)})$ be the variance-minimal weights (6.17) for the ϕ-R-WM $\hat{\boldsymbol{\theta}}^{\phi,h,RWM}$ (6.13). The independence based weights, $\boldsymbol{\omega}_\phi = \boldsymbol{\omega}_\phi(I)$, are independent of ϕ. In the case $h = L$ it holds:

$$\boldsymbol{\omega}_\phi^{(1)}(I) = ... = \boldsymbol{\omega}_\phi^{(I-1)}(I).$$

Proof. Let $\boldsymbol{\pi} = (\pi_{ij}) \in \Delta_{I,J}$ and let $\boldsymbol{\pi}^I = (\pi_{i+}\pi_{+j})$ be the probability vector under independence. The weights (6.17) are defined by

$$\boldsymbol{\omega}_\phi^{(i)}(I) = \frac{\left(\boldsymbol{\Sigma}_{ii}^{\phi,h}(\boldsymbol{\pi}^I)\right)^{-1}\mathbf{1}}{\mathbf{1}^T\left(\boldsymbol{\Sigma}_{ii}^{\phi,h}(\boldsymbol{\pi}^I)\right)^{-1}\mathbf{1}}, \qquad i = 1, ..., I-1,$$

where $\boldsymbol{\Sigma}^{\phi,h}(\boldsymbol{\pi}^I) = (\boldsymbol{\Sigma}_{kl}^{\phi,h}(\boldsymbol{\pi}^I))$ is the partition of the variance-covariance (5.33) as in (6.16) under independence. By Lemma 5.4.4 it holds for any $\phi_1, \phi_2 \in \Phi$:

$$\boldsymbol{\Sigma}^{\phi_1,h}(\boldsymbol{\pi}^I) = \frac{\phi_1''(1)^2}{\phi_2''(1)^2}\boldsymbol{\Sigma}^{\phi_2,h}(\boldsymbol{\pi}^I).$$

Let $\phi_1 = \phi$ and $\phi_2 = \phi_{KL}$ (Kullback-Leibler divergence, Example 4.16) with $\phi_{KL}''(1) = 1$, then

$$\begin{aligned}\boldsymbol{\omega}_\phi^{(i)}(I) &= \frac{\left(\boldsymbol{\Sigma}_{ii}^{\phi,h}(\boldsymbol{\pi}^I)\right)^{-1}\mathbf{1}}{\mathbf{1}^T\left(\boldsymbol{\Sigma}_{ii}^{\phi,h}(\boldsymbol{\pi}^I)\right)^{-1}\mathbf{1}} = \frac{\left((\phi''(1))^2\boldsymbol{\Sigma}_{ii}^{\phi_{KL},h}(\boldsymbol{\pi}^I)\right)^{-1}\mathbf{1}}{\mathbf{1}^T\left((\phi''(1))^2\boldsymbol{\Sigma}_{ii}^{\phi_{KL},h}(\boldsymbol{\pi}^I)\right)^{-1}\mathbf{1}} \\ &= \frac{\left(\boldsymbol{\Sigma}_{ii}^{\phi_{KL},h}(\boldsymbol{\pi}^I)\right)^{-1}\mathbf{1}}{\mathbf{1}^T\left(\boldsymbol{\Sigma}_{ii}^{\phi_{KL},h}(\boldsymbol{\pi}^I)\right)^{-1}\mathbf{1}} = \boldsymbol{\omega}_{\phi_{KL}}^{(i)}(I), \qquad i = 1, ..., I-1.\end{aligned}$$

In the case $h = L$ application of Lemma 4.5.3 completes the proof. □

6.4.1 Examples

Table 6.3 shows the MLE, ϕ-WM (S) and ϕ-WM (I) estimates for the ϕ-uniform local and global association model parameters, together with the standard error estimated. In addition, Table 6.4 shows the ϕ-row-effect model parameters. Notice that for the Streptococcus data (2×3) the R-model is the uniform association model and thus excluded from Table 6.4. The computation times for the different estimators are given in Table 6.5. The hardware used was an Intel(R) Core(TM) i5-3470 CPU 3.20Ghz with 4 GB RAM. For the MLEs, the R function `calc.MLE.R.phi` based on the fitting procedure of `mph.fit` was used. The calculation times can be greatly improved using the closed-form estimators, whose calculation time is nearly independent of $\phi \in \Phi$. For the MLEs, the time consumption increases with increasing λ in the power divergence family (Example 1.10.4). For $\lambda = 1$ and the cannabis data (Table 4.3), the local closed-form estimator ϕ-R-WM took only 1.33% and 1.18% of the MLE calculation time for weight type S and I, respectively. The worst time improvement in the selected data sets is for the intellectual capacity girls data (Table 4.7) in the case $\lambda = 0$, where the closed-form estimation process used 26.08% and 25.52% of the MLE computation time, respectively. The global closed-form estimation seems to give even grater improvements, using only 2.57% and 2.59% for the global ϕ-R-WM (S) and ϕ-R-WM (I) estimators in the case $\lambda = 1$ and the intellectual capacity girls data set.

The observed improvements of calculation times can easily be explained using the formulas of the closed-form estimators, which are obtained using matrix calculations, which is very time efficient. Therefore, the closed-form estimators are calculated faster than the MLE. Their calculation effort is also robust against a scale change, giving similar computation times for all analysed λ, while the MLE procedure takes more time for large $|\lambda|$.

Comparing the estimates using the standard error estimated, most closed-form estimators do not differ from the MLE significantly and only few outlier are observed. For example, the $\lambda = 1$ global uniform association estimation for the Job Satisfaction gives a value of 1.3287 for the WM (S) estimation, while the MLE lies at 0.6523. This effect can be explained due to the sampling zero, since the estimators for $\lambda \geq 1$ take sampling zeros into greater account. This effect can be reduced by taking the independence model as basis for the weights, which by definition is not effected by sampling zeros and gives a value of 0.5656 much closer to the MLE.

The closed-form R-model parameter estimates also seem to be an adequate replacement for the MLE, having only few significant outliers, for example again for the Job Satisfaction data ($\lambda = 1, h = G$). Especially, the ϕ-weighted mean estimators using independence based weights $\boldsymbol{\omega} = \boldsymbol{\omega}(I)$ give estimates close to the MLE and are less sensitive regarding sampling zeros. But notice that sampling zeros, which leads to generalised association factors with zero values can still give undefined or infinite estimates.

The single R-model estimates can be effected by a scale change. For the Intellectual Capacity (Girls) data, the $\theta_{5j}^{\phi,L} = c_{15}$ is 0.315 ($\lambda = 0$) resp. 0.075 ($\lambda = 1$). Notice that the estimates cannot be compared for the different ϕ-scales, but such a comparison can be possible using scores.

The scores $\{\mu_i\}$ of the ϕ-R-association model ($\theta_{ij}^{\phi,h} = \varphi(\mu_{i+1} - \mu_i)(\nu_{j+1} - \nu_j)$) can again be calculated by solving the LSE (4.24) for $\theta_{ij}^{\phi,h} = c_{1i}$. The corresponding R-function `get.phi.scores` can be found in the Appendix B.6. There is a connection between ordering of the scores and stochastic ordering for ϕ-association models based on the local odds ratio ($h = L$), which has already been given in Kateri and Papaioannou [1995, Property 4, p.190]:

Lemma 6.4.3. Let $\phi \in \Phi$ be strictly convex. If the scores $\mu_i, i = 1, ..., I$, and $\nu_j, j =$

$1, ..., J$, are monotonic then under the ϕ-association model (1.80) the conditional probabilities

$$\sum_{i \le a} \pi_{i|j} \qquad \text{and} \qquad \sum_{j \le b} \pi_{j|i}$$

are non-increasing (non-decreasing) in j and i respectively, for any fixed a and b, provided that $\varphi \ge 0$ ($\varphi \le 0$).

The prove of Lemma 6.4.3 is essentially based on the model formulation on $\boldsymbol{\pi}$-level (1.80). Since there is no $\boldsymbol{\pi}$-level formulation for the generalised ϕ-association models, extensions of this result to $h = CO, C, G$ are not available yet. The ϕ-scaled scores are calculated and shown in Figures 6.3 (local) and Figure 6.4 (global) for the power divergence family ($\lambda \in [-1, 2]$) and the selected data sets. The scale change does nearly not effect the score outcome for the Streptococcus, Sexual Fun and Welfare data. Also, MLE of the Job Satisfaction data is robust against a scale change on score level, but the weighted mean estimation gives stable score estimates for $\boldsymbol{\omega} = \boldsymbol{\omega}(S)$ weights for $\lambda \ge 1$ and ϕ-R-WM (I) gives unstable scores for $\lambda \searrow 0$. The IC Girls data is nearly not effected for the local ϕ-scaled R-model estimates, but gives clearer ordering for the global ϕ-scaled R-association scores. The Cannabis data is the most effected data set, heavily changing the distance between the scores for different λ. In this example, $\lambda = 0$ gives equidistant scores, while for $\lambda = 1$, the first and second row scores are nearly the same, $\mu_1 \approx \mu_2$.

6.4.2 Remark on the Loss of Ordering on ϕ-scale

As already known, a generalisation of odds ratio to ϕ-scale loses the ordinary concept of stochastically dependence (Section 5.5), but can sometimes be extended to ϕ-scale (cf. Lemma 6.4.3). However, the loss of ordering can also be observed for the ϕ-R-WM estimator. The local ϕ_λ-R-WM (S) estimator ($h = L$) and the ϕ_λ-scaled scores for the Job Satisfaction data for fixed $\nu_{j+1} - \nu_j = 1, j = 1, ..., J-1$ are

$$\hat{\theta}^{\phi,L,RWM} = (-0.026, 0.163, 0.470), \qquad \boldsymbol{\mu}_1 = (-0.35, -0.40, -0.08, 0.84), \qquad \varphi_1 = 0.51$$

for $\lambda = 1$ and

$$\hat{\theta}^{\phi,L,RWM} = (0.021, 0.107, 0.277), \qquad \boldsymbol{\mu}_2 = (-0.43, -0.36, -0.03, 0.83), \qquad \varphi_2 = 0.32$$

for $\lambda = 2$.

While scores are ordered for $\lambda = 2$, they are not ordered for $\lambda = 1$. Therefore, the overall ordering of row scores changes for different λ. This must not be a disadvantage, as the softening of the order concept opens new opportunities: As already mentioned, the scale variation can change the sampling zero induced sensibility, or revalue the distance between scores (see Cannabis data).

		Strepto.	Cannabis	Sexual Fun	Job Satisf.	Welfare Spending	IC Girls
		n=1398	n=1054	n=91	n=96	n=955	n=8310
		2×3	4×3	4×4	4×4	3×5	6×7
		Local uniform association (U^L)					
$\lambda = -1/3$	ϕ-MLE	0.4300 (0.1598)	0.8153 (0.0864)	0.3024 (0.1052)	0.2034 (0.1279)	-0.0131 (0.0340)	0.1836 (0.0065)
	ϕ-WM (S)	0.4291 (0.1555)	0.8018 (0.0890)	0.2904 (0.1063)	- (-)	-0.0039 (0.0347)	0.1650 (0.0067)
	ϕ-WM (I)	0.4051 (0.1627)	1.0597 (0.0510)	0.2836 (0.0842)	∞ (-)	-0.0195 (0.0340)	0.2223 (0.0050)
$\lambda = 0$	ϕ-MLE	0.4286 (0.1614)	0.8026 (0.0783)	0.2885 (0.0945)	0.2138 (0.1252)	-0.0131 (0.0340)	0.1786 (0.0059)
	ϕ-WM (S)	0.4283 (0.1579)	0.8029 (0.0768)	0.2888 (0.0945)	- (-)	-0.0086 (0.0344)	0.1718 (0.0060)
	ϕ-WM (I)	0.4091 (0.1627)	0.7817 (0.0510)	0.2705 (0.0842)	∞ (-)	-0.0179 (0.0340)	0.1903 (0.0050)
$\lambda = 1/3$	ϕ-MLE	0.4237 (0.1615)	0.6857 (0.0621)	0.2755 (0.0853)	0.2254 (0.1211)	-0.0132 (0.0340)	0.1702 (0.0051)
	ϕ-WM (S)	0.4240 (0.1600)	0.6957 (0.0568)	0.2796 (0.0850)	- (-)	-0.0132 (0.0341)	0.1721 (0.0051)
	ϕ-WM (I)	0.4156 (0.1627)	0.6451 (0.0510)	0.2636 (0.0842)	0.2960 (0.1205)	-0.0163 (0.0340)	0.1717 (0.0050)
$\lambda = 2/3$	ϕ-MLE	0.4162 (0.1600)	0.4720 (0.0283)	0.2640 (0.0780)	0.2385 (0.1145)	-0.0133 (0.0340)	0.1538 (0.0041)
	ϕ-WM (S)	0.4158 (0.1617)	0.5193 (0.0351)	0.2630 (0.0781)	- (-)	-0.0177 (0.0338)	0.1635 (0.0041)
	ϕ-WM (I)	0.4246 (0.1627)	0.5853 (0.0510)	0.2626 (0.0842)	0.2319 (0.1205)	-0.0148 (0.0340)	0.1619 (0.0050)
$\lambda = 1$	ϕ-MLE	0.4071 (0.1577)	0.3667 (0.0146)	0.2543 (0.0725)	0.2556 (0.0993)	-0.0134 (0.0340)	0.1275 (0.0026)
	ϕ-WM (S)	0.4037 (0.1627)	0.3797 (0.0217)	0.2398 (0.0733)	0.3192 (0.0257)	-0.0220 (0.0334)	0.1473 (0.0032)
	ϕ-WM (I)	0.4363 (0.1627)	0.5732 (0.0510)	0.2680 (0.0842)	0.2092 (0.1205)	-0.0132 (0.0340)	0.1586 (0.0050)
		Global uniform association (U^G)					
$\lambda = -1/3$	ϕ-MLE	0.6080 (0.2221)	1.8137 (0.1674)	1.0968 (0.3509)	0.5319 (0.3450)	-0.0165 (0.1107)	1.0522 (0.0333)
	ϕ-WM (S)	0.6071 (0.2203)	1.8045 (0.1671)	1.0199 (0.3514)	- (-)	-0.0047 (0.1110)	1.0350 (0.0322)
	ϕ-WM (I)	0.5720 (0.2253)	2.2237 (0.1452)	1.0992 (0.3425)	∞(0.3438)	-0.0227 (0.1107)	1.2238 (0.0341)
$\lambda = 0$	ϕ-MLE	0.6026 (0.2251)	1.8622 (0.1590)	1.0928 (0.3427)	0.5599 (0.3439)	-0.0165 (0.1107)	1.0960 (0.0334)
	ϕ-WM (S)	0.6022 (0.2241)	1.8692 (0.1571)	1.0667 (0.3420)	- (-)	-0.0152 (0.1109)	1.1031 (0.0327)
	ϕ-WM (I)	0.5726 (0.2253)	1.9495 (0.1452)	1.0658 (0.3425)	∞(0.3438)	-0.0212 (0.1107)	1.1642 (0.0341)
$\lambda = 1/3$	ϕ-MLE	0.5881 (0.2242)	1.8232 (0.1421)	1.0779 (0.3324)	0.5896 (0.3409)	-0.0164 (0.1107)	1.1232 (0.0326)
	ϕ-WM (S)	0.5877 (0.2257)	1.8518 (0.1396)	1.0747 (0.3319)	- (-)	-0.0259 (0.1107)	1.1566 (0.0321)
	ϕ-WM (I)	0.5755 (0.2253)	1.7841 (0.1452)	1.0429 (0.3425)	0.6387 (0.3438)	-0.0196 (0.1107)	1.1241 (0.0341)
$\lambda = 2/3$	ϕ-MLE	0.5676 (0.2201)	1.6229 (0.1034)	1.0505 (0.3188)	0.6207 (0.3349)	-0.0164 (0.1107)	1.1156 (0.0304)
	ϕ-WM (S)	0.5648 (0.2248)	1.6552 (0.1066)	1.0413 (0.3194)	- (-)	-0.0365 (0.1104)	1.1683 (0.0294)
	ϕ-WM (I)	0.5810 (0.2253)	1.6959 (0.1452)	1.0300 (0.3425)	0.6109 (0.3438)	-0.0180 (0.1107)	1.0991 (0.0341)
$\lambda = 1$	ϕ-MLE	0.5451 (0.2142)	1.3260 (0.0563)	1.0124 (0.3025)	0.6523 (0.3243)	-0.0163 (0.1107)	1.0324 (0.0253)
	ϕ-WM (S)	0.5364 (0.2215)	1.3489 (0.0688)	0.9823 (0.3041)	1.3287 (0.0733)	-0.0466 (0.1100)	1.1051 (0.0242)
	ϕ-WM (I)	0.5891 (0.2253)	1.6678 (0.1452)	1.0268 (0.3425)	0.5959 (0.3438)	-0.0164 (0.1107)	1.0865 (0.0341)

Table 6.3: Estimates and variance estimates (in parentheses) for the different closed-form estimators of the local and global ϕ-scaled uniform association parameter ($\theta_{ij}^{\phi,h} = c$, $h = L, G$, $i = 1, ..., I-1, j = 1, ..., J-1$) in the power divergence family (Example 1.10.4). Weights are estimated variance minimal based on the saturated (S) or independence (I) model. Variance-covariance matrix for $\lambda < 1$ cannot be calculated for the Job Satisfaction data due to sampling zeros and give undefined estimates.

	Local ϕ-R-MLE	Local ϕ-R-WM (S)	Local ϕ-R-WM (I)	Global ϕ-R-MLE	Global ϕ-R-WM (S)	Global ϕ-R-WM (I)
	Cannabis					
$\lambda = -1/3$	1.122 (0.597)	1.091 (0.617)	1.492 (0.156)	2.643 (0.661)	2.568 (0.661)	2.914 (0.199)
	0.878 (0.258)	0.836 (0.262)	0.903 (0.133)	2.033 (0.304)	1.981 (0.304)	2.145 (0.162)
	0.758 (0.114)	0.753 (0.112)	0.783 (0.146)	1.762 (0.174)	1.811 (0.177)	1.801 (0.214)
$\lambda = 0$	0.771 (0.370)	0.761 (0.374)	0.779 (0.156)	2.108 (0.395)	2.096 (0.393)	2.202 (0.199)
	0.750 (0.200)	0.731 (0.199)	0.741 (0.133)	1.828 (0.225)	1.823 (0.225)	1.874 (0.162)
	0.840 (0.119)	0.841 (0.119)	0.858 (0.146)	1.865 (0.178)	1.900 (0.179)	1.885 (0.214)
$\lambda = 1/3$	0.456 (0.211)	0.450 (0.206)	0.433 (0.156)	1.744 (0.238)	1.777 (0.232)	1.749 (0.199)
	0.622 (0.150)	0.623 (0.149)	0.621 (0.133)	1.679 (0.171)	1.707 (0.169)	1.680 (0.162)
	0.951 (0.131)	0.956 (0.131)	0.962 (0.146)	2.008 (0.190)	2.027 (0.191)	2.012 (0.214)
$\lambda = 2/3$	0.217 (0.109)	0.242 (0.113)	0.258 (0.156)	1.487 (0.141)	1.521 (0.130)	1.451 (0.199)
	0.516 (0.114)	0.526 (0.115)	0.531 (0.133)	1.571 (0.134)	1.606 (0.129)	1.544 (0.162)
	1.100 (0.152)	1.104 (0.152)	1.102 (0.146)	2.193 (0.215)	2.193 (0.217)	2.190 (0.214)
$\lambda = 1$	0.107 (0.063)	0.132 (0.067)	0.168 (0.156)	1.277 (0.077)	1.284 (0.072)	1.250 (0.199)
	0.439 (0.090)	0.448 (0.093)	0.464 (0.133)	1.480 (0.109)	1.514 (0.104)	1.449 (0.162)
	1.295 (0.185)	1.295 (0.184)	1.290 (0.146)	2.422 (0.255)	2.403 (0.258)	2.432 (0.214)
	Sexual Fun					
$\lambda = -1/3$	0.739 (0.371)	0.625 (0.367)	0.586 (0.298)	1.622 (0.492)	1.790 (0.560)	1.627 (0.467)
	0.357 (0.335)	0.351 (0.335)	0.371 (0.298)	1.138 (0.401)	1.136 (0.428)	1.168 (0.384)
	-0.052 (0.307)	-0.054 (0.317)	-0.050 (0.268)	0.695 (0.407)	0.635 (0.416)	0.761 (0.395)
$\lambda = 0$	0.667 (0.330)	0.597 (0.326)	0.565 (0.298)	1.614 (0.473)	1.717 (0.505)	1.564 (0.467)
	0.339 (0.317)	0.346 (0.322)	0.347 (0.298)	1.138 (0.387)	1.138 (0.402)	1.141 (0.384)
	-0.046 (0.291)	-0.048 (0.300)	-0.044 (0.268)	0.710 (0.402)	0.674 (0.404)	0.739 (0.395)
$\lambda = 1/3$	0.591 (0.300)	0.578 (0.298)	0.559 (0.298)	1.541 (0.447)	1.618 (0.465)	1.527 (0.467)
	0.330 (0.303)	0.332 (0.305)	0.329 (0.298)	1.118 (0.375)	1.138 (0.381)	1.122 (0.384)
	-0.040 (0.274)	-0.042 (0.278)	-0.041 (0.268)	0.710 (0.395)	0.709 (0.390)	0.721 (0.395)
$\lambda = 2/3$	0.522 (0.278)	0.564 (0.284)	0.572 (0.298)	1.420 (0.409)	1.485 (0.435)	1.516 (0.467)
	0.330 (0.291)	0.310 (0.286)	0.316 (0.298)	1.080 (0.365)	1.134 (0.366)	1.109 (0.384)
	-0.036 (0.258)	-0.036 (0.251)	-0.039 (0.268)	0.694 (0.386)	0.736 (0.375)	0.706 (0.395)
$\lambda = 1$	0.466 (0.264)	0.545 (0.281)	0.604 (0.298)	1.306 (0.372)	1.333 (0.409)	1.532 (0.467)
	0.335 (0.283)	0.284 (0.267)	0.308 (0.298)	1.041 (0.357)	1.126 (0.358)	1.103 (0.384)
	-0.032 (0.244)	-0.030 (0.223)	-0.039 (0.268)	0.670 (0.377)	0.754 (0.361)	0.695 (0.395)
	Job Satisfication					
$\lambda = -1/3$	-0.080 (0.362)	- (-)	-0.007 (0.382)	0.290 (0.469)	- (-)	0.351 (0.471)
	0.182 (0.330)	- (-)	0.149 (0.341)	0.449 (0.388)	- (-)	0.514 (0.385)
	0.659 (0.432)	- (-)	∞ (-)	0.853 (0.487)	- (-)	∞ (-)
$\lambda = 0$	-0.071 (0.369)	- (-)	-0.021 (0.382)	0.313 (0.470)	- (-)	0.348 (0.471)
	0.178 (0.333)	- (-)	0.154 (0.341)	0.472 (0.387)	- (-)	0.510 (0.385)
	0.669 (0.407)	- (-)	∞ (-)	0.895 (0.476)	- (-)	∞ (-)
$\lambda = 1/3$	-0.061 (0.377)	- (-)	-0.036 (0.382)	0.340 (0.470)	- (-)	0.345 (0.471)
	0.173 (0.337)	- (-)	0.162 (0.341)	0.498 (0.384)	- (-)	0.507 (0.385)
	0.675 (0.368)	- (-)	0.843 (0.346)	0.946 (0.461)	- (-)	1.052 (0.463)
$\lambda = 2/3$	-0.049 (0.385)	- (-)	-0.055 (0.382)	0.369 (0.467)	- (-)	0.342 (0.471)
	0.168 (0.342)	- (-)	0.172 (0.341)	0.528 (0.380)	- (-)	0.505 (0.385)
	0.664 (0.267)	- (-)	0.621 (0.346)	1.006 (0.430)	- (-)	0.968 (0.463)
$\lambda = 1$	-0.039 (0.394)	-0.026 (0.389)	-0.077 (0.382)	0.396 (0.460)	0.348 (0.459)	0.339 (0.471)
	0.162 (0.349)	0.163 (0.350)	0.186 (0.341)	0.556 (0.371)	0.544 (0.376)	0.504 (0.385)
	0.514 (0.169)	0.470 (0.227)	0.544 (0.346)	1.061 (0.358)	1.326 (0.074)	0.922 (0.463)
	Welfare Spending					
$\lambda = -1/3$	0.115 (0.069)	0.108 (0.067)	0.106 (0.069)	0.160 (0.137)	0.169 (0.135)	0.148 (0.136)
	-0.124 (0.063)	-0.130 (0.065)	-0.130 (0.062)	-0.141 (0.125)	-0.135 (0.126)	-0.144 (0.124)
$\lambda = 0$	0.114 (0.069)	0.110 (0.067)	0.108 (0.069)	0.156 (0.137)	0.163 (0.136)	0.151 (0.136)
	-0.126 (0.063)	-0.129 (0.064)	-0.129 (0.062)	-0.144 (0.124)	-0.138 (0.125)	-0.143 (0.124)
$\lambda = 1/3$	0.114 (0.069)	0.112 (0.068)	0.111 (0.069)	0.152 (0.136)	0.157 (0.137)	0.153 (0.136)
	-0.127 (0.062)	-0.128 (0.063)	-0.128 (0.062)	-0.146 (0.124)	-0.141 (0.124)	-0.142 (0.124)
$\lambda = 2/3$	0.113 (0.069)	0.113 (0.069)	0.113 (0.069)	0.148 (0.136)	0.150 (0.138)	0.155 (0.136)
	-0.129 (0.062)	-0.127 (0.062)	-0.127 (0.062)	-0.149 (0.124)	-0.144 (0.123)	-0.141 (0.124)
$\lambda = 1$	0.113 (0.069)	0.113 (0.070)	0.116 (0.069)	0.144 (0.136)	0.143 (0.139)	0.158 (0.136)
	-0.130 (0.062)	-0.126 (0.061)	-0.127 (0.062)	-0.151 (0.124)	-0.147 (0.122)	-0.140 (0.124)
	Intellectual Capacity Girls					
$\lambda = -1/3$	-0.031 (0.034)	-0.036 (0.034)	-0.083 (0.034)	0.746 (0.080)	0.797 (0.080)	0.861 (0.082)
	0.336 (0.026)	0.343 (0.027)	0.375 (0.023)	1.098 (0.045)	1.149 (0.048)	1.250 (0.045)
	-0.058 (0.025)	-0.057 (0.024)	-0.052 (0.022)	0.852 (0.040)	0.919 (0.042)	0.953 (0.039)
	0.344 (0.027)	0.327 (0.028)	0.343 (0.022)	1.151 (0.042)	1.234 (0.046)	1.330 (0.041)
	0.309 (0.031)	0.286 (0.033)	0.482 (0.025)	1.252 (0.056)	1.322 (0.060)	1.758 (0.057)
$\lambda = 0$	-0.043 (0.035)	-0.046 (0.035)	-0.067 (0.034)	0.794 (0.082)	0.820 (0.082)	0.833 (0.082)
	0.333 (0.025)	0.341 (0.026)	0.341 (0.023)	1.144 (0.045)	1.165 (0.047)	1.197 (0.045)
	-0.053 (0.024)	-0.054 (0.024)	-0.051 (0.022)	0.892 (0.039)	0.922 (0.040)	0.934 (0.039)
	0.321 (0.025)	0.316 (0.025)	0.319 (0.022)	1.194 (0.041)	1.241 (0.043)	1.273 (0.041)
	0.315 (0.031)	0.304 (0.031)	0.343 (0.025)	1.347 (0.057)	1.397 (0.059)	1.562 (0.057)
$\lambda = 1/3$	-0.051 (0.035)	-0.050 (0.034)	-0.056 (0.034)	0.821 (0.081)	0.825 (0.082)	0.816 (0.082)
	0.323 (0.024)	0.326 (0.024)	0.321 (0.023)	1.169 (0.044)	1.165 (0.044)	1.161 (0.045)
	-0.051 (0.023)	-0.051 (0.023)	-0.050 (0.022)	0.919 (0.039)	0.924 (0.038)	0.920 (0.039)
	0.301 (0.023)	0.302 (0.023)	0.301 (0.022)	1.225 (0.040)	1.242 (0.040)	1.233 (0.041)
	0.283 (0.027)	0.281 (0.027)	0.271 (0.025)	1.408 (0.055)	1.437 (0.055)	1.442 (0.057)
$\lambda = 2/3$	-0.050 (0.032)	-0.048 (0.032)	-0.047 (0.034)	0.806 (0.077)	0.809 (0.077)	0.807 (0.082)
	0.308 (0.022)	0.301 (0.022)	0.311 (0.023)	1.144 (0.041)	1.138 (0.041)	1.138 (0.045)
	-0.051 (0.022)	-0.048 (0.022)	-0.049 (0.022)	0.910 (0.037)	0.925 (0.037)	0.909 (0.039)
	0.285 (0.021)	0.285 (0.021)	0.287 (0.022)	1.213 (0.037)	1.227 (0.037)	1.206 (0.041)
	0.161 (0.020)	0.215 (0.021)	0.238 (0.025)	1.371 (0.046)	1.377 (0.045)	1.374 (0.057)
$\lambda = 1$	-0.040 (0.028)	-0.040 (0.028)	-0.040 (0.034)	0.740 (0.069)	0.771 (0.069)	0.806 (0.082)
	0.292 (0.020)	0.268 (0.020)	0.309 (0.023)	1.055 (0.036)	1.082 (0.037)	1.127 (0.045)
	-0.054 (0.022)	-0.044 (0.021)	-0.049 (0.022)	0.841 (0.035)	0.924 (0.036)	0.902 (0.039)
	0.278 (0.020)	0.268 (0.020)	0.277 (0.022)	1.122 (0.033)	1.193 (0.034)	1.190 (0.041)
	0.075 (0.014)	0.145 (0.016)	0.231 (0.025)	1.156 (0.028)	1.202 (0.032)	1.347 (0.057)

Table 6.4: Estimates and standard error estimated (in parentheses) for the ϕ-R-WM based on the local and global R-model on ϕ-scale. Weights are calculated variance optimal based on the independence (I) or saturated (S) model.

	Local			Global		
λ	ϕ-R-MLE	ϕ-R-WM (S)	ϕ-R-WM (I)	ϕ-R-MLE	ϕ-R-WM (S)	ϕ-R-WM (I)
	Streptococcus					
-1/3	0.064	0.000	0.000	0.109	0.000	0.016
0	0.047	0.000	0.000	0.062	0.000	0.000
1/3	0.158	0.000	0.000	0.111	0.000	0.016
2/3	0.282	0.000	0.000	0.162	0.000	0.017
1	0.315	0.000	0.016	0.251	0.000	0.000
	Cannabis					
-1/3	0.281	0.016	0.016	0.484	0.016	0.031
0	0.141	0.016	0.062	0.380	0.016	0.031
1/3	0.961	0.019	0.018	0.477	0.021	0.022
2/3	1.046	0.016	0.018	1.540	0.021	0.023
1	1.351	0.018	0.016	1.383	0.022	0.021
	Sexual Fun					
-1/3	0.201	0.032	0.034	0.228	0.042	0.042
0	0.135	0.032	0.033	0.208	0.041	0.041
1/3	0.183	0.032	0.034	0.242	0.042	0.043
2/3	0.259	0.033	0.033	0.264	0.042	0.042
1	0.271	0.034	0.034	0.367	0.054	0.044
	Job Satisfication					
-1/3	0.290	0.033	0.036	0.297	0.042	0.042
0	0.149	0.034	0.035	0.276	0.042	0.041
1/3	0.310	0.033	0.034	0.302	0.042	0.042
2/3	0.423	0.033	0.032	0.493	0.042	0.045
1	0.513	0.033	0.033	0.536	0.044	0.042
	Welfare Spending					
-1/3	0.231	0.028	0.027	0.916	0.035	0.034
0	0.146	0.028	0.029	0.677	0.037	0.035
1/3	0.196	0.027	0.030	1.884	0.036	0.035
2/3	0.290	0.028	0.026	0.554	0.036	0.035
1	0.410	0.026	0.027	0.455	0.034	0.036
	Intellectual Capacity Girls					
-1/3	1.494	0.287	0.287	4.576	0.357	0.355
0	1.093	0.285	0.279	3.632	0.346	0.344
1/3	1.816	0.291	0.287	4.028	0.356	0.357
2/3	3.048	0.291	0.288	13.318	0.401	0.349
1	4.447	0.289	0.285	13.744	0.353	0.356

Table 6.5: Computation time (seconds) for the ϕ-R-WM based on the local and global R-model on ϕ-scale. Weights are calculated variance optimal based on the independence (I) or saturated (S) model.

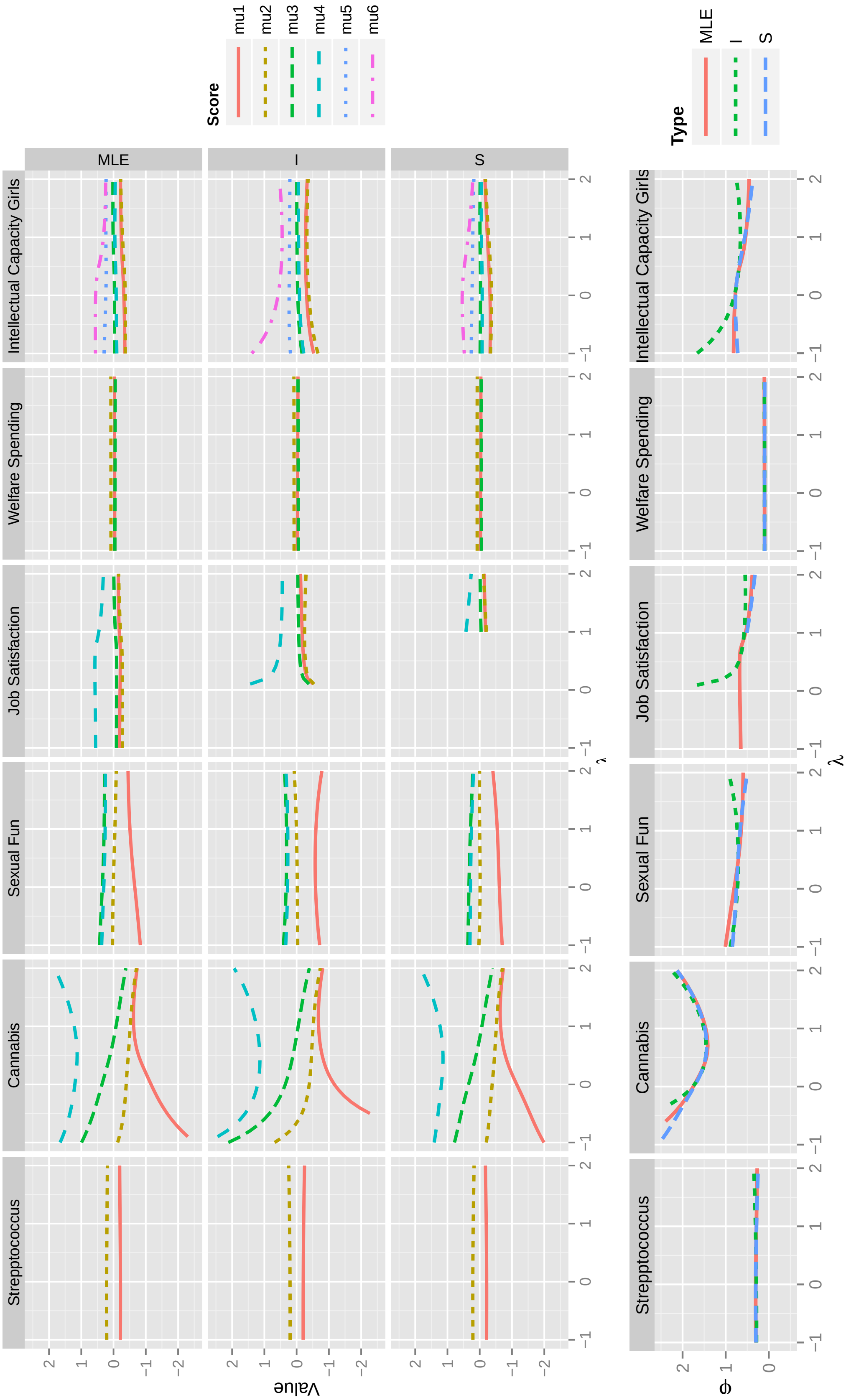


Figure 6.3: Top: Estimated **local** ϕ-R-WM Scores ($\varphi\mu_i$) in the power divergence family ($\lambda \in [-1, 2]$). Variance estimates for the weights are based on the saturated (S) model or the independence (I) model. In addition, scores based on the MLE are calculated. Bottom: Estimated ϕ-R-WM association parameter (φ). ϕ-R-WM variance estimates for the weights are based on the saturated (S) model or the independence (I) model. Again, φ-value under the MLE are given.

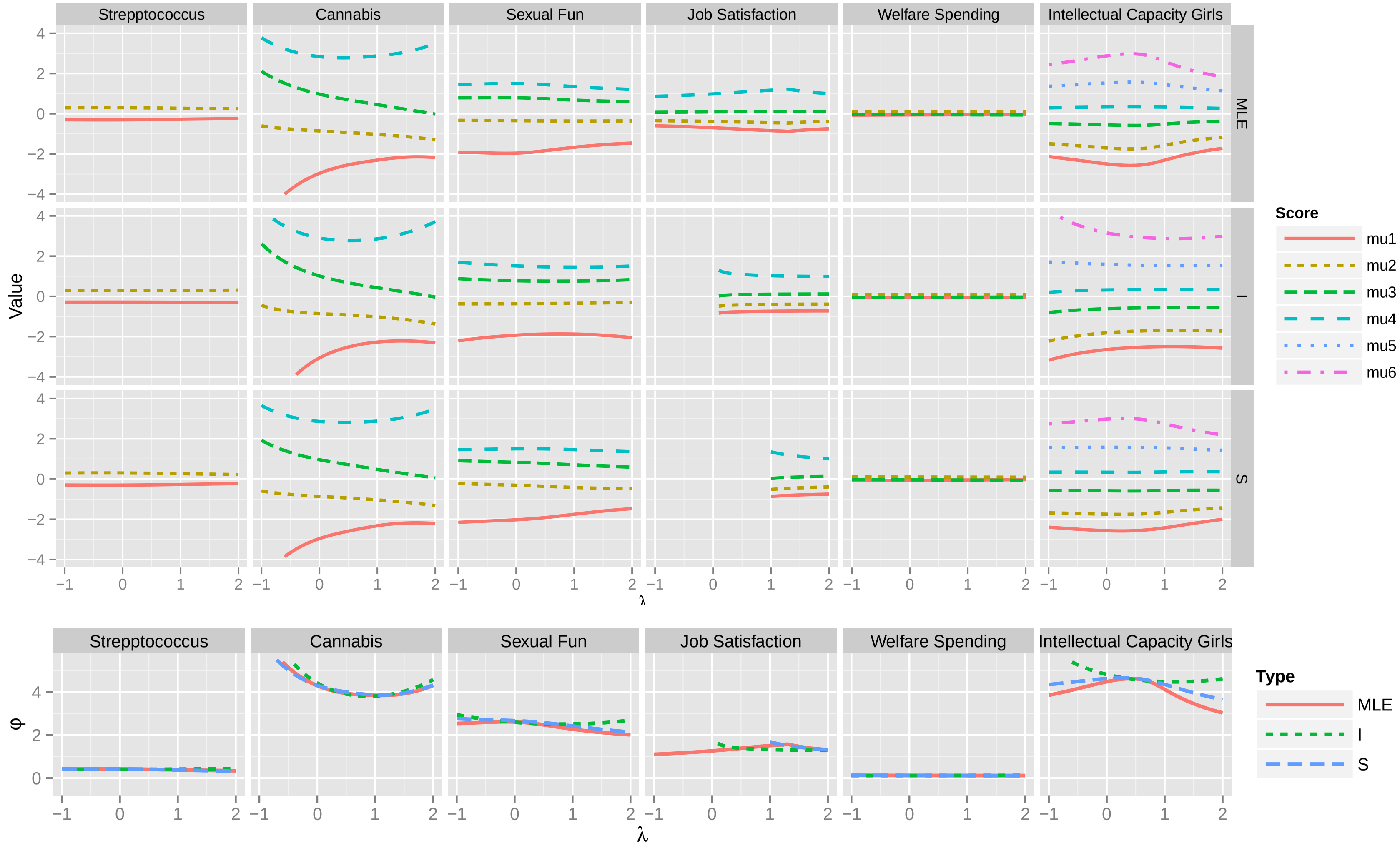


Figure 6.4: Top: Estimated **global** ϕ-R-WM Scores ($\varphi\mu_i$) in the power divergence family ($\lambda \in [-1,2]$). Variance estimates for the weights are based on the saturated (S) model or the independence (I) model. In addition, scores based on the MLE are calculated. Bottom: Estimated ϕ-R-WM association parameter (φ). ϕ-R-WM variance estimates for the weights are based on the saturated (S) model or the independence (I) model. Again, φ-values under the MLE are given.

6.5 ϕ-R-WM Ordering Tests

When measuring association, scores can simplify the analysis as there is a clear connection between the ordering of the scores and stochastic dependencies in the case of the local ϕ-association model, i.e. the model based on local ϕ-scaled odds ratios (Lemma 4.5.4, Lemma 6.4.3). The newly introduced ϕ-R-WM estimators can be used to define an asymptotic test for the ordering of the scores on ϕ-scale. Let $\boldsymbol{\theta}_{ij}^{L} = c_{1i}$ be a ϕ-scaled row effect model with $c_{1i} \geq 0, i = 1, ..., I-1, j = 1, ..., J-1$. Let $\boldsymbol{\mu} = (\mu_i)$ be the corresponding normalized scores which are induced by the solution of the LSE (4.24). Since $c_{i1} \geq 0$ it holds $\mu_{i+1} \geq \mu_i$ for all $i = 1, ..., I-1$ and therefore ordering of the scores can equivalently be expressed as positivity of the ϕ-scaled local odds ratios. Thus, testing the ordering of the ϕ-scaled scores $\boldsymbol{\mu}$ is equivalent to test $\boldsymbol{\theta}^{\phi,L}(\boldsymbol{\pi}) \geq \mathbf{0}$.

This problem can be reduced on a one-sided multivariate testing problem. Techniques using the likelihood ratio test (LRT) are available since the 70's (Kudo [1963]; Nüesch [1966]; Shapiro [1988]). Since the LRT in this case requires advanced computational effort, the union intersection test (UIT) is used in the following. The UIT (briefly introduced in Appendix A.4) is applied to the ϕ-R-WM estimators for testing conditional independence against ordering of the scores. After an example, a simulation study is conducted to give suggestions for the choice of ϕ_λ in the power divergence family (Example 1.10.4).

Let $\mathbf{N} \sim \mathcal{M}(n, \boldsymbol{\pi})$ be a multinomial distributed random variable with realisation $\mathbf{n}$ for an $I \times J$ contingency table for two ordinal classification variables with associated scores. The monotonic ordering of the row scores for ordered column scores is equivalent to the positivity of the local ϕ-R-association parameters ($\theta_{ij}^{\phi,h} = c_{1i} \geq 0$) by the solution of the LSE (4.24). Using the asymptotic normality of the ϕ-R-WM estimator under the R-model (Lemma 6.4.1), the UIT test can be applied to test independence $c_{1i} = 0$ against the positivity of the ϕ-scaled odds ratio $c_{1i} \geq 0, i = 1, ..., I-1$. Let $\boldsymbol{\Sigma}^{\phi,R}(\boldsymbol{\pi}) := \mathbf{A}^T \boldsymbol{\Sigma}^{\phi,L}(\boldsymbol{\pi}) \mathbf{A}$ be the asymptotic variance-covariance matrix of $\hat{\boldsymbol{\theta}}^{\phi,L,RWM} = (\hat{\theta}_1^{\phi,L,RWM}, ..., \hat{\theta}_{I-1}^{\phi,L,RWM})$ as given in (6.14). The test statistic of the ϕ-ordering test becomes

$$\mathbf{T}^{\phi}(\mathbf{n}) = \max_{i=1,...,I-1} T_i(\mathbf{n}), \qquad \text{where} \qquad T_i(\mathbf{n}) = \frac{\hat{\theta}_i^{\phi,h,RWM}(\mathbf{n})}{\sqrt{\boldsymbol{\Sigma}_{ii}^{\phi,R}(\boldsymbol{\pi})}}, \tag{6.18}$$

for $i = 1, ..., I-1$. Again, $\boldsymbol{\Sigma}^{\phi,L}(\boldsymbol{\pi})$ can be estimated using the saturated ($\hat{\boldsymbol{\pi}}$) or independence ($\hat{\boldsymbol{\pi}}^I = (\hat{\pi}_{i+}\hat{\pi}_{+j})$) model. The test rejects independence in favour of ordered scores with significance level $\alpha \in (0,1)$, if $\mathbf{T}^{\phi}(\mathbf{n}) > a_{I-1,\alpha} = \Phi^{-1}((1-\alpha)^{\frac{1}{I-1}})$.

6.5.1 Example

The ϕ-ordering test statistic (6.18) can be calculated using the R-function `UIT.ordering` in Appendix B.6. The values of the test statistic and critical values at significance $\alpha = 0.05$ for the previously used data sets for selected members in the power divergence family (Example 1.10.4) are given in Table 6.6. All but the Job Satisfication data are robust against scale change in the sense that ϕ_λ does not influence the outcome of the ϕ_λ-ordering test in the relevant region $\lambda \in [-1/3, 1]$. In addition, for these data the weight choice

	Strepto.	Cannabis	Sexual Fun	Job Satisf.	Welfare Spending	IC Girls
	2×3	4×3	4×4	4×4	3×5	6×7
Table	4.2	4.3	4.4	4.5	4.6	4.7
$a_{I-1,0.05}$	1.645	2.121	2.121	2.121	1.955	2.319
			$\boldsymbol{\omega}=\boldsymbol{\omega}(S)$			
$\lambda=-1/3$	2.760	6.714	1.704	-	1.625	12.542
$\lambda=0$	2.713	7.088	1.833	-	1.638	13.320
$\lambda=1/3$	2.650	7.289	1.938	-	1.641	13.757
$\lambda=2/3$	2.571	7.275	1.984	-	1.633	13.730
$\lambda=1$	2.481	7.047	1.936	2.072	1.614	13.273
			$\boldsymbol{\omega}=\boldsymbol{\omega}(I)$			
$\lambda=-1/3$	2.490	9.547	1.966	∞	1.552	18.941
$\lambda=0$	2.515	5.861	1.893	∞	1.581	14.669
$\lambda=1/3$	2.555	6.568	1.876	2.438	1.613	13.793
$\lambda=2/3$	2.610	7.525	1.918	1.797	1.648	13.356
$\lambda=1$	2.682	8.810	2.026	1.574	1.686	13.290

Table 6.6: Value of the test statistic $\mathbf{T}^{\phi_\lambda}(\mathbf{n})$ (6.18) and the critical value $a_{I-1,\alpha}, \alpha=0.05$ for the ϕ-ordering test based on the local ϕ-R-WM estimator in the power divergence family (ϕ_λ, Example 1.10.4). Weights are estimated under saturated ($\boldsymbol{\omega}=\boldsymbol{\omega}(S)$) and independence ($\boldsymbol{\omega}=\boldsymbol{\omega}(I)$) model. Independence is rejected in favour of ordered scores if $\mathbf{T}^{\phi_\lambda}(\mathbf{n}) > a_{I-1,\alpha}$.

($\boldsymbol{\omega}=\boldsymbol{\omega}(S)$ or $\boldsymbol{\omega}=\boldsymbol{\omega}(I)$) does not effect the test significantly. The test is not defined in the case of sampling zeros for $\lambda<1$ and $\boldsymbol{\omega}=\boldsymbol{\omega}(S)$ or $\lambda\leq 0$ and $\boldsymbol{\omega}=\boldsymbol{\omega}(I)$. But, as expected due to the higher sensibility to zero cells, the outcome of the test is effected for the Job Satisfication data, rejecting monotonic ordering of the ϕ_λ-scaled scores in favour for independence for $\lambda\geq 2/3$.

6.5.2 Simulation Study

To study the quality of the introduced ϕ-ordering test (Section 6.5) its behaviour is analysed for 3×3 tables with underlying row-effect association. By Remark 1.9.1, a joint probability $\boldsymbol{\pi}\in\Delta_{3\times 3}$ can equivalently be expressed using the triple $\{\log\boldsymbol{\theta}^L, \boldsymbol{\pi}^X, \boldsymbol{\pi}^Y\}$, where $\boldsymbol{\pi}^X\in\Delta_3$ and $\boldsymbol{\pi}^Y\in\Delta_3$ are the marginal distributions of the row and column classification variable, respectively, and where $\log\boldsymbol{\theta}^L=(\log\theta^L_{11},\log\theta^L_{12},\log\theta^L_{21},\log\theta^L_{22})$ are the local log-odds ratios. Under the row-effect model, it holds $\log\theta^L_{11}=\log\theta^L_{12}=c_{11}$ and $\log\theta^L_{21}=\log\theta^L_{22}=c_{12}$. Set $\log\boldsymbol{\theta}^L_R=(c_{11},c_{12})$. Then the joint probability $\boldsymbol{\pi}$ of a row-effect model can equivalently be expressed with the triple $\{\log\boldsymbol{\theta}^L_R,\boldsymbol{\pi}^X,\boldsymbol{\pi}^Y\}$. The probability $\boldsymbol{\pi}$ can be calculated using the algorithm presented in Appendix A.1.

Let $\mathbf{R}^\phi=\{\mathbf{n}\mid \mathbf{T}^\phi(\mathbf{n})>a_{I-1,\alpha}\}$ be the region of the ϕ-ordering test, rejecting independence in favour of the ordering of the scores. The probability $1-\mathbf{P}(\mathbf{R}^\phi)=$

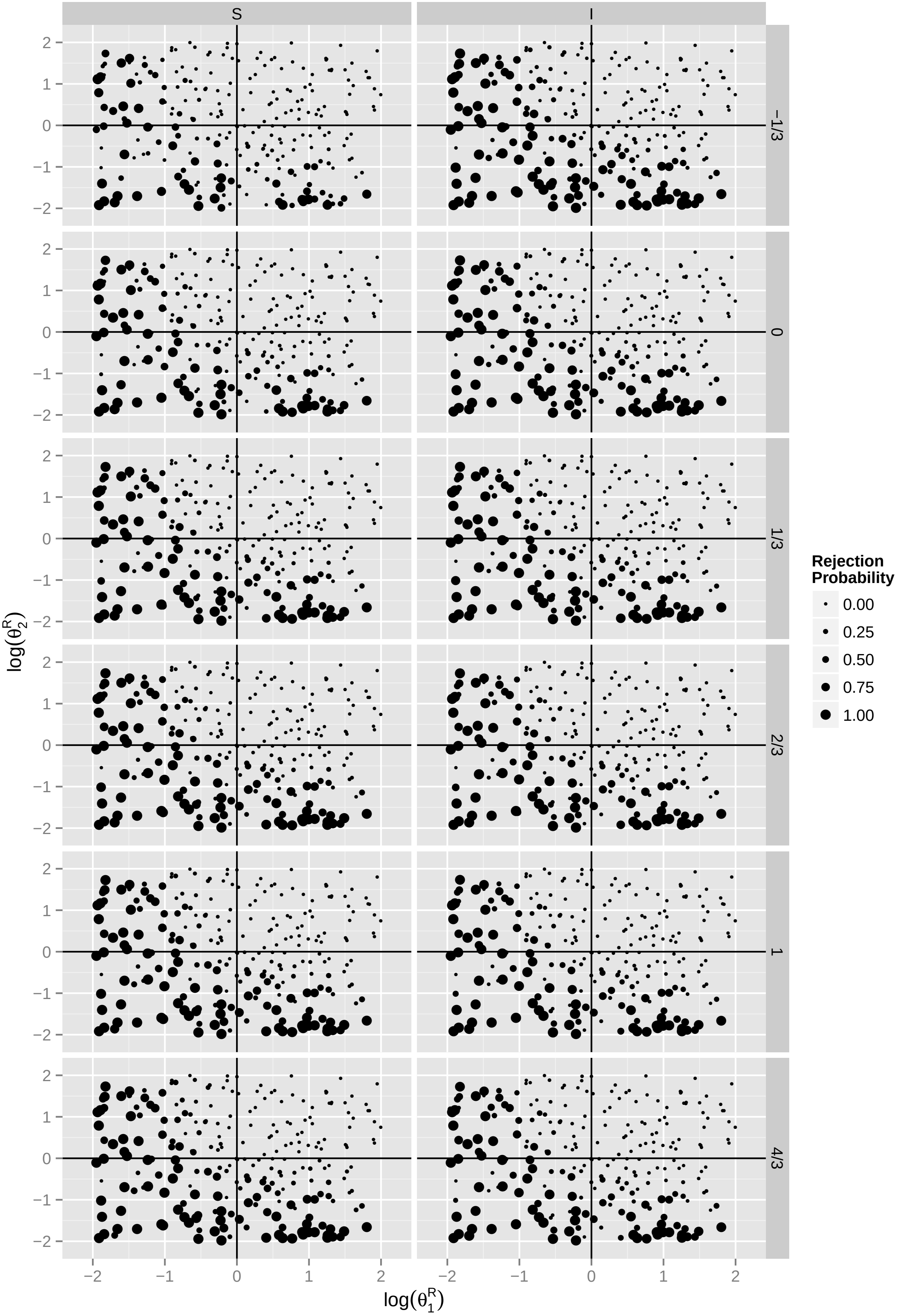


Figure 6.5: Random study for the ϕ-R-WM-ordering test ($\alpha = 5\%$) based on $K_2 = 1000$ generated 3×3 contingency tables ($n = 100$) based on $K_1 = 250$ uniformly generated probabilities $\boldsymbol{\pi} \in \mathcal{R}$ for members in the power divergence family ($\lambda = -\frac{1}{3}, 0, \frac{1}{3}, \frac{2}{3}, 1, \frac{4}{3}$) using single cell correction ($c = 0.5$).

$1-\mathbf{P}(\mathbf{R}^{\phi}|\log\boldsymbol{\theta}_R^L,\boldsymbol{\pi}^X,\boldsymbol{\pi}^Y)$ is calculated under different R-models $\{\log\boldsymbol{\theta}_R^L,\boldsymbol{\pi}^X,\boldsymbol{\pi}^Y\}$ in 3×3 tables. Let $\mathbf{n}_1,...,\mathbf{n}_K$ be a sample of $\mathcal{M}(n,\{\log\boldsymbol{\theta}_R^L,\boldsymbol{\pi}^X,\boldsymbol{\pi}^Y\})$. The probability $1-\mathbf{P}(\mathbf{R}^{\phi})$ can be estimated by

$$1-\hat{\mathbf{P}}(\mathbf{R}^{\phi})=\frac{1}{K}\sum_{k=1}^{K}\mathbf{1}_{\{\mathbf{T}^{\phi}(\mathbf{n}_k)>a_{I-1,\alpha}\}}. \tag{6.19}$$

The simulation study is set up as follows:

1. $K_1=250$ probability models $\{(\log\boldsymbol{\theta}_R^L(i),\boldsymbol{\pi}^X(i),\boldsymbol{\pi}^Y(i))\mid i=1,...,K_1\}$ are sampled, where the parameters $\log\boldsymbol{\theta}_R^L(i)\in[-2,2]^2$, $\boldsymbol{\pi}^X(i),\boldsymbol{\pi}^Y(i)\in\Delta_3$ are uniformly sampled. Notice that multinomial probabilities are uniformly generated using the algorithm presented in the Appendix A.5. Thus, for each $i=1,...,K_1$, one gets a multinomial probability $\boldsymbol{\pi}_i=\boldsymbol{\pi}_i(\log\boldsymbol{\theta}_R^L(i),\boldsymbol{\pi}^X(i),\boldsymbol{\pi}^Y(i))$, fulfilling the row effect model with parameter $\log\boldsymbol{\theta}_R^L(i)$.
2. For each probability $\boldsymbol{\pi}_i$, $K_2=1000$ random samples $\mathbf{n}_1^{(i)},...,\mathbf{n}_{K_2}^{(i)}$ have then been sampled from the distribution $\mathcal{M}(n,\boldsymbol{\pi}_i)$. In addition, single cell correction ($c=0.5$) is applied to also compare ϕ-ordering tests, which are sensitive to sampling zeros.
3. For each $i=1,...,K_1$, the probability (6.19) has been estimated based on the sample $\mathbf{n}_1^{(i)},...,\mathbf{n}_{K_2}^{(i)}$. Call these estimates $p_i, i=1,...,K_1$.
4. Use the sample row-effect parameters $\log\boldsymbol{\theta}_R^L(i)=(\log\theta_1^R,\log\theta_2^R)$ as x- and y-axis of a plot and categorize the p_i's in categories $[0,0.25)$, $[0.25,0.5)$, $[0.5,0.75)$, $[0.75,1)$ and 1, which are graphically represented as points, varying in size according to the category of p_i.

The resulting plot is shown in Figure 6.5 for different members in the power divergence family (Example 1.10.4, $\lambda=-1/3,0,1/3,2/3,1,4/3$) and both weighting schemes (I and S) for the ϕ_λ-R-WM estimator underlying the ϕ_λ-ordering test. To support the analysis, Table 6.7 gives the ordering rejection probability (6.19) of the test for different subregions of $\log\boldsymbol{\theta}_R^L\in[-2,2]^2$. As expected, the ordering is not rejected for positive $\log\boldsymbol{\theta}_R^L$ for all analysed λ and both weighting schemes (I and S). For increasing λ, the test gets a higher selectivity, increasing the rejection probability for negative $\log\boldsymbol{\theta}_R^L$ close to zero. In addition, the weight choice $\boldsymbol{\omega}=\boldsymbol{\omega}(I)$ has a higher rejection probability for non-ordered R-models. Since the simulation study does not show a significant difference between the analysed $\lambda=1$ and $\lambda=4/3$, use of the test based on $\lambda=1$ with weighting scheme $\boldsymbol{\omega}=\boldsymbol{\omega}(I)$ in the ϕ-R-WM estimator is suggested, which does not face the problem of sampling zeros.

6.6 Discussion

Generalized ϕ-linear models were introduced and used to describe the generalised ϕ-association models, giving the possibility to model uniform and row-effect association on ϕ-scale. As seen, a scale change influences the outcome of the model fit and thus gives a greater flexibility when studying association. This effect was already known since the 1980's, when classical association models and correlation models where debated, which are connected by the power divergence as special members of the ϕ_λ-association models. The newly defined GϕLMs unify the theory of association models, by giving the ability to model generalised odds ratios on ϕ-scale.

The closed-form estimator for the association model parameters introduced in Clayton [1974] and extended in Chapter 4 were generalised to ϕ-scale and studied. It turns out that the ϕ-scaled weighted mean estimator is an adequate and computational fast alternative to iterative fitting procedures of the maximum likelihood estimation. The effect of a scale change is greatest, when sampling zeros or small cell frequencies occur.

Since positivity of the local ϕ-scaled odds ratio is equivalent to the ordering of the scores, a union intersection test based on the ϕ-scaled row-effect weighted mean estimates can be constructed to test independence against ordering of the scores. It turns out, that this new ϕ-scaled test performs well for $\lambda = 1$ in the power divergence family without being effected by sampling zeros.

This chapter shows the great potential of the scale changing concept in categorical data analysis. Many problems with the classical CDA tools can be solved by a scale change, for example, the occurrence of sampling zeros is less dramatic on some scales (for example ϕ_P) compared to the log-scale (e.g. ϕ_{KL}). On the other hand, closed-form estimators for generalised ϕ-scaled association model parameters were introduced, which can be used as multidimensional measures of association and can easily be obtained by using matrix calculation.

λ	-1/3	0	1/3	2/3	1	4/3	-1/3	0	1/3	2/3	1	4/3
	I						S					
$\log \boldsymbol{\theta}_R^L$	$n = 100$											
≥ 0	0.000	0.000	0.000	0.001	0.002	0.003	0.002	0.001	0.001	0.000	0.000	0.001
≥ -0.5	0.009	0.014	0.021	0.029	0.037	0.043	0.037	0.029	0.023	0.020	0.019	0.019
≥ -1	0.049	0.071	0.093	0.112	0.127	0.137	0.128	0.112	0.100	0.091	0.087	0.086
≥ -1.5	0.102	0.147	0.185	0.214	0.232	0.241	0.235	0.213	0.195	0.183	0.176	0.177
$\ngeq 0$	0.270	0.365	0.431	0.472	0.495	0.504	0.501	0.473	0.447	0.426	0.415	0.412
$\ngeq -0.5$	0.343	0.463	0.543	0.592	0.616	0.622	0.623	0.591	0.562	0.538	0.524	0.520
$\ngeq -1$	0.440	0.589	0.681	0.731	0.753	0.755	0.763	0.732	0.701	0.675	0.659	0.656
$\ngeq -1.5$	0.549	0.713	0.802	0.844	0.858	0.856	0.867	0.845	0.821	0.796	0.778	0.768
	$n = 250$											
≥ 0	0.008	0.007	0.008	0.010	0.013	0.017	0.012	0.010	0.008	0.008	0.008	0.010
≥ -0.5	0.087	0.090	0.096	0.103	0.110	0.117	0.108	0.103	0.099	0.096	0.095	0.096
≥ -1	0.235	0.241	0.250	0.258	0.264	0.270	0.261	0.256	0.252	0.250	0.249	0.251
≥ -1.5	0.382	0.395	0.408	0.418	0.425	0.428	0.419	0.413	0.409	0.408	0.410	0.414
$\ngeq 0$	0.695	0.714	0.730	0.742	0.748	0.750	0.740	0.735	0.731	0.731	0.733	0.737
$\ngeq -0.5$	0.819	0.841	0.857	0.868	0.871	0.871	0.862	0.858	0.857	0.858	0.862	0.868
$\ngeq -1$	0.858	0.881	0.897	0.908	0.911	0.910	0.900	0.897	0.896	0.899	0.904	0.910
$\ngeq -1.5$	0.942	0.956	0.963	0.967	0.968	0.967	0.962	0.963	0.965	0.966	0.968	0.970
	$n = 500$											
≥ 0	0.005	0.005	0.006	0.010	0.014	0.017	0.011	0.009	0.008	0.007	0.006	0.005
≥ -0.5	0.129	0.131	0.136	0.145	0.154	0.160	0.150	0.146	0.142	0.137	0.135	0.135
≥ -1	0.330	0.340	0.349	0.359	0.367	0.374	0.359	0.355	0.353	0.350	0.350	0.351
≥ -1.5	0.435	0.449	0.460	0.470	0.478	0.483	0.469	0.466	0.464	0.462	0.461	0.461
$\ngeq 0$	0.733	0.752	0.765	0.776	0.784	0.789	0.775	0.772	0.769	0.767	0.767	0.766
$\ngeq -0.5$	0.870	0.894	0.908	0.917	0.921	0.924	0.911	0.909	0.910	0.910	0.910	0.909
$\ngeq -1$	0.937	0.957	0.969	0.975	0.977	0.978	0.973	0.972	0.972	0.971	0.969	0.965
$\ngeq -1.5$	0.979	0.990	0.993	0.996	0.998	0.998	0.996	0.995	0.995	0.994	0.993	0.992

Table 6.7: Simulated rejection probabilities of the ϕ-R-WM-ordering test ($\alpha = 5\%$) based on the random study with each $K_2 = 1000$ 3×3 contingency tables ($n = 100, 250, 500$) randomly generated on $K_1 = 250$ randomly generated probabilities under sample variance (S) or under independence (I) using single cell correction ($c = 0.5$).

Chapter 7

ϕ-based Measures for Asymmetry

In this chapter, square contingency tables with commensurable classification variables are considered. Such tables occur for example in panel studies. The models of symmetry, marginal homogeneity and quasi-symmetry are applicable to such tables. The quasi-symmetry model has been generalised to the ϕ-quasi-symmetry (see Section 1.11). Here, ϕ-scaled measures of asymmetry shall be developed that are related to the ϕ-quasi-symmetry model. In an analogue manner, the ϕ-scaled odds ratios are related to ϕ-association models (cf. (1.81)). In particular, the proposed measures are naturally linked to the asymmetry parameters of the ϕ-quasi-symmetry model (cf. (1.83)).

Basic results on the classical tests for symmetry are briefly reviewed in Section 7.1. For 2×2 tables, the models of symmetry and marginal homogeneity are equivalent and are mostly tested by the McNemar test of symmetry. The McNemar test is generalised to the ϕ-McNemar test in Section 7.2. The properties of the new measure along with its asymptotic inference are studied, while examples and an evaluation study are also provided. The results are generalised to $I \times I$ tables in Section 7.3, leading to a ϕ-scaled generalisation of the symmetry test introduced by Bowker [1948]. The new directed ϕ-scaled measures for asymmetry can be used to define a scalar measure for asymmetry, which is presented in Section 7.4 and compared to the scalar measure of Tomizawa et al. [1998]. Finally, the results are shortly discussed in Section 7.5, which also suggests further fields of study.

7.1 Preliminaries

Let $\boldsymbol{\pi} = (\pi_{ij}) \in \Delta_{I,I}$ be a multinomial probability vector for an $I \times I$ contingency table. Marginal homogeneity holds, if $\pi_{i+} = \pi_{+i}, i = 1, ..., I$, and symmetry holds if $\pi_{ij} = \pi_{ji}, i, j = 1, ..., I$.

However, usually symmetry does not hold and models of symmetry are of interest for example in biomedicine, education and social sciences to measure the asymmetries within the data. For example, *before-and-after studies* are applied to the same group of subjects. The result before and after a treatment are therefore commensurable. Clearly, symmetry is the stronger condition implying marginal homogeneity, i.e. for the before and after group marginal distributions coincide indicating no effect of the treatment. Symmetry models can be used to value the treatment effect by analysing asymmetries.

For $I \times I$ tables, models for symmetry have already been developed and studied (cf. Section 1.9.6). These quasi-symmetry models have been extended using the ϕ-divergence, leading to ϕ-quasi-symmetry models, which are under certain conditions characterized to be the closest models to symmetry in terms of ϕ-divergence (cf. Section 1.11).

Taking differences in the ϕ-quasi-symmetry model (1.82) as given in (1.83) leads to objects of the form:

$$\phi'\left(\frac{\pi_{ij}}{\pi_{ij}^S}\right) - \phi'\left(\frac{\pi_{ji}}{\pi_{ij}^S}\right), \tag{7.1}$$

where $2\pi_{ij}^S = \pi_{ij} + \pi_{ji}, i, j = 1, ..., I$, are the probabilities under symmetry. These objects are also obtained for the ordinal quasi-symmetry models (Kateri and Agresti [2006]). As Kateri and Papaioannou [1997, p. 1126] pointed out, (7.1) occurs as a natural measure of asymmetry and can be used to identify asymmetries among the pairs of the symmetric cells. These resulting objects (7.1), its structure with regard of their measurement abilities and properties to introduce new measures of asymmetry are analysed.

Asymmetry measures have already been defined by Kateri and Papaioannou [2000], Tomizawa [1995] and Tomizawa et al. [1998]. They use the ϕ-divergence $\mathcal{D}^\phi(\boldsymbol{\pi}, \boldsymbol{\pi}^S)$ (see Section 1.10) between the probability $\boldsymbol{\pi}$ and the hypothetical probability under symmetry $\boldsymbol{\pi}^S = (\pi_{ij}^S)$. Applying this ϕ-divergence directly to $I \times I$ tables, the single asymmetry information for each cell in the table is lost. By using the objects occurring in (7.1), another approach using the structural properties can be made, which follow the analysis of the ϕ-scaled odds ratio (Chapter 3) by defining directed ϕ-scaled asymmetry measures. Ideas and methodologies are clearly comparable.

Symmetry in $I \times I$ tables can be tested using the test by Bowker [1948]. For a realisation $\mathbf{n} = (n_{ij})$ it is given by

$$\sum_{i>j} \frac{(n_{ij} - n_{ji})^2}{n_{ij} + n_{ji}}, \tag{7.2}$$

which has a χ^2 distribution with $df = \frac{I(I-1)}{2}$ degrees of freedom under $H_0 : \pi_{ij} = \pi_{ji}, i, j = 1, ..., I$. Bowker's test is the generalisation of the symmetry test introduced by McNemar [1947], which is the first statistical test for symmetry for 2×2 tables. Symmetry is equivalent to marginal homogeneity in this case ($I = 2$). The test statistic (7.2) will be generalised to ϕ-scale by using the directed ϕ-scaled asymmetry measures.

7.2 Directed ϕ-scaled Asymmetry Measures for 2×2 Tables

For 2×2 tables according to (7.1) a directed ϕ-scaled asymmetry measure is defined as follows:

Definition 7.2.1. Let $\boldsymbol{\pi} = (\pi_{ij}) \in \Delta_{2,2}$ and $\phi \in \Phi$ be differentiable. Let $\boldsymbol{\pi}^S = (\pi_{ij}^S), \pi_{ij}^S = \frac{\pi_{ij}+\pi_{ji}}{2}$, be the probability under symmetry corresponding to $\boldsymbol{\pi}$. The *directed ϕ-asymmetry measure* is defined by

$$S^{\phi}(\boldsymbol{\pi}) = \phi'\left(\frac{\pi_{12}}{\pi_{12}^S}\right) - \phi'\left(\frac{\pi_{21}}{\pi_{12}^S}\right) = \phi'\left(\frac{2\pi_{12}}{\pi_{12}+\pi_{21}}\right) - \phi'\left(\frac{2\pi_{21}}{\pi_{12}+\pi_{21}}\right). \tag{7.3}$$

The fraction π_{12}/π_{12}^S resp. π_{21}/π_{12}^S can be called *asymmetry factors* in the same way as Good [1956] introduced the association factors $\pi_{ij}/\pi_{i+}\pi_{+j}$. $S^{\phi}(\boldsymbol{\pi})$ measures the divergence of $\boldsymbol{\pi}$ to the probability of complete symmetry $\boldsymbol{\pi}^S$ by scaling the asymmetry factors with ϕ'. The directed ϕ-asymmetry measure is estimated by $\hat{S}^{\phi} := S^{\phi}(\hat{\boldsymbol{\pi}})$.

One can distinguish between three different symmetries for a probability $\boldsymbol{\pi} = (\pi_{ij}) \in \Delta_{2,2}$ in the 2×2 case: Complete symmetry ($\pi_{12} = \pi_{21}$), positive asymmetry ($\pi_{12} > \pi_{21}$) and negative asymmetry ($\pi_{12} < \pi_{21}$). Depending on the situation, the interpretation of positive or negative can be differed. The measure $S^{\phi} : \Delta_{2,2} \to \mathbb{R}$ should reflect these three symmetries.

Proposition 7.2.2. Let $\boldsymbol{\pi} = (\pi_{ij}) \in \Delta_{2,2}$ and let $\phi \in \Phi$ be differentiable and strictly convex. Then it holds for the directed ϕ-asymmetry measure S^{ϕ}:

1. $S^{\phi}(\boldsymbol{\pi}) = 0$ if and only if $\pi_{12} = \pi_{21}$.
2. For $\boldsymbol{\pi} = [\pi_{ij}] \in \Delta_{2\times 2} \subset (0,1)^{2\times 2}$ it holds $S^{\phi}(\boldsymbol{\pi}^T) = -S^{\phi}(\boldsymbol{\pi})$.
3. For $\epsilon > 0$ define $\boldsymbol{\pi}^{\epsilon} \in \Delta_{2,2}$ by $\boldsymbol{\pi}^{\epsilon} = \boldsymbol{\pi} + (0, \epsilon, -\epsilon, 0)$. It should hold $S^{\phi}(\boldsymbol{\pi}) < S^{\phi}(\boldsymbol{\pi}^{\epsilon})$ for all $\epsilon > 0$, i.e. $|S(\boldsymbol{\pi})|$ increases when asymmetry increases.
4. S^{ϕ} does not depend on the diagonal probabilities π_{11} and π_{22}.
5. S^{ϕ} is invariant under table multiplication.
6. It holds

$$\sup_{\boldsymbol{\pi}\in\Delta_{2,2}} |S^{\phi}(\boldsymbol{\pi})| = \phi'(2) - \phi'(0), \tag{7.4}$$

and thus the directed ϕ-asymmetry measures is bounded if and only if $\phi'(0) := \lim_{t\searrow 0} \phi'(t) > -\infty$.

Proof. 1. Based on the fact that ϕ is strictly convex and thus ϕ' is invertible, 1. is derived straightforward.

2. This is a direct consequence of (7.3).
3. It is straightforward to verify that the probability under symmetry is independent of ϵ, since $(\boldsymbol{\pi}^{\epsilon})^S = \boldsymbol{\pi}^S$. For any $\epsilon > 0$, it holds

$$\frac{\pi_{12}}{\pi_{12}^S} < \frac{\pi_{12}+\epsilon}{\pi_{12}^S} = \frac{\pi_{12}^{\epsilon}}{\pi_{12}^S} \quad \text{and} \quad \frac{\pi_{21}}{\pi_{12}^S} > \frac{\pi_{21}-\epsilon}{\pi_{12}^S} = \frac{\pi_{21}^{\epsilon}}{\pi_{12}^S}.$$

Since ϕ is strictly convex, ϕ' is monotonic increasing. It follows

$$\phi'\left(\frac{\pi_{12}}{\pi_{12}^S}\right) < \phi'\left(\frac{\pi_{12}^{\epsilon}}{\pi_{12}^S}\right) \quad \text{and} \quad \phi'\left(\frac{\pi_{21}}{\pi_{12}^S}\right) > \phi'\left(\frac{\pi_{21}^{\epsilon}}{\pi_{12}^S}\right),$$

such that

$$S^{\phi}(\boldsymbol{\pi}) = \phi'\left(\frac{\pi_{12}}{\pi_{12}^S}\right) - \phi'\left(\frac{\pi_{21}}{\pi_{12}^S}\right) < \phi'\left(\frac{\pi_{12}^{\epsilon}}{\pi_{12}^S}\right) - \phi'\left(\frac{\pi_{21}^{\epsilon}}{\pi_{12}^S}\right) = S^{\phi}(\boldsymbol{\pi}^{\epsilon}).$$

4. It is clear by definition (7.3), that S^{ϕ} does not depend on the diagonal probabilities.
5. Let $\mathbf{n} = (n_{ij})$ be the counts of a 2×2 contingency table. The multiplication invariance can be shown algebraically. For all $k > 0$ it holds

$$\begin{aligned}\hat{S}^{\phi}(\mathbf{n}) &= \phi'\left(\frac{2n_{ij}}{n_{ij}+n_{ji}}\right) - \phi'\left(\frac{2n_{ji}}{n_{ij}+n_{ji}}\right) \\ &= \phi'\left(\frac{2kn_{ij}}{kn_{ij}+kn_{ji}}\right) - \phi'\left(\frac{2kn_{ji}}{kn_{ij}+kn_{ji}}\right) = \hat{S}^{\phi}(k\mathbf{n}).\end{aligned}$$

6. Let $\boldsymbol{\pi} \in \Delta_{2,2}$ and let $\boldsymbol{\pi}^{\epsilon}$ as in 3.. Then $\pi_{21}^{\epsilon} \to 0$ for $\epsilon \to \pi_{21}$. It holds

$$S^{\phi}(\boldsymbol{\pi}^{\epsilon}) \to \phi'(2) - \phi'(0) \qquad \text{for} \qquad \epsilon \to \pi_{21}. \tag{7.5}$$

But since $S^{\phi}(\boldsymbol{\pi}) < S^{\phi}(\boldsymbol{\pi}^{\epsilon})$ by 3. and (7.5) is a monotonic sequence, it holds $\sup_{\boldsymbol{\pi}\in\Delta_{2,2}} |S^{\phi}(\boldsymbol{\pi})|$. The lower bound follows using $S^{\phi}(\boldsymbol{\pi}^T) = -S^{\phi}(\boldsymbol{\pi})$ from property 2..

□

By Proposition 7.2.2 6., the directed ϕ-asymmetry measure is a bounded measure if $\phi'(0) > -\infty$, which for examples holds for the power divergence (Example 1.10.4) for $\lambda > 0$. In such a case a directed asymmetry measure with values in $(-1, 1)$ can be defined by

$$\mathcal{S}^{\phi}(\boldsymbol{\pi}) = \frac{S^{\phi}(\boldsymbol{\pi})}{\phi'(2) - \phi'(0)}.$$

Let $\mathbf{n} = (n_{ij})$ be the counts of a 2×2 contingency table. The MLEs for the symmetry model are $\hat{\pi}_{12}^S = \hat{\pi}_{21}^S = \frac{n_{12}+n_{21}}{2n}$. Under H_0 next to $\hat{\boldsymbol{\pi}} \xrightarrow{P} \boldsymbol{\pi}^S$ it also holds $\hat{\boldsymbol{\pi}}^S \xrightarrow{P} \boldsymbol{\pi}^S$ for $n \to \infty$ (cf. Lemma 1.4.1). This result is useful for calculating the asymptotic distribution of $S^{\phi}(\hat{\boldsymbol{\pi}})$.

Theorem 7.2.3. Let $\phi \in \Phi$ be two times differentiable and strictly convex. Let $n \in \mathbb{N}$ and $\boldsymbol{\pi} = (\pi_{ij}) \in \Delta_{2,2}$ be the parameters of a multinomial distribution $\mathbf{N} \sim \mathcal{M}(n, \boldsymbol{\pi})$. Then $S^{\phi}(\hat{\boldsymbol{\pi}}), \hat{\boldsymbol{\pi}} = \mathbf{N}/n$ is asymptotically normal distributed with mean $S^{\phi}(\boldsymbol{\pi})$ and asymptotic variance

$$\sigma_{S^{\phi}}^2(\boldsymbol{\pi}) = \left(\phi''\left(\frac{2\pi_{12}}{\pi_{12}+\pi_{21}}\right) + \phi''\left(\frac{2\pi_{21}}{\pi_{12}+\pi_{21}}\right)\right)^2 \frac{4\pi_{12}\pi_{21}}{(\pi_{12}+\pi_{21})^3},$$

i.e. it holds

$$\sqrt{n}(S^{\phi}(\hat{\boldsymbol{\pi}}) - S^{\phi}(\boldsymbol{\pi})) \xrightarrow{d} \mathcal{N}(0, \sigma_{S^{\phi}}^2(\boldsymbol{\pi})) \qquad \text{for } n \to \infty.$$

Proof. This is a special case ($I = 2$) of the more general result for $I \times I$ tables in Theorem 7.3.2. □

Example 7.2.4. For the Kullback-Leibler ϕ_{KL} (Example 1.10.2) and Pearson ϕ_P (Example 1.10.3) divergence, it holds

$$S^{\phi_{KL}}(\boldsymbol{\pi}) = \log\left(\frac{\pi_{12}}{\pi_{21}}\right) \qquad \text{and} \qquad S^{\phi_P}(\boldsymbol{\pi}) = \frac{2(\pi_{12}-\pi_{21})}{\pi_{12}+\pi_{21}}$$

with corresponding asymptotic variances

$$\sigma^2_{S^{\phi_{KL}}}(\boldsymbol{\pi}) = \frac{1}{\pi_{12}} + \frac{1}{\pi_{21}} \qquad \text{and} \qquad \sigma^2_{S^{\phi_P}}(\boldsymbol{\pi}) = \frac{16\pi_{12}\pi_{21}}{(\pi_{12}+\pi_{21})^3}, \tag{7.6}$$

respectively.

Lemma 7.2.5. In the set-up of Theorem 7.2.3, the asymptotic variance under $H_0 : \pi_{12} = \pi_{21}$ is estimated by

$$\sigma^2_\phi(\hat{\boldsymbol{\pi}}^S) = \frac{4n}{n_{12}+n_{21}}(\phi''(1))^2.$$

Proof. Inserting the maximum likelihood estimator $\hat{\boldsymbol{\pi}}^S = \left(\frac{n_{11}}{n}, \frac{n_{12}+n_{21}}{n}, \frac{n_{12}+n_{21}}{n}, \frac{n_{22}}{n}\right)$ for the symmetry model gives the result:

$$\begin{aligned}\sigma^2_\phi(\hat{\boldsymbol{\pi}}^S) &= \left(\phi''\left(\frac{\frac{2(n_{12}+n_{21})}{2n}}{\frac{n_{12}+n_{21}}{2n}\frac{n_{12}+n_{21}}{2n}}\right) + \phi''\left(\frac{\frac{2(n_{12}+n_{21})}{2n}}{\frac{n_{12}+n_{21}}{2n}\frac{n_{12}+n_{21}}{2n}}\right)\right)^2 \frac{4\frac{n_{12}+n_{21}}{2n}\frac{n_{12}+n_{21}}{2n}}{\left(\frac{n_{12}+n_{21}}{2n}+\frac{n_{12}+n_{21}}{2n}\right)^3} \\ &= \left(2\phi''(1)\right)^2 \frac{\frac{(n_{12}+n_{21})^2}{n^2}}{\frac{(n_{12}+n_{21})^3}{n^3}} = 4\frac{n}{(n_{12}+n_{21})}\left(\phi''(1)\right)^2.\end{aligned}$$

□

Lemma 7.2.6. In the situation of Theorem 7.2.3 the sequences

$$X^{(n)} = \sqrt{n}\frac{S^\phi(\hat{\boldsymbol{\pi}}) - S_0^\phi}{\sigma_\phi(\boldsymbol{\pi})} \qquad\qquad Z^{(n)} = \sqrt{n}\frac{S^\phi(\hat{\boldsymbol{\pi}}) - S_0^\phi}{\sigma_\phi(\hat{\boldsymbol{\pi}})}$$

are asymptotic standard normal distributed under $H_0 : S^\phi(\boldsymbol{\pi}) = S_0^\phi$. If $S_0^\phi = 0$, the sequence

$$Z_0^{(n)} = \sqrt{n}\frac{S^\phi(\hat{\boldsymbol{\pi}})}{\sigma_\phi(\hat{\boldsymbol{\pi}}^S)}$$

also obtains an asymptotic standard normal distribution under H_0.

Proof. Under $H_0 : S^\phi(\boldsymbol{\pi}) = S_0^\phi$ it holds $\sigma^2_\phi(\hat{\boldsymbol{\pi}}) \xrightarrow{P} \sigma^2_\phi(\boldsymbol{\pi})$ for $n \to \infty$ by the invariance property of the MLEs (cf. Casella and Berger [2002, Theorem 7.2.10, p. 320]), since σ^2_ϕ is a continuous function. Applying Slutsky's Theorem shows that $X^{(n)}$ and $Z^{(n)}$ have the same asymptotic distribution and by Theorem 7.2.3, it holds $X^{(n)} \xrightarrow{d} \mathcal{N}(0,1)$ for $n \to \infty$. The case $H_0 : S^\phi(\boldsymbol{\pi}) = 0$ is equal to symmetry by Proposition 7.2.2 1., such that it holds $\sigma^2_\phi(\hat{\boldsymbol{\pi}}^S) \xrightarrow{P} \sigma^2_\phi(\boldsymbol{\pi}) = \sigma^2_\phi(\boldsymbol{\pi}^S)$ for $n \to \infty$. Again, applying Slutsky's theorem, the asymptotic normality of $Z_0^{(n)}$ is derived. □

Lemma 7.2.6 can be used to define asymptotic tests for symmetry based on the directed ϕ-asymmetry measure. The Wald statistic is

$$Z_\phi = \sqrt{n}\frac{S^\phi(\hat{\boldsymbol{\pi}}) - S_0^\phi}{\sigma_\phi(\hat{\boldsymbol{\pi}})}, \tag{7.7}$$

which obtains an asymptotic standard normal distribution under $H_0 : S^\phi(\boldsymbol{\pi}) = S_0^\phi$. The score statistic is

$$Z_{\phi,0} = \sqrt{n}\frac{S^\phi(\hat{\boldsymbol{\pi}})}{\sigma_\phi(\hat{\boldsymbol{\pi}}^S)}, \tag{7.8}$$

which is asymptotic standard normal distribution under $H_0 : S^\phi(\boldsymbol{\pi}) = 0$.

Corollary 7.2.7. Let ϕ_P be the Pearson divergence ($\phi_P(x) = \frac{1}{2}(x-1)^2$, Example 1.10.3) and $S_0^{\phi_P} = 0$. The test based on $Z_{\phi_P,0}$ (7.8) is the McNemar test.

Proof. It holds

$$\begin{aligned} S^{\phi_P}(\hat{\boldsymbol{\pi}}) &= \frac{2\frac{n_{12}}{n}}{\frac{n_{12}+n_{21}}{n}} - 1 - \frac{2\frac{n_{21}}{n}}{\frac{n_{12}+n_{21}}{n}} + 1 = \frac{2(n_{12}-n_{21})}{n_{12}+n_{21}} \\ \sigma^2_{\phi_P}(\hat{\boldsymbol{\pi}}^S) &= \frac{4n}{n_{12}+n_{21}}, \end{aligned}$$

such that

$$\sqrt{n}\frac{S^{\phi_P}(\hat{\boldsymbol{\pi}})}{\sigma_{\phi_P}(\hat{\boldsymbol{\pi}}^S)} = 2\sqrt{n}\frac{n_{12}-n_{21}}{n_{12}+n_{21}}\sqrt{\frac{n_{12}+n_{21}}{4n}} = \frac{n_{12}-n_{21}}{\sqrt{n_{12}+n_{21}}},$$

which is the test of McNemar [1947]. □

Tests based on $Z_{\phi,0}$ are denominated as ϕ-McNemar tests. The name is justified by Corollary 7.2.7 as they can be regarded as a ϕ-based generalisation of the McNemar test. Since $\sigma^2_\phi(\boldsymbol{\pi}^S)$ is always defined when at least one off-diagonal is non-zero, the ϕ-McNemar test does not face the problem of sampling zeros, where this is the case for the Wald test (7.7), when $\phi'(0) = \lim_{t\searrow 0}\phi'(t) = -\infty$.

Example 7.2.8. Table 7.1 shows the data of a poll under 1600 voting-age British citizens indicating the approval of the Prime Minister's Performance. A second survey was conducted six month after the first under the same participants. Thus 150 citizens changed their opinion from approve to disapprove during this six month, while 86 changed their opinion in the other direction. Table 7.1 also includes the values of the ϕ-McNemar test statistics for different members of the power divergence family (Example 1.10.4).

The critical value to test $H_0 : \pi_{12} = \pi_{21}$ against $H_1 : \pi_{12} \neq \pi_{21}$ is 1.96 at level $\alpha = 5\%$. Thus all tests reject the hypothesis of symmetry with p-value of about 0, providing strong evidence that the approval of the Prime Minister has changed over time.

First Survey	Second Survey	
	Approve	Disapprove
Approve	794	150
Disapprove	86	570

Divergence	Score	Wald
Kullback-Leibler ($\lambda = 0$)	-4.273	-4.113
Power Divergence ($\lambda = 1/3$)	-4.225	-4.207
Power Divergence ($\lambda = 2/3$)	-4.189	-4.279
Pearson ($\lambda = 1$, McNemar)	-4.166	-4.328
Power Divergence ($\lambda = 4/3$)	-4.155	-4.353
Power Divergence ($\lambda = 5/3$)	-4.155	-4.353

Table 7.1: Rating of Performance of Prime Minister and values of the ϕ-McNemar test statistics (7.8) (Score test) and the Wald test statistic (7.7) for different members of the power divergence family. Data is taken from Agresti [2002, p. 409, Table 10.1].

7.2.1 Interpretation and Analysis of Change

The directed ϕ_P-asymmetry measure for the Pearson divergence (Example 1.10.3) has a simple interpretation. It is defined as:

$$S^{\phi_P}(\boldsymbol{\pi}) = \frac{\pi_{12} - \pi_{21}}{\pi_{12}^S} = \frac{2(\pi_{12} - \pi_{21})}{\pi_{12} + \pi_{21}}.$$

Set $S^P = \frac{1}{2} S^{\phi_P}(\boldsymbol{\pi}) = \frac{\pi_{12}-\pi_{21}}{\pi_{12}+\pi_{21}}$. Then, S^P becomes a measure of asymmetry bounded on $[-1, 1]$. Let X and Y be the commensurable row resp. column variable in the 2×2 contingency table with multinomial probability $\boldsymbol{\pi} = (\pi_{ij})$ and probability function $\mathbf{P}$. Clearly, $\mathbf{P}(X \neq Y) = \pi_{12} + \pi_{21}$ such that S^P can be written as:

$$\begin{aligned} S^P &= \frac{\mathbf{P}(X = 1, Y = 2)}{\mathbf{P}(X \neq Y)} - \frac{\mathbf{P}(X = 2, Y = 1)}{\mathbf{P}(X \neq Y)} = \frac{\mathbf{P}(X < Y)}{\mathbf{P}(X \neq Y)} - \frac{\mathbf{P}(X > Y)}{\mathbf{P}(X \neq Y)} \\ &= \mathbf{P}(X < Y \mid X \neq Y) - \mathbf{P}(X > Y \mid X \neq Y). \end{aligned}$$

Thus, S^P is the difference of the conditional probabilities under change. The value $S^P \to 1$ indicates total positive asymmetry ($\mathbf{P}(X < Y|X \neq Y) \to 0$), $S^P \to -1$ indicates total negative asymmetry ($\mathbf{P}(X > Y|X \neq Y) \to 0$) and values of S^P measure the divergence of the conditional probabilities. For example, $S^P = 0.5 \Leftrightarrow \mathbf{P}(X < Y|X \neq Y) = 0.5 + \mathbf{P}(X > Y|X \neq Y)$ means, that the probability for an outcome $X < Y$ is 0.5 points higher than the probability for the outcome $X > Y$ under the condition of change ($X \neq Y$). It holds

$$\mathbf{P}(X > Y|X \neq Y) = \frac{S^P + 1}{2} \qquad \text{and} \qquad \mathbf{P}(X < Y|X \neq Y) = \frac{1 - S^P}{2}.$$

Revisiting the Prime Minister data (Table 7.1), the change of approval can be measured in more detail using inference. Assume, it shall be tested if three times (or more) people changed their opinion to disapprove than people who changed to approve. In terms of conditional probabilities this hypothesis can be formulated as

$$H_0 : \mathbf{P}(X < Y|X \neq Y) \geq 3\mathbf{P}(X > Y|X \neq Y),$$

which is equivalent to $S^p \geq \frac{1}{2}$. Thus the hypothesis test is $H_0 : S^P \geq \frac{1}{2}$ against $H_1 : S^P < \frac{1}{2}$. Using the Wald test statistic (7.7) to execute the one-sided test for $S_0^{\phi_P} = 2S^P = 1$ gives $Z_{\phi_P} = -3.652$ for the Pearson divergence. Since the critical value is $u_{1-\alpha} = 1.64$, the hypothesis $H_0 : S^P \geq \frac{1}{2}$ is rejected in favour of $H_1 : S^P < \frac{1}{2}$ at a significance level of $\alpha = 5\%$ and the probability that people change their opinion from approve to disapprove is less then three times the probability that people changed their opinion from disapprove to approve.

7.2.2 Type I error rates

Following the techniques of Fagerland et al. [2013], one criterion to analyse the quality of the newly defined ϕ-McNemar tests are Type I error rates (TIER). To prevent the problem of sampling zeros, the study is restricted on the test statistic $Z_{\phi,0}$ (7.8) . Let $\mathbf{n}$ be a realisation of a 2×2 table with underlying multinomial random variable $\mathbf{N} \sim \mathcal{M}(n, \boldsymbol{\pi})$ with multinomial probability vector $\boldsymbol{\pi} \in \Delta_{2,2}$, which can be parametrized by the triple of marginal probabilities and odds ratios, i.e. $\{\theta, \pi_{1+}, \pi_{+1}\}$, using (1.19). In 2×2 tables, symmetry $\pi_{12} = \pi_{21}$ holds if and only if $\pi_{1+} = \pi_{+1}$. Let $\mathbf{R}^\phi = \{\mathbf{n} \mid |Z_{\phi,0}| > u_{1-\alpha/2}\}$ be the rejection region of the ϕ-McNemar test. The TIER for fixed $\pi_{1+} = \pi_{+1}$ and $\theta \in (0, \infty)$ is defined as

$$TIER(\pi_{1+}, \theta) = \mathbf{P}(\mathbf{R}^\phi \mid \pi_{1+} = \pi_{+1}, \theta) = \sum_{\mathbf{n}} \mathbf{1}_{\{\mathbf{n}\in\mathbf{R}^\phi\}} \mathbf{P}(\mathbf{N} = \mathbf{n} \mid \pi_{1+} = \pi_{+1}, \theta),$$

where the sum is taken over all realisations of 2×2 contingency tables $\mathbf{n}$ with sample size n. Since focus lies on small sample sizes ($n \leq 100$), complete enumeration is used to calculate the exact Type I error rates for the ϕ-McNemar tests. The TIER is calculated using the R-function `evalf.error0` in Appendix B.7. Results for the 5% significance level can be seen in Figure 7.1 for different members of the power divergence family (Example 1.10.4, $\lambda \in \{2/3, 1, 5/4\}$) and selected multinomial set-ups ($n \in \{20, 40, 60, 100\}, \theta \in \{1, 2, 5\}$) for $\pi_{1+} = \pi_{+1} = 0.01, ..., 0.99$.

The results for the original asymptotic McNemar test, which is the ϕ-McNemar test for the Pearson divergence ($\lambda = 1$), coincide with Fagerland et al. [2013]. In terms of type I error rates, McNemar's test is close to the nominal level for most parameter combinations. The α-level is violated quite often, but only with a small divergence. If the probabilities $\pi_{1+} = \pi_{+1}$ ly next to 0 or 1, the type I error decreases, leading to very conservative tests.

Due to the discreteness of the sample space, tests based on λ's in the immediate vicinity of 1 give the same tests. For example the members $\lambda = 1$ and $\lambda = \frac{4}{3}$ are equivalent in all analysed scenarios. The tests for $\lambda = \frac{2}{3}, 1$ and $\frac{5}{3}$ are studied, which represent the general result. Parameters far away from 1 ($|\lambda - 1| > \frac{2}{3}$) give highly undesirable type I error rates and are not given here. Especially, the symmetry test based on the Kullback-Leibler divergence (Example 1.10.2, $\lambda = 0$) is not suggested.

Following again Fagerland et al. [2013], the TIER for 95 probability scenarios $(n, \theta) \in \{10, 15, ..., 100\} \times \{1, 2, 3, 5, 10\}$ has been calculated for the steps $\pi_{1+} = \pi_{+1} = 0, 0.01, ..., 1$. Table 7.2 shows the summary statistics of the evaluation study for the 9595 elements (n, θ, π_{1+}). The mean and maximum TIER are given together with the proportions for error rates over 0.05 and below 0.03 for the three selected members ($\lambda = \frac{2}{3}, 1, \frac{5}{3}$) of the ϕ-McNemar tests in the power divergence family. To clarify the effect of the sample size n, the sub-scenarios are given in three subregions ($10 \leq n \leq 30, 35 \leq n \leq 60, 65 \leq n \leq 100$). The results of the McNemar test coincides with Fagerland et al. [2013], who pointed out

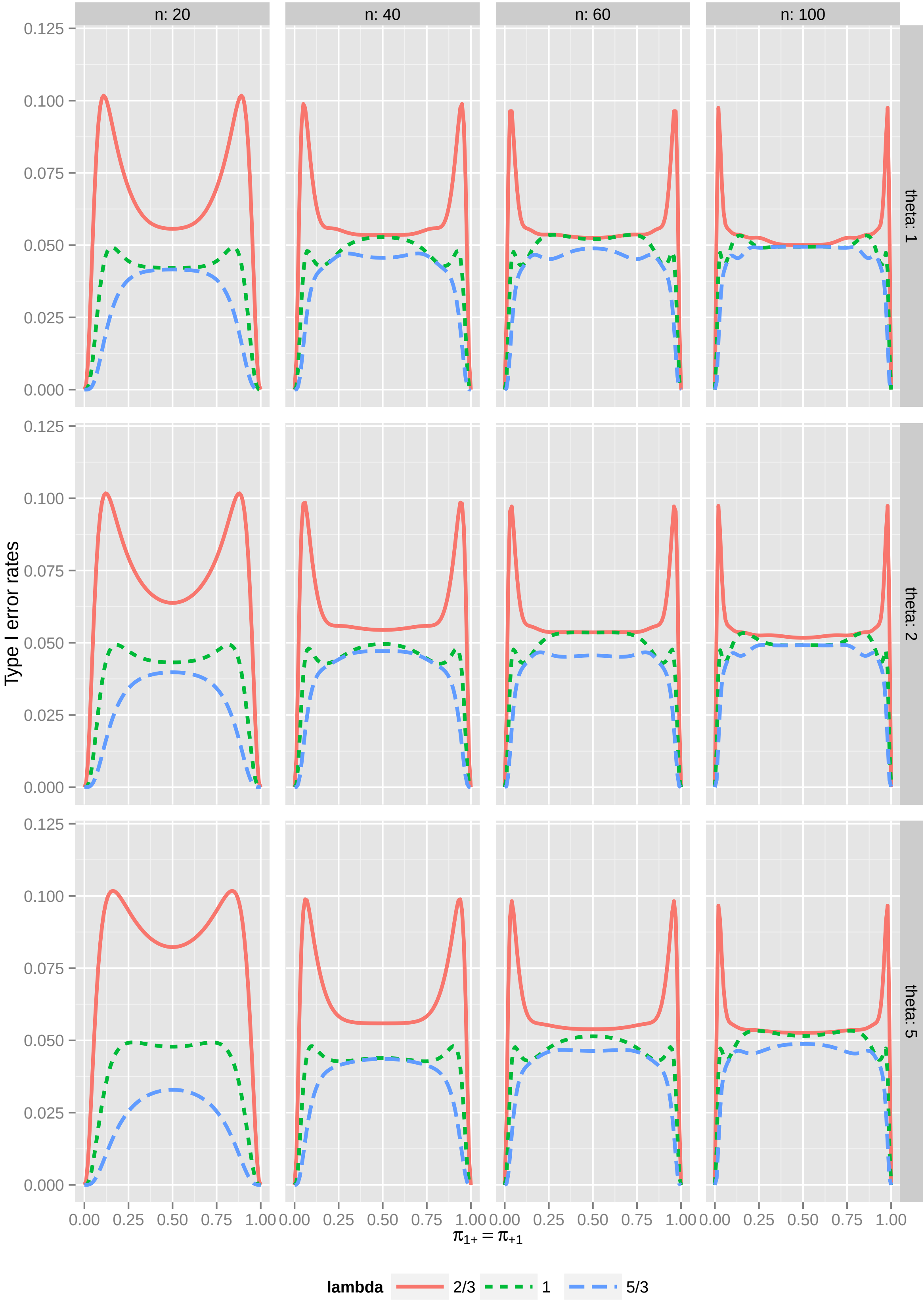


Figure 7.1: Type I error rates for the ϕ-McNemar test for symmetry ($\alpha = 5\%$) for sample sizes $n \in \{20, 40, 60, 100\}$ and odds ratios $\theta \in \{1, 2, 5\}$.

Method	mean TIER	max TIER	proportion TIER > 0.05	proportion TIER < 0.03
McNemar ($\lambda = 1$)	0.0430	0.0537	0.2944	0.1208
McNemar ($\lambda = 2/3$)	0.0617	0.1081	0.9329	0.0496
McNemar ($\lambda = 5/3$)	0.0353	0.0495	0.0000	0.2602
Subregion: $10 \leq n \leq 30$ (2525 scenarios)				
McNemar ($\lambda = 1$)	0.0352	0.0529	0.0372	0.2808
McNemar ($\lambda = 2/3$)	0.0729	0.1081	0.8598	0.1030
McNemar ($\lambda = 5/3$)	0.0212	0.0469	0.0000	0.6269
Subregion: $35 \leq n \leq 60$ (3030 scenarios)				
McNemar ($\lambda = 1$)	0.0435	0.0537	0.2099	0.0838
McNemar ($\lambda = 2/3$)	0.0602	0.0994	0.9498	0.0409
McNemar ($\lambda = 5/3$)	0.0375	0.0489	0.0000	0.1756
Subregion: $65 \leq n \leq 100$ (4040 scenarios)				
McNemar ($\lambda = 1$)	0.0476	0.0535	0.5186	0.0485
McNemar ($\lambda = 2/3$)	0.0558	0.0981	0.9658	0.0228
McNemar ($\lambda = 5/3$)	0.0425	0.0495	0.0000	0.0946

Table 7.2: Evaluation of type I error rates (TIER) over the 19 values of $n \in \{10, 15, 20, ..., 100\}$, the five values $\theta \in \{1, 2, 3, 5, 10\}$, and 101 values $\pi_{1+} = \pi_{+1} \in \{0.00, 0.01, ..., 1.00\}$.

that the McNemar test violates the nominal in 29% of all considered scenarios, where this is the case for only 3.7% for small samples sizes ($10 \leq n \leq 30$) and for over 52% for great sample sizes ($65 \leq n \leq 100$). The maximum TIER is 5.37%, indicating a small distance to the desired nominal.

The ϕ-McNemar test for $\lambda > 1$ is more conservative. The $\lambda = \frac{5}{3}$-test lies always below the McNemar test ($\lambda = 1$) or is at the same nominal level. By Table 7.2, the $\lambda = \frac{5}{3}$-test never violates the nominal with a maximum TIER of 4.95%. As a drawback 26.0% of the values ly below 3%, making the test too conservative. Most of the conservatism is from small samples size scenarios, where 62.7% of the tests have TIER below 3%, this effect decreases with increasing sample size n. The summary statistics support the results drawn from Figure 7.1.

The ϕ-McNemar test for $\lambda < 1$ breaks the α-level for $\pi_{1+} = \pi_{+1}$ close to 0.5 and reaches error rates up to 10% for the $\lambda = \frac{2}{3}$ case and up to 20% for the $\lambda = \frac{1}{3}$ case. There is a peak in the error rates for $\pi_{1+} = \pi_{+1}$ close to 0 and 1. The peak moves closer the edges if the sample size n increases. The ϕ-McNemar test ($\lambda < 1$) violates the nominal quite often. For $\lambda = \frac{2}{3}$ the mean TIER is 6.2% where 93.3% violate the nominal, giving TIER above $\alpha = 5\%$. The maximum TIER is 10.8% which doubles the nominal of $\alpha = 5\%$.

Overall, the TIERs give a very clear suggestion to use the ordinary McNemar test ($\lambda = 1$) to test symmetry. It is the closest to the nominal for different sample sizes, with the drawback to be conservative for values $\pi_{1+} = \pi_{+1}$ close to 0 or 1, which cannot be solved by a ϕ-divergence induced scale change.

7.3 Directed ϕ-Asymmetry Measure for $I \times I$ Tables

The ϕ-symmetry model (1.82) leads to objects of the form (7.1), which can be used to generalise the presented 2×2 directed ϕ-asymmetry measures to $I \times I$ tables. In the same way the McNemar test generalises to a ϕ-scaled test, the test of Bowker [1948] can be generalised to ϕ-scale.

For $\mathbf{x} = (x_{ij}) \in \mathbb{R}^{I^2}$, denote by $(x_{ij})_{i<j}$ the vector of the $d = I(I-1)/2$ upper diagonal values, i.e. $(x_{ij})_{i<j} = (x_{12}, x_{13}, ..., x_{1J}, x_{23}, x_{24}, ..., x_{2J}, x_{34}, ..., x_{I-1J}) \in \mathbb{R}^d$.

Definition 7.3.1. Let $\boldsymbol{\pi} \in \Delta_{I,I}$ be a multinomial probability vector and $\phi \in \Phi$ be differentiable. The *directed ϕ-scaled asymmetry measure* is defined as

$$S^\phi(\boldsymbol{\pi}) = (S^\phi_{ij}(\boldsymbol{\pi}))_{i<j}, \quad \text{where} \quad S^\phi_{ij}(\boldsymbol{\pi}) = \phi'\left(\frac{2\pi_{ij}}{\pi_{ij}+\pi_{ji}}\right) - \phi'\left(\frac{2\pi_{ji}}{\pi_{ij}+\pi_{ji}}\right), \quad i<j.$$

Similar to the generalised association factors (5.2), one can call the terms $\frac{2\pi_{ij}}{\pi_{ij}+\pi_{ji}}$ and $\frac{2\pi_{ji}}{\pi_{ij}+\pi_{ji}}$ *symmetry factors*, which measure the divergence of the cells (i,j) and (j,i) from symmetry. Each $S^\phi_{ij}(\boldsymbol{\pi})$ weights these symmetry factors by applying the monotonic function ϕ'.

Theorem 7.3.2. Let $\phi \in \Phi$ be two times differentiable and strictly convex. Let $\mathbf{N} \sim \mathcal{M}(n, \boldsymbol{\pi})$ with $\boldsymbol{\pi} \in \Delta_{I,I}$ and MLE $\hat{\boldsymbol{\pi}} = \mathbf{N}/n$. Then $S^\phi(\hat{\boldsymbol{\pi}})$ is asymptotic normal distributed with mean $S^\phi(\boldsymbol{\pi})$ and variance-covariance matrix

$$\boldsymbol{\Sigma}^{S^\phi}(\boldsymbol{\pi}) = \text{diag}((\rho^{S^\phi}_{ij}(\boldsymbol{\pi}))_{i<j}), \tag{7.9}$$

where

$$\rho^{S^\phi}_{ij}(\boldsymbol{\pi}) = \frac{4\pi_{ij}\pi_{ji}}{(\pi_{ij}+\pi_{ji})^3}\left(\phi''\left(\frac{2\pi_{ij}}{\pi_{ij}+\pi_{ji}}\right) + \phi''\left(\frac{2\pi_{ji}}{\pi_{ij}+\pi_{ji}}\right)\right)^2, \qquad i<j, \tag{7.10}$$

i.e.

$$\sqrt{n}(S^\phi(\hat{\boldsymbol{\pi}}) - S^\phi(\boldsymbol{\pi})) \xrightarrow{d} \mathcal{N}(0, \boldsymbol{\Sigma}^{S^\phi}(\boldsymbol{\pi})) \qquad \text{for } n \to \infty.$$

Proof. It holds

$$\frac{\partial}{\partial \pi_{kl}} S^\phi_{ij}(\boldsymbol{\pi}) = \begin{cases} \frac{2\pi_{ji}}{(\pi_{ij}+\pi_{ji})^2}\left(\phi''\left(\frac{2\pi_{ij}}{\pi_{ij}+\pi_{ji}}\right) + \phi''\left(\frac{2\pi_{ji}}{\pi_{ij}+\pi_{ji}}\right)\right) & \text{if } (k,l) = (i,j) \\ -\frac{2\pi_{ij}}{(\pi_{ij}+\pi_{ji})^2}\left(\phi''\left(\frac{2\pi_{ij}}{\pi_{ij}+\pi_{ji}}\right) + \phi''\left(\frac{2\pi_{ji}}{\pi_{ij}+\pi_{ji}}\right)\right) & \text{if } (k,l) = (j,i) \\ 0 & \text{else.} \end{cases} \tag{7.11}$$

Now, ϕ is strictly convex and $\phi'' > 0$, such that $\frac{\partial S^\phi}{\partial \boldsymbol{\pi}} \neq \mathbf{0}$. The delta method (Theorem 1.4.3) can be applied. Some algebra shows

$$\left(\frac{\partial S^\phi(\boldsymbol{\pi})}{\partial \boldsymbol{\pi}}\right) \text{diag}(\boldsymbol{\pi}) \left(\frac{\partial S^\phi(\boldsymbol{\pi})}{\partial \boldsymbol{\pi}}\right)^T = \text{diag}((\rho^{S^\phi}_{ij}(\boldsymbol{\pi}))_{i<j})$$

and

$$\left(\frac{\partial S^\phi(\boldsymbol{\pi})}{\partial \boldsymbol{\pi}}\right) \boldsymbol{\pi} = \left(\frac{2\pi_{ij}\pi_{ji}}{(\pi_{ij}+\pi_{ji})^2} - \frac{2\pi_{ji}\pi_{ij}}{(\pi_{ij}+\pi_{ji})^2}\right)\left(\phi''\left(\frac{2\pi_{ij}}{\pi_{ij}+\pi_{ji}}\right) + \phi''\left(\frac{2\pi_{ji}}{\pi_{ij}+\pi_{ji}}\right)\right)_{i<j}$$

$$= \mathbf{0},$$

which gives the result with asymptotic variance-covariance matrix

$$\boldsymbol{\Sigma}^{S^\phi}(\boldsymbol{\pi}) = \left(\frac{\partial S^\phi(\boldsymbol{\pi})}{\partial \boldsymbol{\pi}}\right) \text{diag}(\boldsymbol{\pi}) \left(\frac{\partial S^\phi(\boldsymbol{\pi})}{\partial \boldsymbol{\pi}}\right)^T - \left(\frac{\partial S^\phi(\boldsymbol{\pi})}{\partial \boldsymbol{\pi}}\right) \boldsymbol{\pi}\boldsymbol{\pi}^T \left(\frac{\partial S^\phi(\boldsymbol{\pi})}{\partial \boldsymbol{\pi}}\right)^T$$
$$= \text{diag}((\rho_{ij}^{S^\phi}(\boldsymbol{\pi}))_{i<j}).$$

□

Asymptotic tests for symmetry based on the ϕ-scaled asymmetry measures can be derived using the asymptotic properties of $S^\phi(\hat{\boldsymbol{\pi}})$. To test $H_0 : S^\phi(\boldsymbol{\pi}) = \mathbf{S}^{(0)}$ against $H_1 : S^\phi(\boldsymbol{\pi}) \neq \mathbf{S}^{(0)}$, $\mathbf{S}^{(0)} = (S_{ij}^{(0)})_{i<j} \in \mathbb{R}^{I(I-1)/2}$, the two asymptotic test statistics are

$$W_{S^\phi}^2 = n(S^\phi(\hat{\boldsymbol{\pi}}) - \mathbf{S}^{(0)})^T \left(\boldsymbol{\Sigma}^{S^\phi}(\hat{\boldsymbol{\pi}})\right)^{-1} (S^\phi(\hat{\boldsymbol{\pi}}) - \mathbf{S}^{(0)}) = \sum_{i<j} \frac{(S_{ij}^\phi(\hat{\boldsymbol{\pi}}) - S_{ij}^{(0)})^2}{\rho_{ij}^{S^\phi}(\hat{\boldsymbol{\pi}})/n} \tag{7.12}$$

$$W_{S^\phi,0}^2 = n(S^\phi(\hat{\boldsymbol{\pi}}))^T \left(\boldsymbol{\Sigma}^{S^\phi}(\hat{\boldsymbol{\pi}}^S)\right)^{-1} S^\phi(\hat{\boldsymbol{\pi}}) = \sum_{i<j} \frac{(S_{ij}^\phi(\hat{\boldsymbol{\pi}}))^2}{\rho_{ij}^{S^\phi}(\hat{\boldsymbol{\pi}}^S)/n} \quad \text{for} \quad \mathbf{S}^{(0)} = \mathbf{0}. \tag{7.13}$$

They are asymptotic χ^2 distributed with $df = \frac{I(I-1)}{2}$ degrees of freedom. This is a direct consequence of Theorem 7.3.2 using the techniques of Lemma 7.2.6. Remark that $\boldsymbol{\Sigma}^{S^\phi}(\hat{\boldsymbol{\pi}})$ from (7.9) does not exist when $\phi''(0) = -\infty$ and $\hat{\boldsymbol{\pi}}$ has sampling zeros, which is not the case for $\boldsymbol{\Sigma}^{S^\phi}(\hat{\boldsymbol{\pi}}^S)$ when there is no pair (i, j) such that $\hat{\pi}_{ij} = 0 = \hat{\pi}_{ji}$.

Corollary 7.3.3. Let ϕ_P be the Pearson divergence (Example 1.10.3) and $\mathbf{S}^{(0)} = \mathbf{0}$. The test based on $W_{S^{\phi_P},0}^2$ is the test of Bowker [1948].

Proof. The same calculation as in Corollary 7.2.7 for the 2×2 case, shows that

$$W_{S^{\phi_P},0}^2 = \sum_{i<j} \frac{(S_{ij}^{\phi_P}(\hat{\boldsymbol{\pi}}))^2}{\rho_{ij}^{S^{\phi_P}}(\hat{\boldsymbol{\pi}}^S)/n} = \sum_{i<j} \frac{(n_{ij} - n_{ji})^2}{n_{ij} + n_{ji}} = \sum_{i,j=1}^{I,I} \frac{(n_{ij} - n\hat{\pi}_{ij}^S)^2}{n\hat{\pi}_{ij}^S} = X^2,$$

which is the Bowker test statistic and also the classic Pearson goodness-of-fit statistic X^2 (1.10) for the symmetry model (1.62). □

Thus Corollary 7.3.3 motivates the designation of $W_{S^\phi,0}^2$ as the ϕ-Bowker test statistic. In the case of the Kullback-Leibler divergence (Example 1.10.2), the ϕ_{KL}-Bowker test statistic becomes

$$W_{S^{\phi_{KL}},0}^2 = \sum_{i<j} \frac{n_{ij} + n_{ji}}{4} (\log(n_{ij}) - \log(n_{ji}))^2,$$

while for the LR test statistic G^2 (1.8), it holds for the symmetry model:

$$G^2 = \sum_{i,j=1}^{I,J} n_{ij} \log\left(\frac{n_{ij}}{\hat{\mu}_{ij}}\right) = \sum_{i<j} \left(n_{ij} \log n_{ij} - (n_{ij} + n_{ji}) \log\left(\frac{n_{ij} + n_{ji}}{2}\right) + n_{ji} \log n_{ji}\right), \tag{7.14}$$

where the expected cell frequencies are $\hat{\mu}_{ij}^S = n\hat{\pi}_{ij}^S = \frac{n_{ij}+n_{ji}}{2}, i, j = 1, ..., I$. Therefore, the new proposed test statistics (7.12) and (7.13) differ from the classical goodness-of-fit tests (Section 1.5) for the symmetry model with an exception in the case of the Pearson divergence.

Remark 7.3.4. R-functions for Section 7.3 are found in Appendix B.7. The directed ϕ-scaled asymmetry measure (Definition 7.3.1) is obtained using `S.phi.I`. The derivatives (7.11) is calculated using `dS`, while the matrix $\partial S^\phi(\boldsymbol{\pi})/\partial\boldsymbol{\pi}$ is given in `symmetry.derivation.matrix`. The test statistic (7.12) is calculated using `W_S`, while the ϕ-Bowker test is calculated by `W_S0`. The value of the LR test for symmetry (7.14) is calculated using `LR.sym`.

Example 7.3.5. Table 7.3 shows a sample of 55,981 U.S. residents comparing the residence in 1980 and 1985. The ϕ-Bowker tests based on $W^2_{S\phi}$ and $W^2_{S\phi,0}$ have been applied for different members of the power divergence family (Example (5.2), $\lambda \in \{0, 1/3, ..., 5/3\}$). The critical value of the χ^2 distribution with $df = \frac{4\cdot 3}{2} = 6$ at $\alpha = 5\%$ is $\chi^2_\alpha = 12.59$. Thus, the symmetry hypothesis is clearly rejected by all tests. The residuals

$$\left(\frac{S_{ij}^\phi(\hat{\boldsymbol{\pi}})}{\sqrt{\rho_{ij}^{S\phi}(\hat{\boldsymbol{\pi}}^S)/n}}\right)_{i<j}$$

based on the Pearson divergence ($\lambda = 1$) are:

$$\left(\frac{n_{ij}-n_{ji}}{\sqrt{n_{ij}+n_{ji}}}\right)_{i<j} = (0.95, 8.36, 4.46, 9.37, 5.76, -0.68).$$

Thus, the cells $(1,3), (1,4), (2,3)$ and $(2,4)$ seem to diverge from symmetry, significantly, while the cells $(1,2)$ and $(3,4)$ behave as expected under symmetry, giving non-significant residuals. In addition Table 7.4 shows the migration of 1,855 American adults comparing their residence at the age of 16 and in 2010. Sample size is much smaller than in the previous example. While the ϕ-Bowker tests still reject the hypothesis of symmetry at $\alpha = 5\%$ for all selected ϕ_λ, the choice of λ seems to have a greater effect if the sample size is small, more than doubling the values of the test statistic for the Kullback Leibler divergence (70.518) compared to the Pearson (188.759). Thus choice of λ is sensible.

Residence in 1980	Residence in 1985: Northeast	Midwest	South	West
Northeast	11,607	100	366	124
Midwest	87	13,677	515	302
South	172	255	17,819	270
West	63	176	286	10,192

λ	$W^2_{S\phi}$	$W^2_{S\phi,0}$
0	203.925	229.448
1/3	218.560	221.576
2/3	230.475	215.894
1	238.832	212.224
4/3	243.068	210.437
5/3	242.998	210.451

Table 7.3: Migration from 1980 to 1985 (left) and values of the test statistics, $W^2_{S\phi}$ (7.12) and $W^2_{S\phi,0}$ (7.13). Data based on Table 12 of U.S. Bureau of the census, Corrent Population Reports, Series P-20, No. 420, Geographical Mobility: 1985 (Washington, DC: U.S. Government Printing Office cited by Agresti [2002, p. 423, Table 10.6]).

Residence	Residence in 2010			
at Age 16	Northeast	Midwest	South	West
Northeast	266	15	61	28
Midwest	10	414	50	40
South	8	22	578	22
West	7	6	27	301

λ	$W^2_{S^\phi}$	$W^2_{S^\phi,0}$
0	70.518	143.064
1/3	104.406	115.650
2/3	147.603	99.615
1	188.759	90.840
4/3	211.418	87.110
5/3	208.929	87.278

Table 7.4: Migration of American adults comparing their residence at the age of 16 and in 2010 (left) and values of the test statistics, $W^2_{S^\phi}$ (7.12) and $W^2_{S^\phi,0}$ (7.13). Data is from the U.S. General Social Survey 2010 cited by Agresti [2013, p. 425, Table 11.5].

7.4 A Directed Scalar Measure for Asymmetry

As in the case of association measures (Section 1.6), it is very useful to have a one-dimensional measure for asymmetry, which values the asymmetry on an absolute scale $[0,1]$ or a directed scale $[-1,1]$. Tomizawa et al. [1998] introduced a power divergence $\mathcal{D}^{\phi_\lambda}$ (1.74) based $d=\frac{I(I-1)}{2}$-dimensional measure $\mathbf{\Phi}^{(\lambda)}=(\Phi^{(\lambda)}_{ij})_{i<j}$ with entries

$$\Phi^{(\lambda)}_{ij}=\frac{\lambda(\lambda+1)}{2^\lambda-1}\mathcal{D}^{\phi_\lambda}((\pi^c_{ij},\pi^c_{ji}),(0.5,0.5)),\qquad i<j, \tag{7.15}$$

where $\pi^c_{ij}=\frac{\pi_{ij}}{\pi_{ij}+\pi_{ji}}$ and $\pi^c_{ji}=\frac{\pi_{ji}}{\pi_{ij}+\pi_{ji}}$. He also introduced a scalar measure

$$\Phi^{(\lambda)}=\frac{\lambda(\lambda+1)}{2^\lambda-1}\mathcal{D}^{\phi_\lambda}((\boldsymbol{\pi}^*),(\boldsymbol{\pi}^*)^S)), \tag{7.16}$$

where $\boldsymbol{\pi}^*=(\pi^*_{ij}),\pi^*_{ij}=\pi_{ij}\delta^{-1},\delta=\sum_{i\neq j}\pi_{ij}$ and $(\pi^*_{ij})^S=\frac{\pi^*_{ij}+\pi^*_{ji}}{2}$. Using the upper bound of $\mathcal{D}^{\phi_\lambda}$ (Pardo [2006, p. 8, Proposition 1.1]), it is easy to see that $0\le\Phi^{(\lambda)}\le 1$. In addition, it holds

$$\hat{\Phi}^{(\lambda)}=\frac{\lambda(\lambda+1)}{(2^\lambda-1)n^*}n\mathcal{D}^{\phi_\lambda}(\hat{\boldsymbol{\pi}},\hat{\boldsymbol{\pi}}^S)=\frac{\lambda(\lambda+1)}{(2^\lambda-1)}\mathcal{D}^{\phi_\lambda}(\hat{\boldsymbol{\pi}}^*,(\hat{\boldsymbol{\pi}}^*)^S),$$

where $n^*=\sum_{i\neq j}n_{ij}$ and $\hat{\Phi}^{(\lambda)}$ is a weighted function of the power divergence between the off-diagonal sample proportions and the off-diagonal probabilities estimated under symmetry and therefore directly based on the ϕ-divergence (1.68). As Tomizawa et al. [1998, p. 396] pointed out that $\hat{\Phi}$ is useful when comparing departure from symmetry in several different tables due to its restricted value range.

The previously introduced directed ϕ-scaled asymmetry measure (Definition 7.3.1) uses $d=I(I-1)/2$ dimensions to describe the divergence from symmetry and obtain an upper bound for $\phi\in\Phi$ if $\phi'(0)>-\infty$. A directed scalar using $S^\phi(\boldsymbol{\pi})=(S^\phi_{ij}(\boldsymbol{\pi}))_{i<j}$ can be based on

$$\tilde{S}_\phi=\operatorname{sign}\left(\sum_{i<j}S^\phi_{ij}(\hat{\boldsymbol{\pi}})\right)\cdot\sum_{i<j}(S^\phi_{ij}(\hat{\boldsymbol{\pi}}))^2.$$

Since by Proposition 7.2.2 $|S_{ij}^{\phi}(\boldsymbol{\pi})| < \phi'(2) - \phi'(0)$ for all $\boldsymbol{\pi} \in \Delta_{I,I}$, it holds

$$|\tilde{S}_{\phi}| < \sum_{i<j} (S_{ij}^{\phi}(\hat{\boldsymbol{\pi}}))^2 < \frac{I(I-1)}{2}(\phi'(2) - \phi'(0))^2 =: c, \qquad \phi'(0) > -\infty.$$

Thus, a directed normalized scalar measure for asymmetry is obtained by setting

$$S_{\phi}^{*} := \frac{\tilde{S}_{\phi}}{c}. \tag{7.17}$$

This scalar measure takes values in $(-1, 1)$, where 1 (-1) is obtained for tables with lower (upper) triangular equal to zero. In additional, $S_{\phi}^{*} = 0$ holds if and only if symmetry holds.

Remark 7.4.1. The R-function `tomi.scalar` and `scalar.S` for calculating (7.17) and (7.16), respectively, are found in the Appendix B.7.

Table 7.5 shows the results of three consecutive opinion pools held in August 1971, October 1971 and December 1973, which were held in connection with the Danish referendum on whether or not to join the European Common Market (Tomizawa et al. [1998, p. 390]). Figure 7.2 shows the values of the presented measures ((7.16) and (7.17)) and values of the test statistics (7.12) and (7.13) for the data sets of Table 7.3, 7.4 and 7.5. Notice that for the Poll II vs Poll III data set, the measure S_{ϕ}^{*} is negative. As seen in Example 7.3.5, the both test statistics based on S^{ϕ} behave parabolic with a different monotonicity for increasing parameter in the power divergence. While the statistic $W_{S_{\lambda}^{\phi}}^{2}$ is concave, the test statistic $W_{S_{\lambda}^{\phi},0}^{2}$ is a convex function of λ. The introduced S_{ϕ}^{*} scalar measure gives similar (absolute) values compared to Tomizawa's measure, especially for power divergence parameter $\lambda \in [1, 1.5]$. For example, for the Poll I/II data, $\Phi^{(1)} = 0.478$ and $S_{\phi_1}^{*} = 0.489$, while for the Poll II/III data $\Phi^{(1)} = 0.268$ and $S_{\phi_1}^{*} = -0.263$.

Table 7.5: Results from the (a) first and second polls and (b) second and third polls on the question: Do you think Denmark should join the European Common Market?

(a)

	Poll II		
	Yes	Undecided	No
Yes	176	33	40
Undecided	21	94	32
No	21	33	40

(b)

	Poll III		
	Yes	Undecided	No
Yes	167	36	15
Undecided	19	131	10
No	45	50	20

7.5 Discussion

This chapter presented additional generalisations of categorical data analysis tools based on the ϕ-divergence (Section 1.10) and demonstrated the flexibility of the concept.

Based on the ϕ-quasi-symmetry model of Kateri and Papaioannou [1997] and the ϕ-ordinal quasi-symmetry model of Kateri and Agresti [2006], a directed ϕ-scaled asymmetry measure for 2×2 and $I \times I$ tables has been introduced. Use of its asymptotic properties

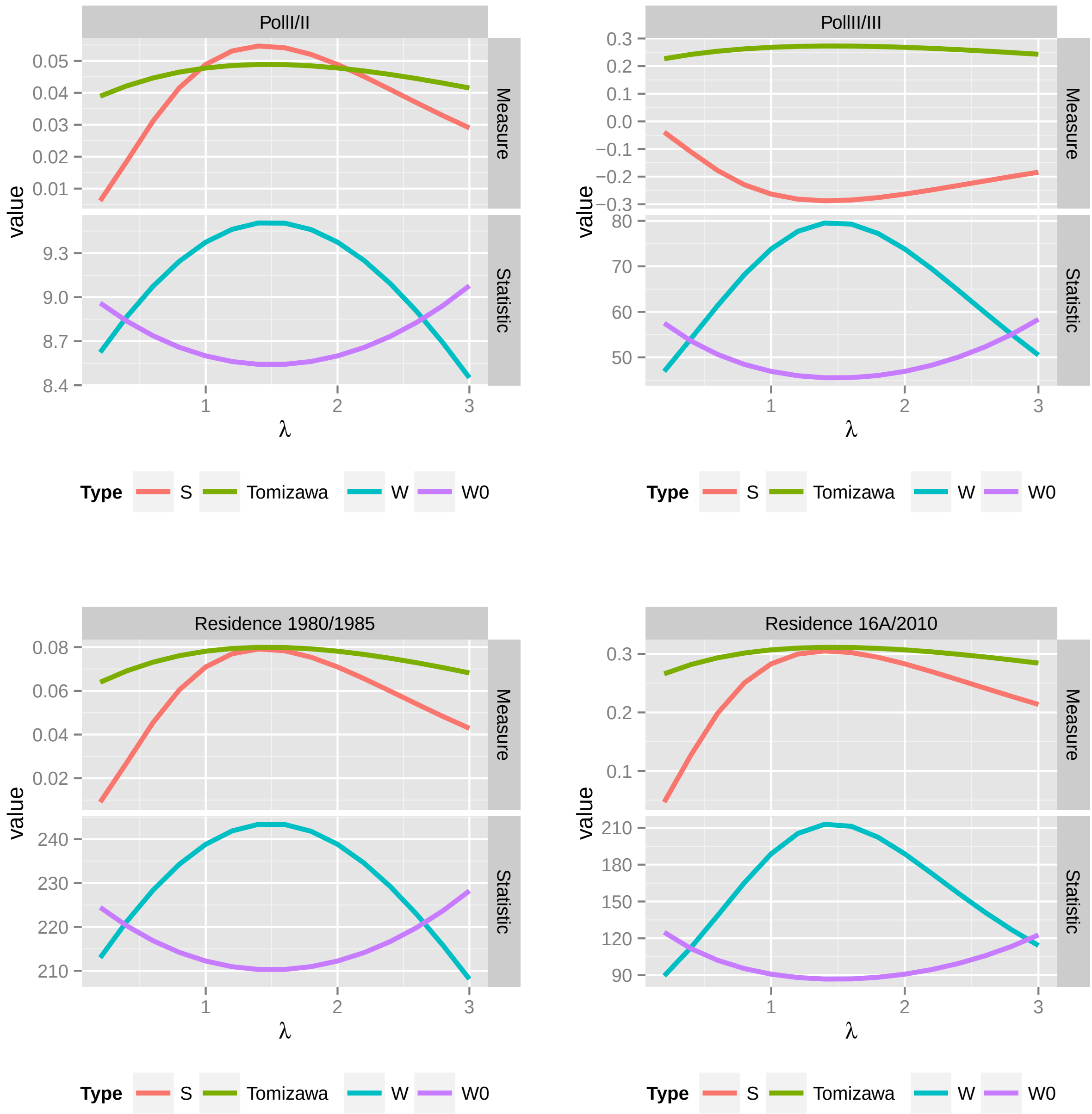


Figure 7.2: Values of the normalized scalar measure for asymmetry (7.17), called S, the measure of Tomizawa et al. [1998] (7.16), called Tomizawa, and the value for the $W^2_{S^\phi}$ (W) and $W^2_{S^\phi,0}$ (W0) test statistics, (7.12) and (7.13), respectively, in the power divergence family with parameter $\lambda \in [0.2, 3]$ for data set Table 7.5 (a) (Poll I/II), (b) (Poll II/III), Table 7.3 (Residence 1980/1985) and Table 7.4 (Residence 16A/2010).

leads to new ϕ-scaled test for symmetry. In the special case of the Pearson divergence, the ordinary symmetry tests of McNemar [1947] and Bowker [1948] are obtained. Thus, the ϕ-scaled tests are called ϕ-McNemar and ϕ-Bowker tests, respectively. The ϕ-McNemar test has been evaluated in terms of the type I error rate. It turns out, that the ordinary McNemar test has the best performance, so that it is suggested to be used in 2×2 tables for testing symmetry (resp. marginal homogeneity). In addition, the directed ϕ-scaled asymmetry measure for $I \times I$ tables can be used to construct a directed scalar measure for asymmetry. Such a measure can be used to compare and value asymmetries through different square tables.

Conclusion and Perspective

This thesis dealt with measuring association between two categorical variables cross-classified in two-way contingency tables. One of the basic problems of the classical association measures like the odds ratio is the occurrence of sampling zeros, which often is solved using the continuity correction. Corrections in contingency tables were presented as an unified theory by using posterior mean estimates in the Bayesian context. In addition, corresponding confidence intervals for the odds ratio can be obtained by inverting test statistics like likelihood ratio, Wald, score or Cressie-Read tests. The last one was, up to date, not used in the context of odds ratios. An extensive study for different correction techniques and confidence intervals for the log-odds ratio has been conducted and suggestions for the use of continuity corrections for small sample sizes are made in Chapter 2.

In the last years, the ϕ-divergence was used to generalise tools of categorical data analysis, introducing for example ϕ-based test statistics, minimum ϕ-divergence estimators (cf. Pardo [2006]), and ϕ-association models (Kateri and Papaioannou [1995]). Association can – next to other interpretations – be regarded as a measurement of divergence from independence in terms of size and direction, which is done by weighting the generalised association factors. These association factors also occur naturally when introducing ϕ-based models and weighting them by ϕ'. Using this concept, another approach of this work to deal with sampling zeros and therefore infinite estimates was to exploit and develop this information-theoretic fundament to introduce a new family of association measures, the ϕ-scaled odds ratio. It is a ϕ-based generalisation of the log-odds ratio and studied in Chapter 3 and the results have partly been published in Espendiller and Kateri [2016] for 2×2 tables. As a main result, the power divergence based odds ratio with parameter $\lambda = 1/3$ can improve the quality of the corresponding Wald confidence interval in terms of coverage probability in some set-ups. In addition, the corresponding independence test based on this odds ratio outperforms the independence test based on the log-odds ratio. Thus, the information-theoretic approach can improve the measurement of association in 2×2 tables.

Measures of association in contingency tables analysis are often linked to models. The most popular such connection is between loglinear models and local log-odds ratios. This connection to models was used in Chapter 4 to define new model-based measures of association that are more sensitive to capture complex association structures without losing their simple interpretability. Before studying such connections on ϕ-scale and the effect of the scale change itself, the classical log-scale is analysed and new measures for the uniform and row-effect association models are presented exemplary. These measures are provided by associated model parameters of the corresponding association model based on the generalised odds ratios, whose estimation is done in closed-form by extending the ideas of Clayton [1974]. By conducting some extensive simulation studies, it turns out that these closed-form estimators can in most situations replace the MLE and are sometimes even agreeable with sampling zeros.

While the information-theoretic approach was first applied to 2×2 tables, the techniques were extended to $I \times J$ tables in Chapter 5 to introduce the generalised ϕ-scaled odds ratios. Properties and asymptotic behaviour were studied. The generalised log-odds ratios are connected to different concepts of stochastic ordering. This connection does not hold in general for the generalised ϕ-scaled odds ratio, reflecting the different ways in which association is weighted. However, this has not to be a disadvantage enabling a greater flexibility of capturing the association. The newly defined objects are – under certain conditions – objects of association, i.e. they encode necessary information to uniquely define the joint distribution when the marginal distribution of the contingency tables are known. This was already known for the generalised log-odds ratio. Therefore, the fundament for multidimensional association measurement in $I \times J$ tables on ϕ-scale is set.

The ϕ-association model of Kateri and Papaioannou [1995] is directly linked with the local ϕ-scaled odds ratios, which measure the divergence of independence in terms of the local association factors. Using the previously introduced generalised ϕ-scaled odds ratio instead of the local ones leads to the introduction of generalised ϕ-association models. These models were studied in Chapter 6. Following the ideas of the closed-form estimators for the generalised log-odds ratios in Chapter 4, a closed-form estimator for parameters of generalised ϕ-scaled association models is introduced exemplary for the uniform and row-effect association. These estimators are used as a (multi-dimensional) measure of association. Again, the introduced closed-form estimators are a good alternative to maximum likelihood estimation for the different ϕ-scales, which require iterative fitting procedures. In addition, the ϕ-row effect measures of association were used to define a ϕ-based ordering test for scores of row categories. After a simulation study for this test, the use of the ordering test based on the Pearson divergence is suggested.

Another ϕ-based generalisation of known CDA tools is the introduction of ϕ-quasi-symmetry models for $I \times I$ contingency tables of two commensurable classification variables (Kateri and Papaioannou [1997]). Similar to the ϕ-association models, they also inherit some objects, that are used to measure asymmetries within the table. This leads to the study of the directed ϕ-scaled asymmetry measures. After proving properties and asymptotic behaviour, they were used to define generalisations of the symmetry tests of McNemar [1947] and Bowker [1948], the ϕ-McNemar and ϕ-Bowker tests. These ϕ-scaled test families include the classical ones in the case of the Pearson divergence. An extensive type I error rate study for 2×2 tables showed that the classical McNemar tests outperforms the other ϕ-generalised tests in the power divergence family and is therefore suggested for testing symmetry in 2×2 tables. In addition, a directed ϕ-based scalar measure to value the overall asymmetry in the tables was introduced, which was compared to the measure introduced by Tomizawa et al. [1998].

When dealing with ϕ-scaled generalisations, often the power divergence was used as parametric family of divergences, where the parameter λ was the control parameter. The previous work can also be conducted for other divergence families, for example the family introduced by Kateri et al. [2015] (Appendix A.2) or Kus et al. [2008]. The first one fulfils the property $\phi''(0) < \infty$ and gives finite variance-covariance estimates even in sampling zero set-ups. An evaluation study for the previously introduced tools can be done for these families to verify weather they offer alternative more powerful ϕ-based generalisations for association measurement in CDA.

In this work, the number of categorical classification variables was restricted to two, leading to two-way contingency tables. Multivariate categorical variables can be represented in multi-way tables. In such tables, the occurrence of sampling zeros is highly probable even for tables with moderate or large sample size and researchers are often confronted with these kind of contingency tables. As Agresti [2013, Section 10.6] summarizes, sampling zeros can lead to non-existing or infinite maximum likelihood estimates in loglinear models. In such cases, the iterative fitting procedures can report convergence but give very large and numerical unstable standard error estimates. Therefore, the multi-way tables with sampling zeros are confronted with the risk of invalid statistical inference. As seen, a ϕ-induced scale change can lead to tools, which are less sensible regarding sampling zeros. The asymptotic goodness-of-fit tests like the likelihood ratio or Pearson's test can be inadequate. Haberman [1977] justified the use of such tests if the expected cell frequency in every cell is at least in the range 5 to 10, which is often not fulfilled for multi-way tables. ϕ-divergence generalised association models face the same problem. It would be of interest to develop tests for ϕ-divergence based models and significance tests for ϕ-divergence based measures when asymptotic goodness-of-fit tests are not valid. The closed-form estimators from Chapter 6 for the ϕ-association models and the models themselves could be extended to multi-way tables and analysed in terms of their estimation ability and inferential properties for tables with many sampling zeros.

However, this very compendious summary already showed many fields of application of ϕ-divergence in CDA. As seen, a ϕ-induced scale change can lead to measures that are more compatible with sampling zeros and exhibiting better asymptotic properties, like improved coverage probability for its confidence intervals or reduced type I error rates of the corresponding tests and can therefore solve classical problems of association measurement. The new models extend the theory of association modelling, which can now be done on ϕ-scale and based on the different generalised odds ratios, which give more flexibility.

Overall, this thesis explores the application of the ϕ-divergence in contingency table analysis in terms of association measures and models.

Appendix A

Mathematical Background

A.1 Inverting Local log-Odds Ratios

Let $\mathcal{D} = \log \boldsymbol{\theta}^L(\Delta_{I,J})$ be the image space of the local log-odds ratio for an $I \times J$ contingency table. There is a bijective function $F : \Delta_{I,J} \to (\mathcal{D}, \Delta_I, \Delta_J), \boldsymbol{\pi} \mapsto (\log \boldsymbol{\theta}^L(\boldsymbol{\pi}), \boldsymbol{\pi}^X, \boldsymbol{\pi}^Y)$, where $\boldsymbol{\pi}^X$ and $\boldsymbol{\pi}^Y$ are the marginal distribution of the multinomial probability $\boldsymbol{\pi}$ (see Osius [2004] or Chapter 5). The image space of $\log \boldsymbol{\theta}^L$ is the image of the IJ-simplex under a multiplicative logarithmic function and therefore hard to handle. In the 2×2 case, $\mathcal{D} = \mathbb{R}$. For any fixed $\log \boldsymbol{\theta}_0^L \in \mathcal{D}, \boldsymbol{\pi}^X \in \Delta_I$ and $\boldsymbol{\pi}^Y \in \Delta_J$, one has to find $\boldsymbol{\pi} \in \Delta_{I,J}$ such that

$$F(\boldsymbol{\pi}) = (\log \theta_0^L, \boldsymbol{\pi}^X, \boldsymbol{\pi}^Y). \tag{A.1}$$

Let $\mathbf{C} = \mathbf{C}^L$ be the $(I-1)(J-1) \times IJ$ matrix described in (1.43) such that $\mathbf{C} \log \boldsymbol{\pi} = \log \boldsymbol{\theta}^L$. In addition, let $\mathbf{B}^X$ and $\mathbf{B}^Y$ be the $I \times IJ$ resp. $J \times IJ$ matrices consisting of 1's and 0's producing the marginal distributions, $\mathbf{B}^X \boldsymbol{\pi} = \boldsymbol{\pi}^X$ and $\mathbf{B}^Y \boldsymbol{\pi} = \boldsymbol{\pi}^Y$. The element $\boldsymbol{\pi} \in \Delta_{I,J}$ in (A.1) is the unique solution to the nonlinear system of equations:

$$\mathbf{C} \log \boldsymbol{\pi} = \log \boldsymbol{\theta}_0^L \tag{A.2}$$

$$\mathbf{B}^X \boldsymbol{\pi} = \boldsymbol{\pi}^X \tag{A.3}$$

$$\mathbf{B}^Y \boldsymbol{\pi} = \boldsymbol{\pi}^Y. \tag{A.4}$$

Notice that $\sum_{i=1}^I \pi_i^X = \sum_{j=1}^J \pi_j^Y = 1$ and thus one equation in (A.3) or (A.4) is redundant. The solution $\boldsymbol{\pi}$ of the nonlinear system of equations can then be found numerically using the R package **nleqslv**, which is based on the Broyden method described in Dennis and Schnabel [1996]. The inverting algorithm is for example applied in Section 6.5.2 for the simulation study of the ϕ-R-WM-ordering test or in Section 6.2 for conducting the simulation study to analyse the effect of the scale change in ϕ-association models. The corresponding R function is `get.model.thetaR` in Appendix B.4.

A.2 Parametric Families of Divergences

The most well-known parametric family of divergences is the power divergence (1.73). Another parametric family of divergences, introduced by Kateri et al. [2015] in the framework of quasi-symmetry models, is given by

$$\phi_t(x) = f_t(x) - f_t(1) - f_t'(1)(x-1), \qquad 0 < t < 1, \tag{A.5}$$

where

$$\begin{aligned} f_t(x) &= \left(x + \frac{2t}{1-t}\right) \log\left(x + \frac{2t}{1-t}\right) \\ f_t'(x) &= \log\left(x + \frac{2t}{1-t}\right) + 1 \\ f_t''(x) &= \frac{1}{x + \frac{2t}{1-t}} = \frac{1-t}{(1-t)x + 2t}. \end{aligned}$$

For $t \to 0$, the divergence becomes the Kullback-Leibler divergence (1.70), i.e. $\lim_{t\searrow 0} \phi_t = \phi_{KL}$. It holds

$$\begin{aligned} \phi_t'(x) &= f_t'(x) - f_t(1) = \log\left(x + \frac{2t}{1-t}\right) - \log\left(1 + \frac{2t}{1-t}\right) \\ \phi_t''(x) &= f_t''(x) = \frac{1-t}{(1-t)x + 2t} > 0 \end{aligned}$$

and

$$\begin{aligned} \phi_t'(0) &= f_t'(0) - f_t(1) = \log\left(\frac{2t}{1-t}\right) - \log\left(1 + \frac{2t}{1-t}\right) > -\infty \\ \phi_t''(0) &= f_t''(0) = \frac{1-t}{2t} > -\infty. \end{aligned}$$

Thus ϕ_t is strictly convex due to $\phi_t'(0) > -\infty$.

The ϕ_t-scaled odds ratio θ^{ϕ_t} (cf. (3.1)) is bounded and well-defined even for sampling zeros. In addition, due to $\phi_t''(0) > 0$ the asymptotic variance is also defined for sampling zeros ($0 < t < 1$) (see Lemma 3.3.1). Therefore, the ϕ_t-scaled odds ratios are close to the log-odds ratios and do not face the problem of sampling zeros. Coverage probabilities and average relative length (ARL) (cf. Section 3.6) for selected probability pairs (π_1, π_2) in the binomial sampling scheme with different sample sizes $(n_1, n_2) = (10, 10), (15, 15), (10, 30)$ for the ϕ_t-scaled odds ratios are given with (Table A.1) and without (Table A.2) continuity correction $c = 0.5$, which can be compared to the results of Table 3.6 for the ϕ_λ-scaled odds ratio of the power divergence ϕ_λ (1.70).

A.3 HLP Models

Lang [2004, 2005] introduced a powerful family of models, the *homogeneous linear predictor* (HLP) models, that include the classical GLMs and GLLMs (see Section 1.9.5).

		Coverage Probability														
		$(n_1,n_2)=(10,10)$					$(n_1,n_2)=(15,15)$					$(n_1,n_2)=(10,30)$				
π_1	π_2	0	1/5	2/5	3/5	4/5	0	1/5	2/5	3/5	4/5	0	1/5	2/5	3/5	4/5
0.700	0.700	0.980	0.923	0.923	0.923	0.923	0.973	0.944	0.944	0.931	0.919	0.972	0.941	0.937	0.928	0.925
	0.500	0.978	0.954	0.937	0.937	0.934	0.966	0.958	0.943	0.943	0.943	0.973	0.940	0.938	0.936	0.936
	0.300	0.971	0.932	0.932	0.932	0.932	0.959	0.943	0.949	0.949	0.949	0.971	0.938	0.946	0.941	0.941
	0.100	0.962	0.967	0.974	0.974	0.974	0.960	0.945	0.954	0.954	0.955	0.956	0.950	0.950	0.938	0.943
	0.050	0.952	0.964	0.967	0.975	0.975	0.942	0.964	0.968	0.973	0.974	0.956	0.948	0.948	0.948	0.944
	0.025	0.909	0.981	0.994	0.972	0.972	0.927	0.984	0.988	0.992	0.992	0.961	0.965	0.973	0.976	0.983
0.500	0.500	0.960	0.948	0.948	0.948	0.948	0.957	0.954	0.954	0.954	0.943	0.968	0.944	0.944	0.942	0.930
	0.300	0.978	0.954	0.937	0.937	0.934	0.966	0.958	0.943	0.943	0.943	0.967	0.943	0.942	0.942	0.935
	0.100	0.959	0.954	0.963	0.963	0.963	0.969	0.946	0.946	0.951	0.954	0.974	0.951	0.940	0.940	0.934
0.400	0.400	0.965	0.940	0.940	0.940	0.940	0.961	0.951	0.951	0.950	0.934	0.971	0.943	0.943	0.937	0.930
	0.200	0.979	0.938	0.917	0.917	0.917	0.969	0.930	0.930	0.924	0.924	0.973	0.951	0.942	0.932	0.941
	0.100	0.973	0.971	0.954	0.954	0.954	0.969	0.959	0.960	0.935	0.935	0.975	0.952	0.941	0.932	0.928
	0.050	0.947	0.977	0.980	0.980	0.974	0.945	0.976	0.982	0.982	0.982	0.978	0.942	0.942	0.942	0.942
	0.025	0.903	0.981	0.988	0.993	0.993	0.934	0.985	0.992	0.992	0.992	0.959	0.979	0.980	0.976	0.976
0.300	0.100	0.980	0.964	0.964	0.963	0.963	0.976	0.947	0.946	0.953	0.953	0.988	0.955	0.922	0.922	0.905
	0.050	0.964	0.971	0.974	0.974	0.974	0.962	0.979	0.974	0.974	0.974	0.983	0.951	0.940	0.943	0.943
0.200	0.200	0.995	0.913	0.913	0.913	0.913	0.989	0.933	0.933	0.906	0.904	0.971	0.944	0.927	0.926	0.906
	0.100	0.995	0.956	0.956	0.956	0.956	0.987	0.939	0.938	0.938	0.938	0.992	0.939	0.907	0.907	0.902
	0.050	0.984	0.983	0.987	0.983	0.983	0.978	0.977	0.971	0.971	0.971	0.997	0.947	0.947	0.947	0.946
	0.025	0.961	0.968	0.968	0.969	0.969	0.955	0.968	0.979	0.979	0.979	0.985	0.977	0.964	0.956	0.966
0.100	0.100	1.000	0.950	0.950	0.950	0.950	0.999	0.922	0.922	0.915	0.915	0.981	0.970	0.963	0.963	0.947
	0.050	0.999	0.969	0.969	0.962	0.962	0.997	0.957	0.957	0.957	0.957	0.999	0.975	0.917	0.917	0.917
	0.025	0.990	0.991	0.991	0.991	0.991	0.986	0.987	0.987	0.987	0.987	1.000	0.981	0.980	0.980	0.980
0.050	0.050	1.000	0.986	0.986	0.986	0.986	1.000	0.966	0.966	0.966	0.966	0.993	0.977	0.977	0.977	0.975
	0.025	1.000	0.985	0.985	0.984	0.984	1.000	0.975	0.975	0.975	0.975	1.000	0.994	0.959	0.956	0.956

		Average Relative Length														
		$(n_1,n_2)=(10,10)$					$(n_1,n_2)=(15,15)$					$(n_1,n_2)=(10,30)$				
π_1	π_2	0	1/5	2/5	3/5	4/5	0	1/5	2/5	3/5	4/5	0	1/5	2/5	3/5	4/5
0.700	0.700	0.322	0.435	0.467	0.480	0.485	0.231	0.346	0.377	0.389	0.394	0.222	0.314	0.329	0.326	0.316
	0.500	0.306	0.407	0.435	0.446	0.450	0.220	0.323	0.349	0.360	0.364	0.217	0.288	0.296	0.290	0.279
	0.300	0.318	0.388	0.403	0.409	0.411	0.231	0.311	0.327	0.333	0.335	0.222	0.281	0.288	0.286	0.280
	0.100	0.375	0.350	0.341	0.337	0.335	0.295	0.286	0.280	0.278	0.277	0.258	0.278	0.278	0.279	0.282
	0.050	0.403	0.328	0.311	0.304	0.302	0.338	0.267	0.253	0.248	0.246	0.293	0.261	0.250	0.247	0.250
	0.025	0.420	0.314	0.293	0.285	0.282	0.367	0.253	0.235	0.229	0.226	0.324	0.240	0.221	0.215	0.215
0.500	0.500	0.289	0.403	0.437	0.451	0.456	0.209	0.319	0.350	0.363	0.368	0.201	0.290	0.305	0.303	0.293
	0.300	0.306	0.407	0.435	0.446	0.450	0.220	0.323	0.349	0.360	0.364	0.206	0.297	0.315	0.316	0.310
	0.100	0.385	0.403	0.405	0.406	0.406	0.294	0.319	0.322	0.323	0.323	0.247	0.315	0.330	0.338	0.343
0.400	0.400	0.297	0.411	0.444	0.457	0.463	0.214	0.325	0.356	0.369	0.374	0.206	0.295	0.310	0.308	0.298
	0.200	0.337	0.428	0.453	0.462	0.466	0.244	0.341	0.364	0.374	0.377	0.219	0.320	0.343	0.348	0.345
	0.100	0.393	0.434	0.442	0.446	0.447	0.296	0.342	0.351	0.354	0.355	0.251	0.338	0.359	0.370	0.375
	0.050	0.439	0.433	0.428	0.426	0.425	0.349	0.335	0.330	0.328	0.327	0.299	0.340	0.351	0.359	0.366
	0.025	0.468	0.430	0.417	0.412	0.410	0.387	0.327	0.312	0.307	0.305	0.346	0.333	0.332	0.335	0.340
0.300	0.100	0.406	0.475	0.491	0.496	0.499	0.303	0.376	0.391	0.397	0.399	0.262	0.366	0.393	0.404	0.407
	0.050	0.452	0.484	0.487	0.488	0.488	0.355	0.378	0.378	0.379	0.379	0.310	0.380	0.398	0.409	0.416
0.200	0.200	0.376	0.488	0.518	0.530	0.534	0.271	0.390	0.421	0.433	0.437	0.258	0.355	0.369	0.367	0.357
	0.100	0.430	0.532	0.555	0.563	0.566	0.321	0.425	0.449	0.458	0.461	0.287	0.405	0.431	0.439	0.438
	0.050	0.475	0.558	0.571	0.575	0.576	0.370	0.443	0.455	0.458	0.460	0.332	0.436	0.461	0.472	0.477
	0.025	0.503	0.571	0.578	0.578	0.578	0.407	0.452	0.453	0.452	0.452	0.377	0.450	0.465	0.473	0.481
0.100	0.100	0.479	0.613	0.640	0.648	0.651	0.365	0.499	0.529	0.539	0.543	0.337	0.456	0.473	0.470	0.459
	0.050	0.520	0.665	0.691	0.697	0.699	0.410	0.550	0.579	0.586	0.589	0.377	0.519	0.544	0.550	0.548
	0.025	0.546	0.695	0.719	0.723	0.723	0.443	0.580	0.605	0.609	0.611	0.417	0.555	0.578	0.587	0.591
0.050	0.050	0.557	0.738	0.769	0.774	0.775	0.451	0.629	0.668	0.675	0.678	0.414	0.574	0.596	0.596	0.587
	0.025	0.582	0.780	0.813	0.817	0.817	0.481	0.678	0.723	0.727	0.729	0.452	0.628	0.655	0.663	0.662

Table A.1: Coverage probabilities and average relative length for the 95% CIs of θ^{ϕ_t} based on the ϕ_t-divergence (A.5), $t \in \{0, 1/5, 2/5, 3/5, 4/5\}$, under the independent binomial sampling scheme, for three choices of tables sample sizes, using continuity correction $c = 0.5$.

These models have, next to their great generality, the ability to model contingency tables using the multinomial-Poisson distribution, i.e. a combination of multinomial and Poisson

		Coverage Probability											
		$(n_1, n_2) = (10, 10)$				$(n_1, n_2) = (15, 15)$				$(n_1, n_2) = (10, 30)$			
π_1	π_2	1/5	2/5	3/5	4/5	1/5	2/5	3/5	4/5	1/5	2/5	3/5	4/5
0.700	0.700	0.903	0.885	0.885	0.885	0.918	0.918	0.918	0.913	0.923	0.909	0.908	0.908
	0.500	0.916	0.913	0.913	0.913	0.932	0.926	0.926	0.926	0.924	0.910	0.910	0.910
	0.300	0.913	0.913	0.913	0.856	0.935	0.935	0.904	0.904	0.909	0.909	0.909	0.909
	0.100	0.740	0.841	0.843	0.845	0.795	0.826	0.868	0.872	0.899	0.909	0.881	0.881
	0.050	0.893	0.895	0.890	0.893	0.852	0.853	0.855	0.933	0.789	0.801	0.779	0.768
	0.025	0.866	0.869	0.869	0.869	0.905	0.897	0.900	0.900	0.801	0.806	0.817	0.692
0.500	0.500	0.948	0.911	0.911	0.911	0.943	0.943	0.943	0.918	0.930	0.929	0.917	0.917
	0.300	0.916	0.913	0.913	0.913	0.932	0.926	0.926	0.926	0.935	0.927	0.922	0.922
	0.100	0.630	0.630	0.630	0.636	0.774	0.774	0.774	0.774	0.899	0.899	0.896	0.893
0.400	0.400	0.938	0.906	0.906	0.906	0.934	0.934	0.934	0.915	0.930	0.924	0.917	0.917
	0.200	0.836	0.836	0.833	0.833	0.895	0.895	0.895	0.895	0.924	0.917	0.900	0.907
	0.100	0.626	0.626	0.626	0.621	0.771	0.771	0.776	0.776	0.895	0.888	0.888	0.888
	0.050	0.392	0.489	0.618	0.618	0.523	0.526	0.572	0.572	0.759	0.761	0.759	0.758
	0.025	0.708	0.864	0.864	0.952	0.585	0.850	0.847	0.848	0.522	0.522	0.522	0.517
0.300	0.100	0.626	0.626	0.626	0.626	0.774	0.770	0.770	0.770	0.886	0.873	0.873	0.869
	0.050	0.383	0.383	0.383	0.383	0.519	0.524	0.524	0.524	0.753	0.753	0.753	0.736
0.200	0.200	0.788	0.784	0.784	0.784	0.878	0.878	0.878	0.878	0.860	0.856	0.856	0.856
	0.100	0.600	0.584	0.584	0.581	0.759	0.737	0.737	0.737	0.830	0.825	0.825	0.825
	0.050	0.379	0.379	0.379	0.379	0.511	0.519	0.519	0.519	0.701	0.701	0.701	0.688
	0.025	0.210	0.210	0.210	0.210	0.302	0.302	0.304	0.304	0.498	0.498	0.489	0.489
0.100	0.100	0.481	0.481	0.481	0.481	0.649	0.649	0.649	0.649	0.616	0.616	0.616	0.616
	0.050	0.330	0.330	0.325	0.325	0.469	0.469	0.457	0.457	0.548	0.547	0.547	0.547
	0.025	0.199	0.199	0.199	0.199	0.290	0.290	0.290	0.290	0.413	0.413	0.413	0.413
0.050	0.050	0.251	0.251	0.251	0.251	0.366	0.366	0.366	0.366	0.357	0.357	0.357	0.357
	0.025	0.168	0.168	0.168	0.168	0.248	0.248	0.248	0.248	0.296	0.296	0.296	0.296

		Average Relative Length											
		$(n_1, n_2) = (10, 10)$				$(n_1, n_2) = (15, 15)$				$(n_1, n_2) = (10, 30)$			
π_1	π_2	1/5	2/5	3/5	4/5	1/5	2/5	3/5	4/5	1/5	2/5	3/5	4/5
0.700	0.700	0.387	0.427	0.442	0.448	0.327	0.363	0.377	0.382	0.293	0.311	0.308	0.297
	0.500	0.367	0.402	0.415	0.420	0.304	0.334	0.346	0.350	0.267	0.277	0.271	0.260
	0.300	0.347	0.368	0.375	0.378	0.290	0.309	0.315	0.318	0.260	0.270	0.269	0.263
	0.100	0.273	0.276	0.276	0.275	0.245	0.244	0.243	0.242	0.256	0.258	0.260	0.264
	0.050	0.223	0.224	0.223	0.223	0.203	0.200	0.198	0.198	0.225	0.218	0.217	0.219
	0.025	0.189	0.190	0.190	0.189	0.171	0.169	0.168	0.167	0.185	0.173	0.168	0.166
0.500	0.500	0.370	0.411	0.426	0.433	0.301	0.336	0.350	0.355	0.274	0.291	0.289	0.278
	0.300	0.367	0.402	0.415	0.420	0.304	0.334	0.346	0.350	0.282	0.302	0.304	0.297
	0.100	0.291	0.307	0.312	0.314	0.264	0.274	0.277	0.278	0.293	0.312	0.321	0.327
0.400	0.400	0.375	0.416	0.432	0.438	0.307	0.343	0.356	0.362	0.279	0.296	0.294	0.283
	0.200	0.365	0.398	0.410	0.415	0.316	0.343	0.354	0.358	0.305	0.332	0.338	0.334
	0.100	0.301	0.323	0.331	0.334	0.280	0.295	0.301	0.302	0.315	0.341	0.352	0.358
	0.050	0.226	0.242	0.248	0.250	0.217	0.226	0.229	0.230	0.279	0.296	0.306	0.314
	0.025	0.172	0.184	0.189	0.191	0.162	0.169	0.172	0.173	0.218	0.226	0.231	0.236
0.300	0.100	0.312	0.340	0.351	0.355	0.303	0.325	0.333	0.336	0.335	0.367	0.380	0.384
	0.050	0.229	0.249	0.256	0.259	0.230	0.245	0.250	0.252	0.301	0.326	0.339	0.348
0.200	0.200	0.377	0.416	0.431	0.436	0.355	0.391	0.405	0.410	0.309	0.330	0.329	0.319
	0.100	0.315	0.348	0.361	0.366	0.328	0.359	0.370	0.375	0.340	0.373	0.384	0.385
	0.050	0.230	0.253	0.263	0.266	0.249	0.270	0.278	0.281	0.312	0.344	0.359	0.365
	0.025	0.161	0.178	0.185	0.187	0.171	0.186	0.192	0.194	0.241	0.264	0.276	0.282
0.100	0.100	0.279	0.310	0.322	0.327	0.319	0.352	0.365	0.370	0.279	0.305	0.311	0.308
	0.050	0.213	0.238	0.247	0.251	0.251	0.278	0.289	0.293	0.270	0.300	0.312	0.315
	0.025	0.152	0.169	0.176	0.179	0.174	0.193	0.200	0.203	0.221	0.246	0.258	0.263
0.050	0.050	0.177	0.198	0.206	0.210	0.213	0.238	0.248	0.251	0.191	0.212	0.220	0.222
	0.025	0.135	0.152	0.158	0.161	0.159	0.177	0.185	0.188	0.170	0.190	0.199	0.202

Table A.2: Coverage probabilities and average relative length for the 95% CIs of θ^{ϕ^t} based on the ϕ_t-divergence (A.5), $t \in \{1/5, 2/5, 3/5, 4/5\}$, under the independent binomial sampling scheme, for three choices of tables sample sizes **without** using continuity correction, conditional on $n_{+j} \neq 0, j = 1, 2$.

sampling scheme spread over any selection of cells.

The following matrices and vectors are defined:

$$\mathbf{Z} = \mathbf{1}_{IJ} \in \mathbb{R}^{IJ\times 1} \quad \mathbf{Z}_F = \mathbf{0}_{IJ} \in \mathbb{R}^{IJ\times 1} \quad \mathfrak{n} = 0 \quad \text{(Poisson)} \tag{A.6}$$

$$\mathbf{Z} = \mathbf{1}_{IJ} \in \mathbb{R}^{IJ\times 1} \quad \mathbf{Z}_F = \mathbf{Z} \in \mathbb{R}^{IJ\times 1} \quad \mathfrak{n} = n \in \mathbb{N} \quad \text{(Multinomial)} \tag{A.7}$$

$$\mathbf{Z} = \oplus_{j=1}^{J} \mathbf{1}_I \in \mathbb{R}^{I\times J} \quad \mathbf{Z}_F = \mathbf{Z} \in \mathbb{R}^{I\times J} \quad \mathfrak{n} = (n_1, ..., n_I) \in \mathbb{N}^I \quad \text{(Product Multinomial)}, \tag{A.8}$$

where $\mathbf{Z}$ is called *population matrix*, $\mathbf{Z}_F$ is the *sampling constraint matrix* and $\mathfrak{n}$ is the *vector of sample sizes*. Overall, $(\mathbf{Z}, \mathbf{Z}_F, \mathfrak{n})$ is called *sampling plan*. A sampling plan can be used to construct a distribution for an $I \times J$ contingency table $\mathbf{N}$ by conditioning independent Poisson distributions (Haberman [1974a]).

Let $(\mathbf{Z}, \mathbf{Z}_F, \mathfrak{n})$ be a sampling plan and let $\mathbf{Y} = (Y_{ij})$ follow the independent Poisson sampling scheme in an $I \times J$ table, i.e. $\mathbf{Y} \sim \mathcal{P}(\boldsymbol{\gamma})$ as in (1.1). Then

$$\mathbf{N} \sim \mathbf{Y} \mid \{\mathbf{Z}_F^T \mathbf{Y} = \mathfrak{n}\} \tag{A.9}$$

obtains a multinomial-Poisson distribution. Since focus lied on the multinomial and the product multinomial sampling scheme, sampling plans are not introduced formally and the introduction is restricted to the sampling plans needed in this work. Clearly, (A.9) becomes the Poisson sampling scheme for (A.6), the multinomial sampling scheme for (A.7) and the product multinomial sampling scheme for (A.8).

Let $\boldsymbol{\mu} = \mathbb{E}\mathbf{N}$ be the expected value of (A.9). To define a HLP model, additional homogeneous constraints in terms of a function $\mathbf{h}(\boldsymbol{\mu}) = \mathbf{0}$ are introduced.

Definition A.3.1. [Lang, 2004, p. 351]

Let $\Omega = \{\boldsymbol{\mu} \in \mathbb{R}^c \mid \boldsymbol{\mu} > 0\}$. A function $\mathbf{h} : \Omega \to \mathbb{R}^u$ is $\mathbf{Z}$-homogeneous if

$$\mathbf{h}(\operatorname{diag}(\mathbf{Z}\boldsymbol{\delta})\boldsymbol{\mu}) = \mathbf{h}(\boldsymbol{\mu}), \qquad \forall \boldsymbol{\delta} > 0, \forall \boldsymbol{\mu} \in \Omega.$$

Here $\mathbf{Z}$ is a $c \times K$ population matrix and $\boldsymbol{\delta} = (\delta_1, ..., \delta_K)$. Define $\mathcal{H}(\mathbf{Z})$ as the set of all $\mathbf{Z}$-homogeneous functions.

Definition A.3.2. [Lang, 2004, p. 352]

The set $\mathcal{H}''(\mathbf{Z})$ contains all functions $\mathbf{h} : \Omega \to \mathbb{R}^u$ that satisfy the following four conditions:

(i) $\omega(\mathbf{h} \mid \mathbf{0}) := \{\boldsymbol{\mu} \mid \boldsymbol{\mu} > 0, \mathbf{h}(\boldsymbol{\mu}) = \mathbf{0}\} \neq \emptyset$

(ii) $\mathbf{h}$ has continuous second-order derivatives on $\Omega = \{\boldsymbol{\mu} \mid \boldsymbol{\mu} > 0\}$

(iii) $\mathbf{H}(\boldsymbol{\mu}) = \partial \mathbf{h}(\boldsymbol{\mu}) / \partial \boldsymbol{\mu}$ is of full column rank u on $\Omega = \{\boldsymbol{\mu} \mid \boldsymbol{\mu} > 0\}$

(iv) $\mathbf{h} \in \mathcal{H}(\mathbf{Z})$.

In addition, $\mathbf{H} = \mathbf{0}$ is accepted as a full column rank matrix.

HLP models can now be defined:

Definition A.3.3. [Lang, 2005, p. 125, Definition 3]

Let $(\mathbf{Z}, \mathbf{Z}_F, \mathfrak{n})$ be a sampling plan and let $\mathbf{N}$ follow (A.9) with expected value $\boldsymbol{\mu} = \mathbb{E}\mathbf{N}$. An homogeneous linear predictor (HLP) model has the form

$$\mathbf{L}(\boldsymbol{\mu}) = \mathbf{X}\boldsymbol{\beta}$$

for a model design matrix $\mathbf{X}$, a parameter vector $\boldsymbol{\beta}$ and a function $\mathbf{L}$, that satisfies the following two conditions:

HLP1 $\mathbf{L}(\boldsymbol{\mu}) = \mathbf{a}(\boldsymbol{\gamma}) + \mathbf{L}(\boldsymbol{\pi})$ for some function $\mathbf{a}$, with $\mathbf{a}(\boldsymbol{\gamma}_1) - \mathbf{a}(\boldsymbol{\gamma}_2) = \mathbf{a}(\boldsymbol{\gamma}_1/\boldsymbol{\gamma}_2) - \mathbf{a}(\mathbf{1})$

HLP2 $\mathbf{U}^T\mathbf{L} \in \mathcal{H}''(\mathbf{Z})$ for a full column rank $\mathbf{U}$ that is an orthogonal complement of $\mathbf{X}$,

where $\boldsymbol{\gamma} = \mathbb{E}\mathbf{N}$ for the Poisson, $\boldsymbol{\gamma} = n$ for multinomial and $\boldsymbol{\gamma} = (n_1, ..., n_I)$ for the product multinomial sampling scheme.

Notice that this definition can be extended to general multinomial-Poisson distributions. In addition, one can introduce a model with

$$\mathbf{h}(\boldsymbol{\mu}) = \mathbf{0},$$

which is called MPH model.

A.4 Union Intersection Test

A simple technique for constructing multivariate one-sided test is the union intersection method (cf. Casella and Berger [2002]). Let $\mathbf{X} = (X_1, ..., X_k) \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ be a k-dimensional multivariate normal distributed random variable with mean $\boldsymbol{\mu} = (\mu_1, ..., \mu_k)$ and $k \times k$ variance-covariance matrix $\boldsymbol{\Sigma} = (\sigma_{ij})$. Let $\mathbf{x} = (x_1, ..., x_k)$ be a realisation of $\mathbf{X}$. The union intersection test can be used to test

$$H_0 : \boldsymbol{\mu} = \mathbf{0} \qquad \text{vs.} \qquad H_1 : \boldsymbol{\mu} \geq \mathbf{0} \quad \text{with at least one inequality strict} \tag{A.10}$$

by testing the hypothesis on each subspace. For fixed $i = 1, ..., k$, the i-th component of $\mathbf{X}$ is distributed as $X_i \sim \mathcal{N}(\mu_i, \sigma_{ii})$. Regard the hypotheses $H_{0i} : \mu_i = 0$ vs $H_{1i} : \mu_i > 0$, which are rejected if

$$T_i(\mathbf{X}) = \frac{x_i}{\sqrt{\sigma_{ii}}} > u_\alpha,$$

where u_α is the $(1-\alpha)$-quantile of a standard normal distribution. The hypothesis (A.10) can be written as $H_0 : \bigcap_{i=1}^k H_{0i}$ and the union intersection tests rejects H_0 at level $\alpha \in (0, 1)$ if

$$\max_{i=1,\dots,k} T_i(\mathbf{X}) \geq a_{k,\alpha} \tag{A.11}$$

for some $a_{k,\alpha}$. Let $\mathbf{R} = \{\max_{i=1,\dots,k} T_i(\mathbf{X}) \geq a_{k,\alpha}\}$ be the rejection region. It holds

$$\begin{aligned}\mathbf{P}(\mathbf{R}) &= \mathbf{P}(\max_{i=1,\dots,k} T_i(\mathbf{X}) \geq a_{k,\alpha}) = 1 - \mathbf{P}(\max_{i=1,\dots,k} T_i(\mathbf{X}) \leq a_{k,\alpha}) \\ &= 1 - \prod_{i=1}^{k} \Phi(a_{k,\alpha}) = 1 - (\Phi(a_{k,\alpha}))^k.\end{aligned}$$

Thus, the critical value $a_{k,\alpha}$ in the union intersection test in (A.11) is the solution to $1 - \Phi(a_{k,\alpha})^k = \alpha$ given by $a_{k,\alpha} = \Phi^{-1}((1-\alpha)^{\frac{1}{k}})$, where Φ is the distribution function of the standard normal distribution. The union intersection test is used in Section 6.5 to define a ϕ-scaled test for the ordering of the row scores in a R-model.

A.5 Uniform Multinomial Probability Sampling

To value the quality of new proposed models, estimates and association measures, often random studies are conducted on the particular probability parameter (sub)space. Such a studies require the generation of multinomial samples and - dependent on the simulation study - the generation of the multinomial probabilities themselves. While the first is simple and implemented in most statistical software, the probability sampling for multinomials requires uniform sampling on the $n := IJ$-simplex $\Delta_{I,J}$. A simple algorithm to produce uniformly distributed $\boldsymbol{\pi} \in \Delta_n$, which is available as R function `gen.uniform.probability` in Appendix B.8, was proposed by Smith and Tromble [2004]: Let $M \in \mathbb{N}$ be large.

1. Select $x_1, ..., x_{n-1}$ uniformly on $\{0, 1, ..., M\}$.
2. Let $x_{(1)}, ..., x_{(n-1)}$ be the ordered sample and add $x_{(0)} = 0$ and $x_{(n)} = M$ such that $0 = x_{(0)} \leq x_{(1)} \leq ... \leq x_{(n)} = M$.
3. Take the differences $y_i = x_{(i)} - x_{(i-1)}, i = 1, ..., n$ and set $\mathbf{y} = (y_1, ..., y_n)$.
4. The vector $\boldsymbol{\pi} = \frac{\mathbf{y}}{M}$ is element of the n-simplex Δ_n.

Example A.5.1. (Sampling in R-models) Let $\mathcal{R} \subseteq \Delta_{I,J}$ be the subspace of probability vectors fulfilling the R-model, $\log \theta^L_{ij} = c_{i,0}, i = 1, ..., I-1$. Assume that the log-odds ratios are restricted on a subspace $\mathbf{c} = (c_{1,0}, ..., c_{I-1,0}) \in \mathcal{C} \subseteq \mathbb{R}^{I-1}$ due to the aim of the study. The marginal distributions $\boldsymbol{\pi}^X \in \Delta_I$ and $\boldsymbol{\pi}^Y \in \Delta_J$ together with $\mathbf{c} \in \mathcal{C}$ give a joint distribution $\boldsymbol{\pi} = \boldsymbol{\pi}(\mathbf{c}, \boldsymbol{\pi}^X, \boldsymbol{\pi}^Y) \in \mathcal{R}$ by inverting a nonlinear system of equations (Appendix A.1). Thus for random sampling from $\boldsymbol{\pi} \in \mathcal{R}$, it is sufficient to generate $\mathbf{c} \in \mathcal{C}$ uniformly and use the proposed algorithm to generate $\boldsymbol{\pi}^X \in \Delta_I$ and $\boldsymbol{\pi}^Y \in \Delta_J$.

A.6 Selected ϕ-Divergences

Kullback-Leibler	$\phi_{KL}(x) = x \log x - x + 1$	$\phi'_{KL}(x) = \log x$ $\phi'_{KL}(1) = 0$	$\phi''_{KL}(x) = \frac{1}{x}$ $\phi''_{KL}(1) = 1$
Pearson	$\phi_P(x) = \frac{1}{2}(x-1)^2$	$\phi'_P(x) = x - 1$ $\phi'_P(1) = 0$	$\phi''_P(x) = 1$ $\phi''_P(1) = 1$
Power Divergence ($\lambda \neq -1, 0$)	$\phi_\lambda(x) = \frac{x^{\lambda+1} - x - \lambda(x-1)}{\lambda(\lambda+1)}$	$\phi'_\lambda(x) = \frac{x^\lambda - 1}{\lambda}$ $\phi'_\lambda(1) = 0$	$\phi''_\lambda(x) = x^{\lambda-1}$ $\phi''_\lambda(1) = 1$
J-Divergence	$\phi(x) = (x-1)\log x$	$\phi'(x) = \log(x) - \frac{1}{x} + 1$ $\phi'(1) = 0$	$\phi''(x) = \frac{1}{x^2} + \frac{1}{x}$ $\phi''(x) = 2$
Balakrishnan and Sanghri	$\phi(x) = \frac{(x-1)^2}{(x+1)^2}$	$\phi'(x) = \frac{4(x-1)}{(x+1)^3}$ $\phi'(1) = 0$	$\phi''(x) = -\frac{8(x-2)}{(x+1)^4}$ $\phi''(x) = 1$
Minimum Discrimination Information	$\phi(x) = -\log x + x - 1$	$\phi'(x) = -\frac{1}{x} + 1$ $\phi'(1) = 0$	$\phi''(x) = \frac{1}{x^2}$ $\phi''(x) = 1$

Appendix B

R-Code

B.1 Chapter 1

`log.OR`	Calculates the log-odds ratio $\log\theta$ (1.17) for a 2×2 contingency table.
`calc.log.CI`	Calculates the two-sided Wald confidence interval (1.29) of the log-odds ratio $\log\theta$ (1.17) on log-scale for a 2×2 contingency table.
`OR.pearson.test`	Calculates the value of the Pearson test statistic (1.25) for the odds ratio θ for a 2×2 contingency table in the binomial sampling scheme. Notice that this is equivalent to the score test statistic (1.26) calculated by `OR.score.test`.
`OR.score.test`	Calculates the value of the scores test statistic (1.26) for the odds ratio θ for a 2×2 contingency table in the product binomial sampling scheme.
`OR.LR.test`	Calculates the value of the Likelihood Ratio test statistic (1.27) for the odds ratio θ for a 2×2 contingency table in the product binomial sampling scheme.
`OR.CR.test`	Calculates the value of the Cressie-Read test statistic (1.28) with parameter $\lambda \neq -1, 0$ for the odds ratio θ in the product binomial sampling scheme. For $\lambda = 1$ this is equivalent to `OR.pearson.test`.
`get.Pearson.CI`	Calculates the $(1-\alpha)$ Pearson CI (1.30) for the odds ratio θ by inverting the Pearson test statistic (1.25) in the product binomial sampling scheme.
`get.score.CI`	Calculates the $(1-\alpha)$ score CI (1.30) for the odds ratio θ by inverting the score test statistic (1.26) in the product binomial sampling scheme. Notice that this is equivalent to `get.Pearson.CI`.
`get.Pearson.CI2`	Calculates the $(1-\alpha)$ score CI (1.30) for the odds ratio θ by inverting the score test statistic (1.26) in the product binomial sampling scheme **including** the correction term $\frac{n_1+n_2-1}{n_1+n_2}$ suggested by Miettinen and Nurminen [1985].

`get.LR.CI`	Calculates the $(1-\alpha)$ Likelihood ratio CI (1.30) ($T = G^2$) for the odds ratio θ by inverting the Likelihood Ratio test statistic (1.27) in the product binomial sampling scheme.
`get.CR.CI`	Calculates the $(1-\alpha)$ Cressie-Read CI (1.30) ($T = CR_\lambda, \lambda \neq -1, 0$) for the odds ratio θ by inverting the Cressie-Read test statistic (1.28) in the product binomial sampling scheme. For $\lambda = 1$ this is equivalent to `get.Pearson.CI`.
`log.genOR`	Calculates the maximum likelihood estimate (1.44) and asymptotic variance estimate (1.45) of the generalised log-odds ratio for an $I \times J$ contingency table.

B.2 Chapter 2

Remark B.2.1. A simulation and evaluation study for the CI types (Wald, score, LR, CR, see Section 1.7.1) and the continuity correction types of Section 2.1 requires calculation of the CIs in many varying probability set-ups ($\pi_i \in (0,1), i = 1,2$) for fixed sample sizes ($n_i \in \mathbb{N}, i = 1,2$). Calculating score, LR and CR CIs is time consumptive, as it requires a numerical solution of $T = \chi^2_{1,\alpha}$ within (1.30). To speed up computation time, the CIs are calculated once for fixed sample sizes in the product binomial sampling scheme and saved in a separate list within the working directory of R, which then is used to calculate (2.10) and (2.11). The required functions to calculate the lists of CIs for the Wald, score, LR and CR ($\lambda \neq -1, 0$) using the correction techniques presented in Section 2.1 is `generate.*.CIs.list`. After generating the CI lists, the coverage probability (2.10) is calculated using `cov.prob.*.LIST`, and in the same way the mean length (2.11) is calculated using `mean.length.*.LIST`.

`constant.cell.correction`	Calculates the constant cell correction ($\mathbf{C}c$) for a 2×2 contingency table (n_{ij}) with correction size $c \geq 0$ as given in (2.3).
`DD.const`	Calculates the data dependent constant correction ($\mathbf{DC}$) for a 2×2 contingency table (n_{ij}) as given in (2.4).
`ind.smooth`	Calculates the independence smoothed cell correction ($\mathbf{I}c$) for a 2×2 contingency table (n_{ij}) with correction size $c \geq 0$ as given in (2.5).
`DD.ind.smooth`	Calculates the data dependent independence smoothed correction ($\mathbf{DI}$) for a 2×2 contingency table (n_{ij}) as given in (2.6).
`single.cell.correction`	Calculates the single cell correction ($\mathbf{S}c$) for a 2×2 contingency table (n_{ij}) with correction size $c \geq 0$ as given in (2.8).
`generate.CI.cor`	Calculates the 95% Wald, score, LR or CR ($\lambda \neq -1, 0$) CIs using the correction types $\mathbf{C}c, \mathbf{DC}, \mathbf{I}c, \mathbf{DI}, \mathbf{S}c$ ($c > 0$) or no correction.

`generate.wald.CIs.list`	Generates list of 95% Wald CIs for the odds ratio for given sample sizes n_1 and n_2 in a 2×2 table with product binomial sampling scheme using a given correction. Saves the output in a file within the working directory.
`generate.score.CIs.list`	Generates list of 95% score CIs for the odds ratio for given sample sizes n_1 and n_2 in a 2×2 table with product binomial sampling scheme using a given correction. Saves the output in a file within the working directory.
`generate.CIs.LR.list`	Generates list of 95% Likelihood Ratio CIs for given sample sizes n_1 and n_2 in a 2×2 table with product binomial sampling scheme using a given correction. Saves the output in a file within the working directory.
`generate.CIs.CR.list`	Generates list of 95% Cressie-Read CIs ($\lambda \neq -1, 0$) for the odds ratio for given sample sizes n_1 and n_2 in a 2×2 table with product binomial sampling scheme using a given correction. Saves the output in a file within the working directory.
`cov.prob.wald.LIST`	Calculates the coverage probability (2.10) in the product binomial sampling scheme for the 95% Wald CI for different correction techniques, conditional on *only one zero cell*. Requires list of CIs generated by `generate.wald.CIs.list`.
`cov.prob.score.LIST`	Calculates the coverage probability (2.10) in the product binomial sampling scheme for the 95% score CI for different correction techniques, conditional on *only one zero cell*. Requires list of CIs generated by `generate.score.CIs.list`.
`cov.prob.LR.LIST`	Calculates the coverage probability (2.10) in the product binomial sampling scheme for the 95% Likelihood Ratio CI for different correction techniques, conditional on *only one zero cell*. Requires list of CIs generated by `generate.LR.CIs.list`.
`cov.prob.CR.LIST`	Calculates the coverage probability (2.10) in the product binomial sampling scheme for the 95% Cressie-Read CI ($\lambda \neq -1, 0$) for different correction techniques, conditional on *only one zero cell*. Requires list of CIs generated by `generate.CR.CIs.list`.
`mean.length.wald.LIST`	Calculates the mean length (2.11) in the product binomial sampling scheme for the 95% Wald CI for different correction techniques, conditional on *only one zero cell*. Requires list of CIs generated by `generate.wald.CIs.list`.

`mean.length.score.LIST`	Calculates the mean length (2.11) in the product binomial sampling scheme for the 95% score CI for different correction techniques, conditional on *only one zero cell.* Requires list of CIs generated by `generate.score.CIs.list`.
`mean.length.LR.LIST`	Calculates the mean length (2.11) in the product binomial sampling scheme for the 95% Likelihood Ratio CI for different correction techniques, conditional on *only one zero cell.* Requires list of CIs generated by `generate.LR.CIs.list`.
`mean.length.CR.LIST`	Calculates the mean length (2.11) in the product binomial sampling scheme for the 95% Cressie-Read CI ($\lambda \neq -1, 0$) for different correction techniques, conditional on *only one zero cell.* Requires list of CIs generated by `generate.CR.CIs.list`.

B.3 Chapter 3

`por`	Calculates the ϕ-scaled odds ratio for a 2×2 table, either in the multinomial (3.1) or product binomial sampling scheme (3.3) (cf. Lemma 3.1.1).
`por.nabla`	Calculates the partial derivatives (3.5) to (3.8) of the ϕ-scaled odds ratio in the multinomial sampling scheme.
`por.nabla.binomial`	Calculates the partial derivatives (3.11) and (3.12) of the ϕ-scaled odds ratio in the product binomial sampling scheme.
`por.SE`	Calculates the standard error estimated for the ϕ-scaled odds ratio with underlying multinomial sampling scheme based on (3.4).
`por.SE.binomial`	Calculates the standard error estimated for the ϕ-scaled odds ratio with underlying product binomial sampling scheme based on (3.9).
`Z.phi`	Calculates Wald test statistic based on the ϕ-scaled odds ratio in the product binomial sampling scheme (3.18).
`Z.phi.multinomial`	Calculates Wald test statistic based on the ϕ-scaled odds ratio in the multinomial sampling scheme (3.18).

`calc.CI.phi`	Calculates the two-sided 95% Wald CI for the ϕ-scaled odds ratio (3.19) in the product binomial sampling scheme.
`calc.CI.phi.multi`	Calculates the two-sided 95% Wald CI for the ϕ-scaled odds ratio (3.19) in the multinomial sampling scheme.
`CI.cov.percent.corrected`	Calculates the relative length (3.23) of the constant cell corrected ($c \geq 0$) two-sided 95% Wald CI for the ϕ-scaled odds ratio (3.19) in the product binomial sampling scheme.
`CI.cov.percent.corrected.multi`	Calculates the relative length (3.23) of the constant cell corrected ($c \geq 0$) two-sided 95% Wald CI for the ϕ-scaled odds ratio (3.19) in the multinomial sampling scheme.
`cov.prob.phi`	Calculates the coverage probability (3.26) of the ϕ-scaled odds ratio $(1-\alpha)$ Wald CI in the binomial sampling scheme with using continuity correction $(c > 0)$).
`cov.prob.log`	Calculates the coverage probability (3.26) of the log-odds ratio $(1-\alpha)$ Wald CI in the product binomial sampling scheme with using continuity correction $(c > 0)$).
`cov.prob.phi.multi`	Calculates the coverage probability (3.26) of the ϕ-scaled odds ratio $(1-\alpha)$ Wald CI in the multinomial sampling scheme with using continuity correction $(c > 0)$).
`cov.prob.log.multi`	Calculates the coverage probability (3.26) of the log-odds ratio $(1-\alpha)$ Wald CI in the multinomial sampling scheme with using continuity correction $(c > 0)$).
`cov.prob.phi.no.correction`	Calculates the coverage probability (3.26) of the ϕ-scaled odds ratio $(1-\alpha)$ Wald CI in the binomial sampling scheme conditional, that no sampling zeros occur.
`calc.ARL`	Calculates the average relative length (3.27) of the ϕ-scaled odds ratio $(1-\alpha)$ Wald CI in the binomial sampling scheme with using continuity correction $(c > 0)$).
`get.RL.dist`	Categorized the relative length (3.23) of all ϕ-scaled odds ratio 95% Wald CI in the product binomial sampling scheme into the categories $[0\%, 10\%], (10\%, 20\%], ..., (90\%, 100\%]$ and 100% as used in Table 3.4.

`is.RL.censored`	Indicates if a $(1-\alpha)$ Wald CI of the ϕ-scaled odds ratio had to be truncated.
`count.RL.censoring`	Counts the number of $(1-\alpha)$ Wald CIs of the ϕ-scaled odds ratio in the product binomial sampling scheme, that had to be censored as used in Table 3.4.
`rescale4`	Solves equation (3.24), i.e. for given row and column marginal probability and given value of the 2×2 ϕ-scaled odds ratio θ_0, calculates the solution π^*_{11} of the 2×2 probability table $\boldsymbol{\pi}^* \in \Delta_{2\times2}(\boldsymbol{\pi}^X, \boldsymbol{\pi}^Y)$ with fixed marginal probabilities $\boldsymbol{\pi}^X = (\pi_{1+}, 1-\pi_{1+})$ and $\boldsymbol{\pi}^Y = (\pi_{+1}, 1-\pi_{+1})$, such that $\theta^\phi(\boldsymbol{\pi}^*) = \theta_0$. Truncates if necessary.
`eval.error0`	Calculates the Type I error rate (3.28) of the independence test based on the ϕ-scaled odds ratio, i.e. $H_0 : \theta^{\phi_\lambda} = 0$ against $H_1 : \theta^{\phi_\lambda} \neq 0$ for the power divergence (1.73) in the product binomial sampling scheme using constant cell correction.
`RR.phi`	Calculates the ϕ-scaled relative risk (Definition 3.29) for a 2×2 table, either in the multinomial or product binomial sampling scheme.
`RR.nabla`	Calculates the partial derivatives of the ϕ-scaled relative risk in the multinomial sampling scheme.
`RR.nabla.binomial`	Calculates the partial derivatives (3.37) and (3.38) of the ϕ-scaled relative risk in the product binomial sampling scheme.
`RR.SE`	Calculates the standard error estimated for the ϕ-scaled relative risk with underlying multinomial sampling scheme based on (3.31).
`RR.SE.binomial`	Calculates the standard error estimated for the ϕ-scaled relative risk with underlying product binomial sampling scheme based on (3.36).
`RR.CI.phi`	Calculates the two-sided 95% Wald CI for the ϕ-scaled relative risk in the product binomial sampling scheme.
`RR.CI.phi.multi`	Calculates the two-sided 95% Wald CI for the ϕ-scaled relative risk in the multinomial sampling scheme.

B.4 Chapter 4

`get.sums`	Calculates the terms $\mathfrak{F}(i1 : i2, j1 : j2, \mathbf{x})$ from (1.40).
`model.fits`	Calculates the value of the likelihood ratio goodness-of-fit test (1.11) with degrees of freedom and p-value for the independence, local uniform association, row and column-effect association models on log-scale.
`get.MLE.log`	Calculates the MLEs of the local or global uniform association model using Lang's `mph.fit`. Requires `local.por.phi` and `global.por.phi`.
`zero.cell.prob`	Simulates the zero cell probability for a given $I \times J$ table based on the independence, saturated and local or global uniform association model on log scale using K simulated tables under the model fit. Requires `global.por.phi`
`logNI`	Calculates the logNI estimator (4.5) of the local uniform association model.
`BDNI`	Calculates the BDNI estimator (4.7) of the local uniform association model.
`MNI`	Calculates the MNI estimator (4.8) of the local uniform association model.
`WM.calc.log`	Calculates the Weighted Mean (WM) estimator (4.9) and its standard error estimated for the local or global uniform association model. Requires function from Chapter 3 (see Appendix B.3).
`MH.calc.log`	Calculates the Mantel-Haenszel (MH) estimator (4.10) for the local or global uniform association model. Requires function from Chapter 3 (see Appendix B.3).
`calc.MLE.R`	Calculates the MLE of the local or global row-effect association model ($\log \theta_{ij}^h = c_{i1}, i = 1, ..., I-1, j = 1, ..., J-1$) using Lang's `mph.fit`. Requires `local.por.phi` and `global.por.phi` from Chapter 5.
`WM.calc.R.model`	Calculates the R-Weighted Mean (R-WM) estimator (4.17) and its standard error estimated for the local or global uniform association model. Requires function from Chapter 3. (see Appendix B.3).
`MH.calc.R.model`	Calculates the R-Mantel-Haenszel (R-MH) estimator (4.20) for the local or global uniform association model. Requires function from Chapter 3 (see Appendix B.3).
`get.scores`	Calculates the normalized scores (1.56) by solving (4.24) based on the estimates of the row log-odds ratios (1.57).

`invert.global.log.odds`	Calculates the probabilities $\pi_{ij} = P(X = i, Y = j)$ in an $I \times J$ contingency table with given cumulative row and column marginal probabilities, i.e. $\nu_i = \mathbf{P}(X \leq i), i = 1, ..., I$, $\eta_j = \mathbf{P}(Y <= j), j = 1, ..., J$, and given global log-odds ratio $\log \boldsymbol{\theta}_0^G = (\theta_{ij})$ based on the inversion formula by Dale (1964) as given in (5.32).
`get.global.U.fits`	Calculates the non-iterative goodness-of-fit tests (Section 4.7) for the global uniform association model on log-scale. Returns value of the LR test statistic G^2 (1.27) and the p-value.
`local.odds.DM`	Function from the web-appendix of Kateri [2014]. Produces the matrix $\mathbf{C}$ as in (A.2), needed for producing the vector of local log-odds ratios of an $I \times J$ table.
`get.model.thetaR`	Calculates the probability table $\boldsymbol{\pi} \in \Delta_{I,J}(\boldsymbol{\pi}^X, \boldsymbol{\pi}^Y)$ with fixed marginal probabilities $\boldsymbol{\pi}^X, \boldsymbol{\pi}^Y$ and fixed local log-odds ratio $\log \boldsymbol{\theta}_0$, i.e. solves the non-linear system of equation (A.2) to (A.4).
`get.local.U.fits`	Calculates the non-iterative goodness-of-fit tests (Section 4.7) for the local uniform association model on log-scale. Returns value of the LR test statistic G^2 (1.27) and the p-value.
`get.local.R.fits`	Calculates the non-iterative goodness-of-fit tests (Section 4.7) for the local row-effect association model (1.57) on log-scale. Returns value of the LR test statistic G^2 (1.27) and the p-value.
`get.global.R.fits`	Calculates the non-iterative goodness-of-fit tests (Section 4.7) for the global ($h = G$) row-effect association model (4.16) on log-scale. Returns value of the LR test statistic G^2 (1.27) and the p-value.

B.5 Chapter 5

`local.por.phi`	Calculates the local ϕ-scaled odds ratio estimates $\hat{\boldsymbol{\theta}}^L$ (5.1.1) for an $I \times J$ contingency table.
`continuation.por.phi`	Calculates the continuation ϕ-scaled odds ratio estimates $\hat{\boldsymbol{\theta}}^{CO}$ (5.1.1) for an $I \times J$ contingency table.
`cumulative.por.phi`	Calculates the cumulative ϕ-scaled odds ratio estimates $\hat{\boldsymbol{\theta}}^C$ (5.1.1) for an $I \times J$ contingency table.
`global.por.phi`	Calculates the global ϕ-scaled odds ratio estimates $\hat{\boldsymbol{\theta}}^G$ (5.1.1) for an $I \times J$ contingency table.

`get.lambda.ij`	Calculates the solution (5.29) $\mathbb{T}^{(0)}(i,j) = \mathbf{P}(X \leq i, Y \leq j), i = 1, ..., I, j = 1, ..., J$, for given global ϕ-scaled odds ratio $\boldsymbol{\theta}_0^{\phi,G}$ and given marginal $\boldsymbol{\pi}^X$ and $\boldsymbol{\pi}^Y$.
`get.pi`	For given global essential sums $\mathbb{T}^{(0)}(i,j) = \mathbf{P}(X \leq i, Y \leq j), i = 1, ..., I, j = 1, ..., J$, and fixed marginal probabilities, calculates the corresponding $\boldsymbol{\pi} \in \Delta_{I,J}$ by (5.30).
`inv.global.phi`	Calculates the inverse of the global ϕ-scaled odds ratio, i.e. the solution $\boldsymbol{\pi} = (\pi_{ij}) \in \Delta_{I,J}(\boldsymbol{\pi}^X, \boldsymbol{\pi}^Y)$ of (5.28) using the functions `rescale4`, `get.lambda.ij` and `get.pi`.
`D.div.local`	Calculates the partial derivative $\partial\theta_{i_1j_1}^{\phi,L}(\boldsymbol{\pi})/\partial\pi_{i_2j_2}$, which are the components of $\mathbf{D}^L$ from (5.34).
`local.derivation.matrix`	Calculates the Jacobian matrix of the local ϕ-scaled odds ratio $\mathbf{D}^L$ from (5.34).
`cov.matrix.local`	Calculates the asymptotic variance-covariance matrix $\boldsymbol{\Sigma}^{\phi,L}(\boldsymbol{\pi})$ of the local ϕ-scaled odds ratio from (5.33).
`D.div.continuation`	Calculates the partial derivative $\partial\theta_{i_1j_1}^{\phi,CO}(\boldsymbol{\pi})/\partial\pi_{i_2j_2}$, which are the components of $\mathbf{D}^{CO}$ from (5.34).
`continuation.derivation.matrix`	Calculates the Jacobian matrix of the continuation ϕ-scaled odds ratio $\mathbf{D}^{CO}$ from (5.34).
`cov.matrix.continuation`	Calculates the asymptotic variance-covariance matrix $\boldsymbol{\Sigma}^{\phi,CO}(\boldsymbol{\pi})$ of the continuation ϕ-scaled odds ratio from (5.33).
`D.div.cumulative`	Calculates the partial derivative $\partial\theta_{i_1j_1}^{\phi,C}(\boldsymbol{\pi})/\partial\pi_{i_2j_2}$, which are the components of $\mathbf{D}^C$ from (5.34).
`cumulative.derivation.matrix`	Calculates the Jacobian matrix of the cumulative ϕ-scaled odds ratio $\mathbf{D}^C$ from (5.34).
`cov.matrix.cumulative`	Calculates the asymptotic variance-covariance matrix $\boldsymbol{\Sigma}^{\phi,C}(\boldsymbol{\pi})$ of the cumulative ϕ-scaled odds ratio from (5.33).
`D.div.global`	Calculates the partial derivative $\partial\theta_{i_1j_1}^{\phi,G}(\boldsymbol{\pi})/\partial\pi_{i_2j_2}$, which are the components of $\mathbf{D}^G$ from (5.34).
`global.derivation.matrix`	Calculates the Jacobian matrix of the global ϕ-scaled odds ratio $\mathbf{D}^G$ from (5.34).
`cov.matrix.global`	Calculates the asymptotic variance-covariance matrix $\boldsymbol{\Sigma}^{\phi,G}(\boldsymbol{\pi})$ of the global ϕ-scaled odds ratio from (5.33).

B.6 Chapter 6

`fit.phi.U.model`	Fits the local uniform association model (6.7) on ϕ-scale, where $\phi = \phi_\lambda$ is a member of the power divergence ($\lambda \neq 0, -1$) (see Example 1.10.4), for $\lambda = 0$, the Kullback-Leibler divergence (Example 1.10.2) is selected. Gives MLEs, p-value and count estimates under the model. Requires `mph.fit`.
`get.model.prob.U`	Simulates uniformly $K_1 \in \mathbb{N}$ multinomial probabilities $\boldsymbol{\pi}_k \in \Delta_{I,J}, k = 1, ..., K_1$, of an $I \times J$ contingency table fulfilling the underlying uniform association model (6.7) on log-scale. Uniform association parameter c lies in the interval [-4,4]. Algorithm is based on generating uniform multinomial probabilities (Smith and Tromble [2004]) and inverting generalised log-odds ratios (Appendix A.1). Compare also Example A.5.1. Requires `gen.uniform.probability` from the Appendix function file (Appendix B.8) and `get.model.thetaR` from the function file presented in Appendix B.5.
`calc.WM.U.phi`	Calculates the ϕ-WM (6.8) and its asymptotic variance based on (6.9) for the ϕ-scaled uniform model parameter of the local or global ϕ-scaled odds ratio.
`calc.MLE.R.phi`	Calculates the fit of the ϕ-scaled row-effect model based on the local or global odds ratio. Requires `mph.fit`, `local.por.phi` and `global.por.phi`.
`WM.calc.R.model.phi`	Calculates the ϕ-R-weighted-mean (ϕ-R-WM) estimator (6.12) and its asymptotic variance (6.14) for the local or global row-effect association model. Requires `local.por.phi` and `global.por.phi`.
`get.phi.scores`	Calculates the normalized row scores (4.24) based on the local or global ϕ-scaled row-effect model parameter (WM (S), WM (I) or MLE) in the power divergence (Example 1.10.4). Requires `get.scores` and `WM.calc.R.model.phi`.
`UIT.ordering`	Calculates the union intersection test (UIT) for the ordering of the scores (6.18), based on ϕ-R-WM estimation. Requires `WM.calc.R.model.phi`.

B.7 Chapter 7

`S.phi.I`	Returns the directed ϕ-scaled asymmetry measure as $I \times I$ matrix and or as composite vector $(S_{12}^{\phi}, S_{13}^{\phi}, ...)$ (Definition 7.3.1).
`dS`	Returns the derivative $\frac{\partial}{\partial \pi_{kl}} S_{ij}^{\phi}(\boldsymbol{\pi})$ as given in (7.11).
`symmetry.derivation.matrix`	Returns the Jacobian matrix $\frac{\partial S^{\phi}(\boldsymbol{\pi})}{\partial \boldsymbol{\pi}}$ based on (7.11
`W_S`	Return the value of the Wald test statistic based on the directed ϕ-scaled asymmetry measure (7.12).
`W_S0`	Return the value of the ϕ-Bowker test statistic (7.13).
`LR.sym`	Calculates the likelihood ratio statistic G^2 (1.8) for testing symmetry in an $I \times I$ contingency table.
`eval.error0`	Calculates the TIER (Type I Error Rates) for the symmetry test based on the directed ϕ-scaled asymmetry measure under $H_0 : \pi_{1+} = \pi_{+1}$ for fixed $\theta_0 \geq 1$ in the multinomial sampling scheme with probability induced by $(\theta_0, \pi_{1+}, \pi_{1+})$. Requires `rescale4` from B.4.
`tomi.scalar`	Calculates the one-dimensional asymmetry measure by Tomizawa et al. [1998] (see (7.16)) for an $I \times I$ table in the power divergence.
`scalar.S`	Calculates the value of the directed normalized scalar measure S_{ϕ}^{*} (see (7.17)) based on the directed ϕ-scaled asymmetry measure.

B.8 Appendix

`gen.uniform.probability`	Generates uniformly K_1 multinomial probabilities for an $I \times J$ contingency table using the algorithm of Smith and Tromble [2004] as presented in Appendix A.5.

References

Acock, A. C. and Stavig, G. R. [1979]. A measure of association for nonparametric statistics, *Social Forces* **57**: 1381–1386. 15

Agresti, A. [1980]. Generalized odds ratios for ordinal data, *Biometrics* **36**: 59–67. 99

Agresti, A. [1999]. On logit confidence intervals for the odds ratio with small samples, *Biometrics* **55**: 597–602. 1, 37, 38, 42, 44, 47, 48, 58, 61, 82, 85, 87, 91

Agresti, A. [2002]. *Categorical Data Analysis*, 2nd edn, John Wiley & Sons, Hoboken, NJ. 30, 102, 103, 177, 183

Agresti, A. [2003]. Dealing with discreteness: making exact confidence intervals for proportions, differences of proportions, and odds ratios more exact, *Statistical Methods in Medical Research* **12**: 3–21. 20

Agresti, A. [2007]. *An Introduction to Categorical Data Analysis*, 2nd edn, John Wiley & Sons, Hoboken, NJ. 8

Agresti, A. [2010]. *Analysis of ordinal categorical data*, 2nd edn, John Wiley & Sons, Hoboken, NJ. 8

Agresti, A. [2011]. Score and pseudo-score confidence intervals for categorical data analysis, *Statistics in Biopharmaceutical Research* **3**: 163–172. 20

Agresti, A. [2013]. *Categorical Data Analysis*, 3rd edn, John Wiley & Sons, Hoboken, NJ. 1, 8, 11, 13, 18, 22, 27, 32, 37, 43, 97, 101, 123, 184, 191

Agresti, A., Bini, M., Bertaccini, B. and Ryu, E. [2008]. Simultaneous confidence intervals for comparing binomial parameters, *Biometrics* **64**: 1270–1275. 20

Agresti, A. and Coull, B. A. [1998a]. Approximate is better than exact for interval estimation of binomial proportions, *The American Statistician* **52**: 119–126. 37

Agresti, A. and Coull, B. A. [1998b]. Order-restricted inference for monotone trend alternatives in contigency tables, *Computational Statistics and Data Analysis* **28**: 139–155. 43

Agresti, A. and Min, Y. [2002]. Unconditional small-sample confidence intervals for the odds ratio, *Biostatistics* **3**: 379–386. 20, 37

Ali, S. and Silvey, S. [1966]. A general class of coefficients of divergence of one distribution from another, *Journal of the Royal Statistical Society, Series B (Methodological)* **28**: 131–142. 2, 32

Altham, P. M. [1969]. Exact bayesian analysis of a 2×2 contingency table, and fisher's "exact" significance test, *Journal of the Royal Statistical Society. Series B (Methodological)* **31**: 261–269. 38

Anscombe, F. J. [1956]. On estimating binomial response relations, *Biometrika* **43**: 461–464. 2, 21, 38, 40, 44, 91

Balakrishnan, V. and Sanghvi, L. [1968]. Distance between populations on the basis of attribute data, *Biometrics* **24**: 859–865. 34, 66, 92

Bedrick, E. J. [1987]. A family of confidence intervals for the ratio of two binomial proportions, *Biometrics* **43**: 993–998. 20

Beh, E. J. and Davy, P. J. [2004]. A non-iterative alternative to ordinal log-linear models, *Journal of Applied Mathematics and Decision Sciences* **8**: 67–86. 3, 104, 105, 106

Beh, E. J. and Farver, T. B. [2009]. An evaluation of non-iterative methods for estimating the linear-by-linear parameter of ordinal log-linear models, *Australian and New Zealand Journal of Statistics* **51**: 335–352. 99, 100, 104, 105, 106, 127

Bhattacharya, A. K. [1943]. On a measure of divergence between two statistical populations defined by their probability distributions, *Bulletin of the Calcutta Mathematical Society* **35**: 99–109. 32

Bhattacharya, A. K. [1946]. On a measure of divergence between two multinomial populations, *The Indian Journal of Statistics* **7**: 401–406. 32

Birch, M. [1963]. Maximum likelihood in three-way contingency tables, *Journal of the Royal Statistical Society. Series B (Methodological)* **25**: 220–233. 30

Bishop, Y. M., Fienberg, S. E. and Holland, P. W. [1975]. *Discrete Multivariate Analysis: Theory and Practice*, MIT Press, Cambridge. 1, 8, 38, 40, 44

Bowker, A. H. [1948]. A test for symmetry in contingency tables, *Journal of the american statistical association* **43**: 572–574. 4, 171, 172, 181, 182, 186, 190

Brown, L. D., Cai, T. T. and DasGupta, A. [2001]. Interval estimation for a binomial proportion, *Statistical science* **16**: 101–117. 1, 2, 37

Casella, G. and Berger, R. L. [2002]. *Statistical Inference*, Duxbury. 12, 16, 19, 62, 72, 73, 153, 175, 198

Caussinus, H. [1965]. Contribution à l'analyse statistique des tableaux de corrélation, *Annales de la Faculté des Sciences de Toulouse* **29**: 77–183. 32

Clayton, D. [1974]. Some odds ratio statistics for the analysis of ordered categorical data, *Biometrika* **61**: 525–531. 3, 99, 100, 102, 103, 104, 106, 107, 108, 110, 118, 127, 168, 190

Cochran, W. G. [1954]. Some methods for strengthening the common χ^2 tests, *Biometrics* **10**: 417–451. 72, 107

Cornfield, J. [1951]. A method of estimating comparative rates from clinical data: Applications to cancer in the lung, breast, and cervix, *Journal of the National Cancer Institute* **11**: 1269–1275. 21

Cornfield, J. [1956]. A statistical problem arising from retrospective studies, *Proceedings of the third Berkeley symposium on mathematical statistics and probability*, Vol. 4, University of California Press Berkeley, pp. 135–148. 19

Costner, H. L. [1965]. Criteria for measures of association, *American Sociological Review* **30**: 341–353. 14

Cox, D. R. and Hinkley, D. V. [1974]. *Theoretical statistics*, Chapman and Hall, Boca Raton, Florida. 11

Cramér, H. [1924]. Remarks on correlation, *Scandinavian Actuarial Journal* **1924**: 220–240. 14

Cressie, N. and Read, T. R. C. [1984]. Multinomial goodness-of-fit tests, *Royal Statistical Society* **46**: 440–464. 2, 13, 33, 34, 89

Csiszár, I. [1963]. Eine informationstheoretische ungleichung und ihre anwendung auf den beweis der ergodizität von markoffschen ketten, *Publications of the Mathematical Institute of the Hungarian Academy of Science* **8**: 84–108. 2, 32

Dale, J. R. [1984]. Local versus global association for bivariate ordered responses, *Biometrika* **71**: 507–514. 30

Dale, J. R. [1986]. Global cross-ratio models for bivariate, discrete, ordered responses, *Biometrics* **42**: 909–917. 30, 140, 150

Davenport, E. C. and El-Sanhurry, N. A. [1991]. Phi/phimax: review and synthesis, *Educational and Psychological Measurement* **51**: 821–828. 14

Dennis, J. E. and Schnabel, R. B. [1996]. *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Society for Industrial and Applied Mathematics, Englewood Cliffs, NJ. 193

Diaconis, P. and Efron, B. [1985]. Testing for independence in a two-way table: new interpretations of the chi-square statistic, *The Annals of Statistics* **13**: 845–874. 37

Douglas, R., Fienberg, S. E., Lee, M.-L. T., Sampsons, A. R. and Whitaker, L. R. [1990]. Positive dependence concepts for ordinal contigency tables, *Lecture Notes-Monograph Series* **16**: 189–202. 24, 25

Espendiller, M. and Kateri, M. [2016]. A family of association measures for 2×2 contingency tables based on the ϕ-divergence, *Statistical Methodology* **35**: 45–61. 92, 147, 189

Fagerland, M. W., Lydersen, S. and Laake, P. [2013]. The mcnemar test for binary matched-pairs data: mid-p and asymptotic are better than exact conditional, *BMC medical research methodology* **13**. 178

Fagerland, M. W., Lydersen, S. and Laake, P. [2015]. Recommended confidence intervals for two independent binomial proportions, *Statistical Methods in Medical Research* **24**: 224–254. 91

Fienberg, S. E. and Holland, P. W. [1970]. Methods for eliminating zero counts in contingency tables, *Random Counts in Models and Structures* **1**: 233–260. 38, 41, 44

Fienberg, S. E. and Holland, P. W. [1972]. On the choice of flattening constants for estimating multinomial probabilities, *Journal of Multivariate Analysis* **2**: 127–134. 38, 39, 41, 44

Firth, D. [1993]. Bias reduction of maximum likelihood estimates, *Biometrika* **80**: 27–38. 41

Fisher, R. A. [1922]. On the interpretation of χ^2 from contingency tables, and the calculation of p, *Journal of the Royal Statistical Society* **85**: 87–94. 1

Fleiss, J. L. and Cohen, J. [1973]. The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability, *Educational and psychological measurement* **33**: 613–619. 99

Gart, J. J. [1962]. Approximate confidence limits for the relative risk, *Journal of the Royal Statistical Society. Series B (Methodological)* **24**: 454–463. 2, 21, 38, 40, 44, 91

Gilula, Z. and Haberman, S. J. [1986]. Canonical analysis of contingency tables by maximum likelihood, *Journal of the American Statistical Association* **81**: 780–788. 35

Gilula, Z., Krieger, A. M. and Ritov, Y. [1988]. Ordinal association in contingency tables: some interpretive aspects, *Journal of the American Statistical Association* **83**: 540–545. 35

Good, I. [1956]. On the estimation of small frequencies in contingency tables, *Journal of the Royal Statistical Society. Series B (Methodological)* **18**: 113–124. 62, 130, 173

Goodman, L. A. [1979]. Simple models for the analysis of association in cross-classifications having ordered categories, *Journal of the American Statistical Association* **74**: 537–552. 29

Goodman, L. A. [1981]. Association models and canonical correlation in the analysis of cross-classifications having ordered categories, *Journal of the American Statistical Association* **76**: 320–334. 123

Goodman, L. A. [1985]. The analysis of cross-classified data having ordered and/or unordered categories: Association models, correlation models, and asymmetry models for contingency tables with or without missing entries, *The Annals of Statistics* **13**: 10–69. 35

Goodman, L. A. and Kruskal, W. H. [1954]. Measures of association for cross classifications, *Journal of the American Statistical Association* **49**: 732–764. 14, 99, 101

Goodman, L. A. and Kruskal, W. H. [1959]. Measures of association for cross classifications. ii: Further discussion and references, *Journal of the American Statistical Association* **54**: 123–163. 14

Goodman, L. A. and Kruskal, W. H. [1963]. Measures of association for cross classifications iii: Approximate sampling theory, *Journal of the American Statistical Association* **58**: 310–364. 14

Goodman, L. A. and Kruskal, W. H. [1972]. Measures of association for cross classifications, iv: Simplification of asymptotic variances, *Journal of the American Statistical Association* **67**: 415–421. 14

Haberman, S. J. [1974a]. *The Analysis of Frequency Data*, University of Chicago Press, Chicago. 141, 197

Haberman, S. J. [1974b]. Log-linear models for frequency tables with ordered classifications, *Biometrics* **30**: 589–600. 29

Haberman, S. J. [1977]. Log-linear models and frequency tables with small expected cell counts, *The Annals of Statistics* **5**: 1148–1169. 191

Haldane, B. J. [1956]. The estimation and significance of the logarithm of a ratio of frequencies, *Annals of human genetics* **20**: 309–311. 2, 21, 38, 40, 44, 91

Harville, D. A. [1998]. Matrix algebra from a statistician's perspective, *Technometrics* **40**: 164–164. 108

Holmes, M. C. and Williams, R. [1954]. The distribution of carriers of streptococcus pyogenes among 2,413 healthy children, *Journal of Hygiene* **52**: 165–179. 103

Hout, M., Duncan, O. D. and Sobel, M. E. [1987]. Association and heterogeneity: Structural models of similarities and differences, *Sociological methodology* **17**: 145–184. 103

Kateri, M. [2014]. *Contingency Table Analysis*, Springer, Aachen. 12, 17, 23, 26, 32, 102, 103, 123, 132, 150, 208

Kateri, M. and Agresti, A. [2006]. A class of ordinal quasi-symmetry models for square contingency tables, *Statistics & Probability Letters* **77**: 598–603. 4, 36, 172, 185

Kateri, M. and Agresti, A. [2010]. A generalized regression model for binary response, *Statistics and Probability Letters* **80**: 89–95. 36, 63, 92

Kateri, M., Mohammadi, F. and Sturmfels, B. [2015]. A family of quasisymmetry models, *Journal of Algebraic Statistics* **6**: 1–16. 34, 92, 190, 194

Kateri, M. and Papaioannou, T. [1995]. f-divergence association models, *International Journal of Mathematical and Statistical Science* **3**: 179–203. 3, 35, 61, 62, 129, 133, 147, 150, 157, 189, 190

Kateri, M. and Papaioannou, T. [1997]. Asymmetry models for contingency tables, *Journal of the American Statistical Association* **92**: 1124–1131. 4, 36, 172, 185, 190

Kateri, M. and Papaioannou, T. [2000]. Measures of symmetry-asymmetry for square contingency tables, *Proceedings of the 13th Conference of the Greek Statistical Institute, Greece* . 172

Kristensen, B., B., E., Holmegaard, S. N., Krarup-Hansen, A., Transbøl, I. and H., M. [1992]. Prednisolone in the treatment of severe malignant hypercalcaemia in metastatic breast cancer: a randomized study, *Journal of Internal Medicine* **232**: 237–245. 10

Kudo, A. [1963]. A multivariate analogue of the one-sided test, *Biometrika* **50**: 403–418. 164

Kullback, S. [1959]. *Information Theory and Statistics*, John Wiley & Sons, New York. 32, 33

Kullback, S. and Leibler, R. A. [1951]. On information and sufficiency, *Annals of Mathematical Statistics* **22**: 79–86. 33, 34

Kus, V., Morales, D. and Vajda, I. [2008]. Extensions of the parametric families of divergences used in statistical inference, *Kybernetika* **44**: 95–112. 34, 190

Lang, J. B. [1996]. Maximum likelihood methods for a generalized class of log-linear models, *The Annals of Statistics* **24**: 726–752. 30

Lang, J. B. [2004]. Multinomial-poisson homogeneous models for contingency tables, *Annals of statistics* **32**: 340–383. 31, 102, 118, 147, 148, 194, 197

Lang, J. B. [2005]. Homogeneous linear predictor models for contingency tables, *Journal of the American Statistical Association* **100**: 121–134. 31, 102, 118, 147, 148, 194, 197

Lang, J. B. and Agresti, A. [1994]. Simultaneously modeling joint and marginal distributions of multivariate categorical responses, *Journal of the american statistical association* **89**: 625–632. 30

Lehmann, E. L. [1966]. Some concepts of dependence, *The Annals of Mathematical Statistics* **37**: 1137–1153. 24, 25

Lévy, P. [1925]. Calcul des probabilités, *Gauthiers-Villars, Paris* . 32

Liese, F. and Vajda, I. [1987]. *Convex Statistical Distances*, Teubner Verlagsgesellschaft, Leipzig. 33

Lindley, D. V. [1965]. *Introduction to Probability and Statistics from Bayesian Viewpoint. Part 2: Inference*, CUP Archive. 42

Lui, K.-J. and Lin, C.-D. [2003]. A revisit on comparing the asymptotic interval estimators of odds ratio in a single 2× 2 table, *Biometrical journal* **45**: 226–237. 47, 91

Mahalanobis, P. C. [1936]. On the generalised distance in statistics, *Proceedings of the National Institute of Science of India* **2**: 49–55. 32

Mantel, N. and Haenszel, W. [1959]. Statistical aspects of the analysis of data from retrospective studies of disease, *Journal of the National Cancer Institute* **22**: 719–748. 72, 107

Marselos, M., Boutsouris, K., Liapi, H., Malamas, M., Kateri, M. and Papaioannou, T. [1997]. Epidemiological aspects of the use of cannabis among university students in greece, *European Addiction Research* **3**: 184–191. 102

McCullagh, P. and Nelder, J. A. [1989]. *Generalized Linear Models*, 2nd edn, Chapman and Hall, London. 27

McNemar, Q. [1947]. Note on the sampling error of the difference between correlated proportions or percentages, *Psychometrika* **12**: 153–157. 4, 172, 176, 186, 190

Mendenhall, W. M., R., M. R., Sharkey, D. and Cassisi, N. J. [1984]. Stage t3 squamous cell carcinoma of the glottic larynx treated with surgery and/or radiation therapy, *Int. J. Radiat. Oncol. Biol. Phys.* **10**: 357–363. 9

Miettinen, O. and Nurminen, M. [1985]. Comparative analysis of two rates, *Statistics in Medicine* **4**: 213–226. 17, 19, 20, 201

Nüesch, P. E. [1966]. On the problem of testing location in multivariate populations for restricted alternatives, *The Annals of Mathematical Statistics* **37**: 113–119. 164

Nurminen, M. and Mutanen, P. [1987]. Exact bayesian analysis of two proportions, *Scandinavian Journal of Statistics* **14**: 67–77. 38

Osius, G. [2004]. The association between two random elements: A complete characterization and odds ratio models, *Metrika* **60**: 261–277. 129, 140, 145, 193

Pardo, L. [2006]. *Statistical Inference Based on Divergence Measures*, Chapman and Hall, New York. 33, 34, 184, 189

Pearson, K. [1900]. On the criterion that a given system of deviations from the probable in the case of a correlated in system of variables is such that it can be reasonably supposed to have arisen from random sampling, *Philos. Mag. Ser. 5* **50**: 157–175. (Reprinted in Karl Pearson's Early Satistical Papers, ed. E.S. Pearson. Cambridge, UK: Cambridge University Press, 1948.). 1, 12, 33, 34

Pearson, K. [1904]. Mathematical contributions to the theory of evolution xiii: On the theory of contingency and its relation to association and normal correlation, *Draper's Co. Research Memoirs, Biometric Series* **1**. (Reprinted in Karl Pearson's Early Satistical Papers, ed. E.S. Pearson. Cambridge, UK: Cambridge University Press, 1948.). 1, 8

Pearson, K. [1922]. On the χ^2 test of goodness of fit, *Biometrika* **14**: 186–191. 1

Plackett, R. [1962]. A note on interactions in contingency tables, *Journal of the Royal Statistical Society. Series B (Methodological)* **24**: 162–166. 38, 43, 44, 91

Rao, C. R. [1973]. *Linear Statistical Inference and Its Applications*, Wiley, New York. 11

Rao, R. C. [1948]. Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation, *Mathematical Proceedings of the Cambridge Philosophical Society* **44**: 50–57. 12

Rényi, A. [1961]. On measures of entropy and information, *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability* **1**: 547–561. 32

Robert, C. P. [2001]. *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation*, 2nd edn, Springer Science and Business Media, Paris. 39, 42

Rom, D. and Sarkar, S. K. [1992]. A generalized model for the analysis of association in ordinal contigency tables, *Journal of Statistical Planning and Inference* **33**: 205–212. 35, 61, 63

Serfling, R. J. [1980]. *Approximation Theorems of Mathematical Statistics*, Wiley, Baltimore, Maryland. 12, 45

Shaked, M. and Shanthikumar, J. G. [2007]. *Stochastic Orders*, Springer Science and Business Media, Tucson, Arizona. 25

Shannon, C. E. [1948]. The mathematical theory of communication, *The Bell System Technical Journal* **27**: 379–423. 2, 32, 33

Shapiro, A. [1988]. Towards a unified theory of inequality constrained testing in multivariate analysis, *International Statistical Review* **56**: 49–62. 164

Sinkhorn, R. [1967]. Diagonal equivalence to matrices with prescribed row and column sums, *American Mathematical Monthly* **74**: 402–405. 141

Smith, N. A. and Tromble, R. W. [2004]. Sampling uniformly from the unit simplex, *Johns Hopkins University, Tech. Rep* **1**: 1–6. 199, 210, 211

Tomizawa, S. [1995]. Measures of departure from marginal homogeneity for contingency tables with nominal categories, *The Statistician* **44**: 425–439. 172

Tomizawa, S., Seo, T. and Yamamoto, H. [1998]. Power-divergence-type measure of departure from symmetry for square contingency tables that have nominal categories, *Journal of Applied Statistics* **25**: 387–398. 171, 172, 184, 185, 186, 190, 211

Tutz, G. [2000]. *Die Analyse kategorialer Daten*, Oldenbourg Verlage München, Wien. 8

Tutz, G. [2012]. *Regression for Categorical Data*, Cambridge University Press, Cambridge, UK. 41

Vajda, I. [1989]. *Theory of Statistical Inference and Information*, Kluwer Academic Publishers. 32, 33

Wald, A. [1943]. Tests of statistical hypotheses concerning several parameters when the number of observations is large, *Transactions of the American Mathematical society* **54**: 426–482. 12

Warrens, M. J. [2008]. On association coefficients for the 2×2 tables and properties that do not depend on the marginal distribution, *Psychometrika* **73**: 777–789. 14

Wilks, S. S. [1938]. The large-sample distribution of the likelihood ratio for testing composite hypotheses, *The Annals of Mathematical Statistics* **9**: 60–62. 12

Woolf, B. [1955]. On estimating the relation between blood group and disease, *Annals of Human Genetics* **19**: 251–253. 107

Yule, G. U. [1900]. On the association of attributes in statistics, *Philosophical Transactions of the Royal Society* **194**: 257–319. 1, 13, 15, 16

Yule, G. U. [1903]. Notes on the theory of association of attributes in statistics, *Biometrika* **2**: 121–134. 1

Yule, G. U. [1912]. On the methods of measuring association between two attributes, *Royal Statistical Society* **75**: 579–642. 1, 13, 15, 16