

Frühwarnsysteme für Technik und Wirtschaft

Online-Detektionsverfahren für stochastische Prozesse sind verborgenen Strukturbrüchen auf der Spur

RWTH THEMEN 2/2008

Die Stochastik ist die Mathematik des Zufalls, also die formale Beschreibung und Analyse von zufallsbehafteten Phänomenen. Nach solchen muss man nicht lange suchen: Vorgänge in der Natur oder von Menschenhand geschaffene technische Systeme liefern unzählige Beispiele, bei denen der Zufall und somit die Stochastik ins Spiel kommen. Zu nennen sind hier etwa das Wetter, Wirtschafts- und Finanzdaten oder industrielle Produktionsprozesse. Häufig ist man an zeitlichen Entwicklungen interessiert und erhebt daher an ausgewählten Zeitpunkten Datenmaterial. Eine wichtige Problemstellung ist dann die Frage, ob strukturelle Zusammenhänge – zumindest über gewisse Zeitabschnitte – unverändert Bestand haben oder Änderungen unterworfen sind. Die Zeitpunkte, an denen Strukturbrüche erfolgen und neue Entwicklungen einsetzen, sind dann zu detektieren. Nur so können aus dem Ruder gelaufene technische Prozesse wieder unter Kontrolle und in ihren Optimalzustand gebracht werden. Auch die Einleitung notwendiger Maßnahmen zur Anpassung an neue ökonomische Gegebenheiten erfordert die frühzeitige Erkennung von Strukturbrüchen. Es geht also um die Konstruktion von Frühwarnsystemen, die uns rechtzeitig auf neue Entwicklungen hinweisen und somit eine Ampel-Funktion übernehmen. Hierzu dienen stochastische Monitoring-Verfahren zur Überwachung von Datenströmen. Ein Schwerpunkt der Aachener Stochastik liegt in der Erforschung der nötigen Grundlagen mit dem Ziel, die Resultate zur Konstruktion und Analyse innovativer Verfahren für komplexe technische und ökonomische Datenströme anzuwenden.

Von der Anwendung in der Industrie...

Über Erfolg oder Nichterfolg eines Produktes entscheidet oft die Produktqualität, die erheblich schwanken kann. Wichtige Beispiele sind Hightech-Produkte wie Computerchips oder Solarmodule. Hier ist die zu öko-

nomisch vertretbaren Kosten technisch erreichbare Qualität gleichzeitig die treibende Kraft für die Forschung. Die erreichbare Produktqualität schlägt sich direkt auf die Rentabilität der Milliardeninvestitionen nieder. Bei Solarmodulen ist etwa die dauerhaft abgegebene Stromleistung die relevante Größe. Für Betreiber von Solarkraftwerken ist die Qualität einer einzelnen Lieferung entscheidend. Hierzu hat der Lehrstuhl für Stochastik in Zusammenarbeit mit der TÜV Rheinland Group spezielle innovative Prüfverfahren entwickelt. Um Reklamationen und hieraus resultierenden Preisabschlägen vorzubeugen, setzt die Industrie zudem seit Jahrzehnten auf die kontinuierliche Überwachung von Produktionsprozessen durch stochastische Verfahren.

Der Weg zu guten Monitoring-Verfahren

- 1. Stochastische Modellierung:** Anwendungssituation und Problemstellung müssen durch ein geeignetes Modell mathematisch erfasst werden. Wie sieht die Welt vor und nach einem Strukturbruch aus?
- 2. Stochastische Methodenentwicklung:** Zur Lösung des gestellten Problems muss eine konkrete Methode gefunden werden. Wie sehen geeignete Prozeduren (Algorithmen) aus, um Strukturbrüche frühzeitig zu erkennen?
- 3. Mathematische Analyse:** Welche analytischen Eigenschaften hat die Lösung? Welchen Einfluss hat etwa die Abhängigkeitsstruktur des Datenstroms?
- 4. Simulation:** Durch Computersimulationen werden die finiten Eigenschaften des Verfahrens für realitätsnahe Modellwelten untersucht, auch um Empfehlungen für die Praxis ableiten zu können.

Die erfassten Qualitätsdaten werden in Echtzeit auf Auffälligkeiten und Abweichungen von der Spezifikation untersucht. Eine abrupte Änderung kann oftmals mit einem technischen Defekt in Verbindung gebracht werden, beispielsweise

dem Ausfall eines Sensors. Kontinuierliche Änderungen deuten hingegen eher auf Verschleißerscheinungen hin. Bei einem Fehlalarm können jedoch durch Produktionsverzögerungen immense Kosten entstehen. Somit besteht der Zielkonflikt, einerseits Abweichungen mit minimaler Verzögerung erkennen zu wollen und andererseits die Fehlalarmrate so gering wie möglich zu halten. Die Überwachung von realen Produktionsdaten aus der Fertigung von Mikroprozessoren (CPUs), für Computer zeigt Bild 1.

...und der Finanzökonomie...

Die modernen Finanzmärkte liefern Datenströme, in denen sich auf komplexe Weise strukturelle ökonomische Zusammenhänge und Zufallseinflüsse überlagern. Wichtige Größen wie Preisnotierungen von Rohstoffen und Aktien werden auf verschiedensten Zeitskalen – von Quartalsdaten bis hin zu Tickdaten – erfasst und stehen online zur Verfügung. Die Zufallskomponente speist sich aus verschiedenen Quellen: Konsumentenverhalten, Stimmungen von Investoren aber auch Naturkatastrophen nehmen Einfluss. Ohne die moderne Stochastik geht hier nichts!

Strukturelle Zusammenhänge zwischen ökonomischen Variablen müssen jedoch nicht zeitlich konstant sein; abrupte oder kontinuierliche Änderungen sind möglich. Dies können unerwartete Bilanzzahlen, feindliche Übernahmen oder Änderungen des Konsumentenverhaltens sowie neue Gesetze aber auch technologische Innovationen sein. Die Anwendung von Monitoring-Verfahren liegt auch hier auf der Hand. Nur so können Strukturbrüche, beispielsweise mögliche Risikofaktoren für die Rendite eines Riesen-Fonds, rechtzeitig erkannt werden.

Ein Forschungsschwerpunkt des Lehrstuhls für Stochastik liegt in der Entwicklung von Online-Detektionsverfahren für Probleme, die in einem engen Zusammenhang mit der so genannten Kointegrationstheorie stehen. Diese stellt einen ma-

thematischen Ansatz dar, ökonomische Gleichgewichtsbeziehungen nicht-stationärer Zeitreihen zu modellieren. Für seine Bahn brechenden Arbeiten wurde Clive W. Granger zusammen mit seinem Kollegen und Koautoren Robert F. Engle im Jahr 2003 mit dem ökonomischen Pendant zum Nobelpreis ausgezeichnet. Beide studierten zunächst Mathematik und Physik; Granger studierte Wirtschaftsmathematik an der Universität Nottingham. Nach seinem Ph.D. in Statistik war er als Lecturer für Statistik am Mathematik-Department in Nottingham tätig, bevor er Professor an der Universität von Kalifornien in San Diego (UCSD) wurde.

...über mathematische Modelle und Methoden...

Betrachten wir zunächst einen technischen Produktionsprozess, bei dem ein Input X_t zurzeit t theoretisch zu einem Output $Y_t = gX_t$ mit einer Proportionalitätskonstante g führt. Input und Output werden an diskreten Zeitpunkten $t=0,1,2,\dots$ erfasst. In der Realität sind nun zufällige Störeinflüsse wie Messfehler zu berücksichtigen. Ein Modell der Form $Y_t = gX_t + E_t$ mit einem stochastischen Störterm E_t , der die zufälligen Abweichungen vom wahren Wert nach oben und unten beschreibt, ist schon realistischer. Meist werden Input-Änderungen nur zeitverzögert umgesetzt. So dauert es beim Färben von flüssigen Kunststoffen bis Farbzusatz und Kunststoff homogen vermischt sind. Ein geeignetes Modell hierfür ist: $Y_{t+1} = \delta Y_t + (1-\delta)gX_{t+1} + E_{t+1}$. Der Parameter $\delta \in [0,1]$ misst die Trägheit des Prozesses. Für $\delta=0$ reagiert das System unmittelbar auf den Input. Große Werte von δ korrespondieren zu einem trägen System. Setzt man den Input X_t konstant auf a und die Proportionalitätskonstante g auf 1, dann erhält man die Gleichung $Y_{t+1} = \delta Y_t + (1-\delta)a + E_{t+1}$ die man bei einem Startwert $Y_0=0$ explizit auflösen kann:

$$Y_{t+1} = a(1-\delta^{t+1}) + \delta^t E_1 + \dots + \delta E_t + E_{t+1}$$



Angewandte Mathematik erfordert ein präzises Problemverständnis und auch die Diskussion mit Experten vor Ort. Dann können neue mathematische Lösungen zur Handlungsanweisung für die Praxis werden.
Foto: Peter Winandy

Für $|\delta| < 1$ konvergiert der nicht-stochastische Term gegen a und die Stochastik lehrt, dass sich dann die auftretende gewichtete Summe der E_t nahezu wie ein stationärer Prozess verhält. Um diese genau zu verstehen, braucht man einiges an fortgeschrittener Stochastik. Hier reicht die folgende Anschauung: Stationäre Prozesse beschreiben zufälliges „Rauschen“ gleicher Struktur um eine Mittellage, wobei sich Struktur und Mittellage nicht ändern. Die schwarze Zeitreihe in Bild 4 zeigt einen mustergültigen stationären Prozess. Im Fall $\delta = 1$ nimmt das Modell die Form $Y_{t+1} = Y_t + E_{t+1} = E_1 + \dots + E_{t+1}$ an. Jetzt werden die stochastischen Fehlerterme sukzessive aufsummiert, so dass sich die Verteilung von Y_t mit der Zeit ändert. Man spricht von einem integrierten Prozess oder einer zufälligen Irrfahrt. Die blaue und die rote Reihe in Bild 4 folgen solchen Irrfahrten. Ihr Verhalten ist grundverschieden von dem stationären Prozesse und daher bei weiteren Analysen zu berücksichtigen.

Integrierte Prozesse treten auch in der Ökonomie auf. Logarithmierte Aktienpreise folgen typischerweise einer Irrfahrt. Bei anderen Zeitreihen ist dies jedoch fraglich, so dass mathematische Methoden zur Klärung nötig sind. Bei der Kointegrationstheorie von Engle und Granger stehen sie sogar im Vordergrund.

Wagen wir einen anschaulichen Einblick:
Erfüllen zwei Variablen an allen Zeitpunkten t eine lineare Gleichung, sagen wir wieder in Form $Y_t = gX_t$ mit einer Konstanten g , dann sprechen Ökonomen von einem Gleichgewicht. Man muss jedoch Abweichungen vom Gleichgewichtspfad $y = gx$ zulassen. Welcher Natur können diese Abweichungen sein? Manche Paare integrierter ökonomischer Zeitreihen scheinen „einander zu folgen“ und um einen Gleichgewichtspfad zu schwanken. Kann aber überhaupt zwischen zwei integrierten Zeitreihen eine Langfristbeziehung der

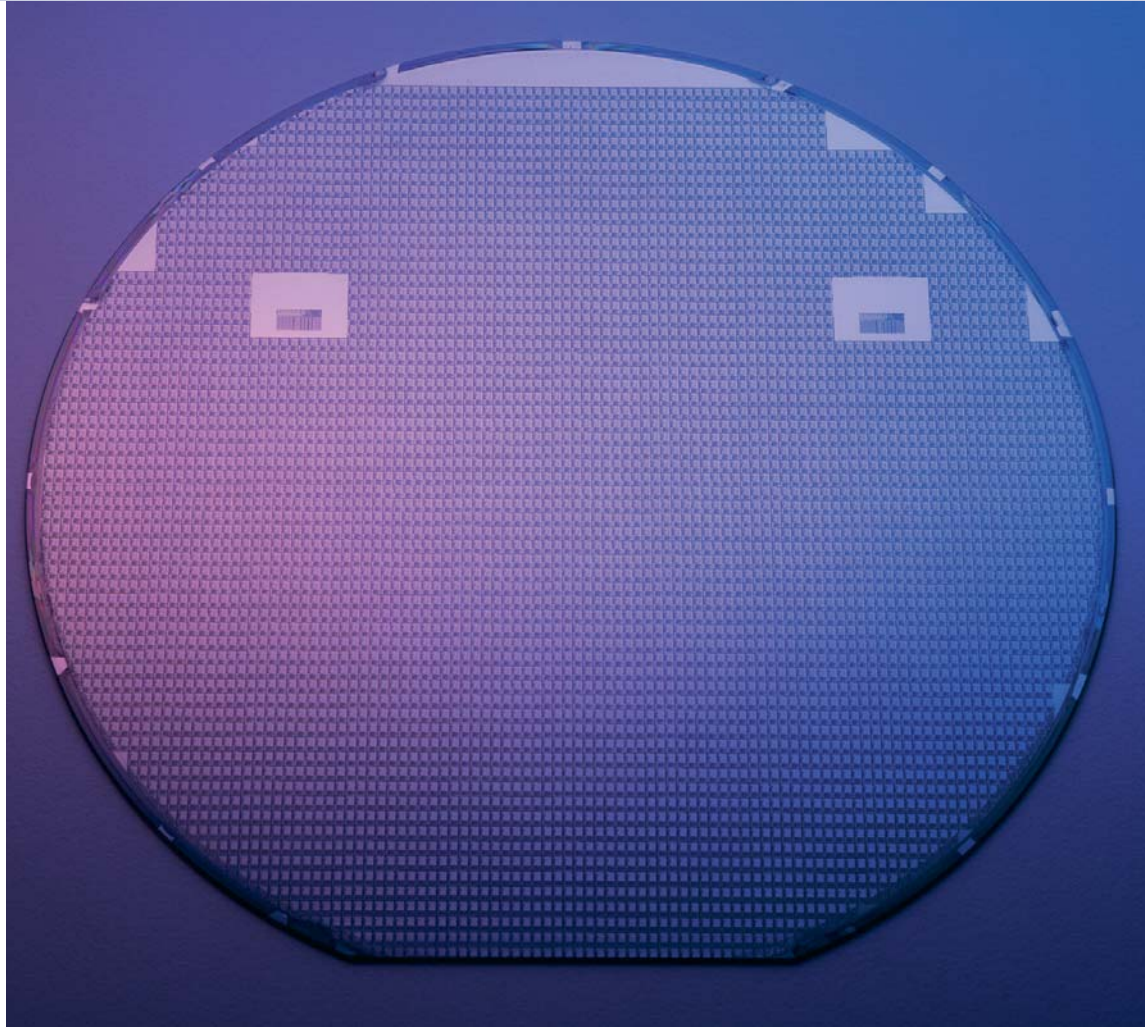
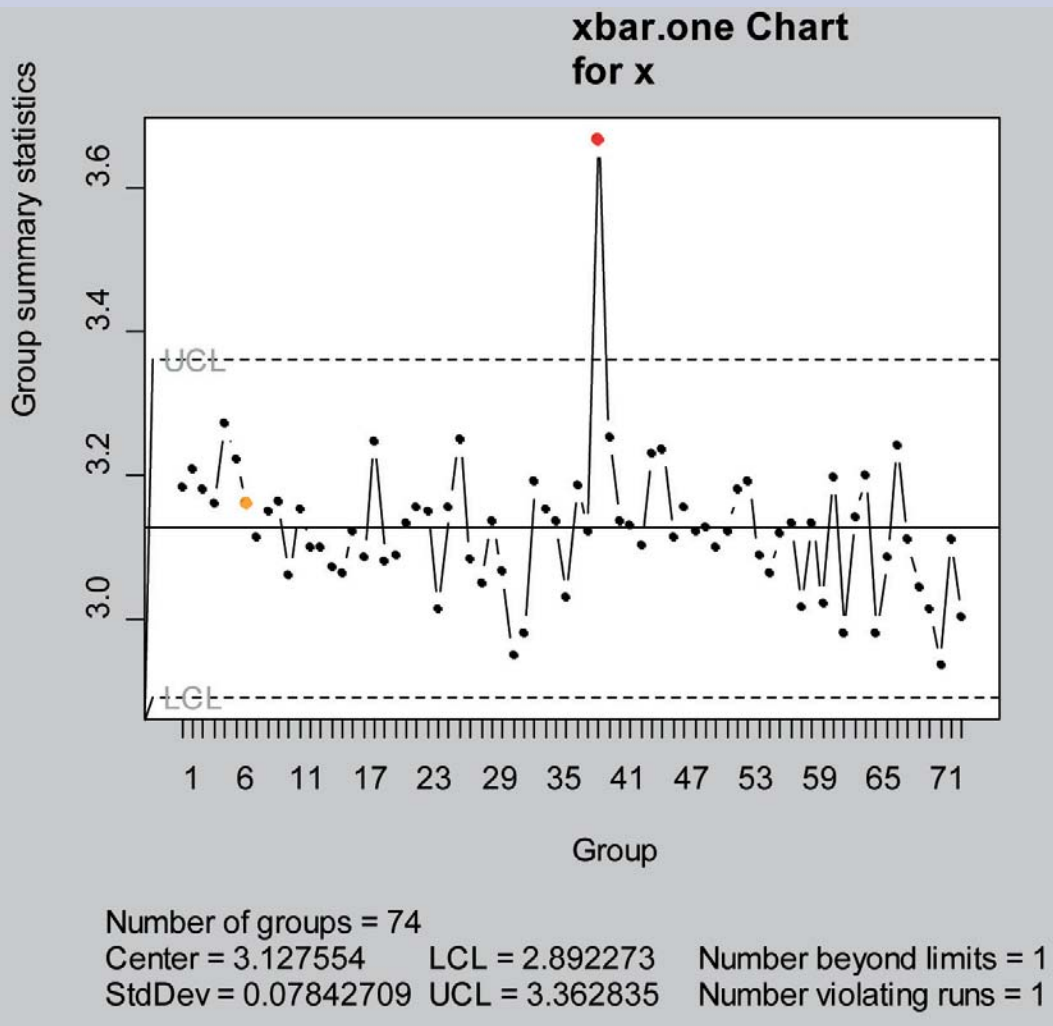


Bild 1: Ein Wafer (links) mit Dutzenden von CPUs. Ein wichtiges Qualitätsmaß ist die Strukturbreite (CD) der kleinsten Mikro- und Nanostrukturen. Die Überwachung von realen CD-Messungen durch ein



Standard-Verfahren ist rechts
gezeigt. Bis auf einen Wafer
(rot) ist der Prozess unter
Kontrolle.
Quelle:
© Christian Delbert –
Fotolia.com.

Form $y = gx$ mit stationären Abweichungen bestehen? Diese Frage hat auch Engle und Granger beschäftigt und sie zu ihrer Kointegrationstheorie geführt.

Für eine mathematisch rigore Beantwortung benötigt man profunde Kenntnisse aus diversen Gebieten der Stochastik. Die Grundideen können aber anhand der folgenden Analogie verstanden werden – für den Zusammenhang mit Perlenketten sei auf die Nobelpreisrede von Granger verwiesen.

Ein Betrunkener und sein Hund: Ein Betrunkener verlässt eine Kneipe. Er kann noch laufen, macht aber nur noch ziel- und planlos mal einen Schritt in diese Richtung und mal in jene, ohne eine spezielle Richtung zu präferieren. Seine Schritte sind zufällig und bilden einen stationären Prozess. Wie verhält es sich mit seiner Position? Bezeichnen wir mit U_i die Schrittweiten in horizontaler Richtung (x-Achse), so ist seine Entfernung von der Kneipe in x-Richtung nach n Schritten durch $X_n = U_1 + \dots + U_n$ gegeben. Dies ist eine Irrfahrt. Irrfahrten wurden bereits vom Botaniker

Robert Brown (1773-1858) beobachtet: Unter dem Mikroskop vollziehen Pollen in einer Flüssigkeit eine scheinbar regellose, unvorhersehbare Zick-Zack-Bewegung. Albert Einstein (1879-1955) gelang es 1905 dieses Phänomen physikalisch zu erklären und mathematisch zu modellieren: Die sichtbaren Partikel erhalten unregelmäßige Stöße der sich ständig bewegenden Atome und Moleküle der Flüssigkeit. Es ist das Verdienst von Norbert Wiener (1894-1966), die Verteilung der Brown'schen Bewegung mathematisch exakt konstruiert zu haben. Tragen wir die skalierten Positionen $X_i/\sqrt{n}$ gegen die Zeitpunkte i auf und verbinden die Punkte durch Linien, so lehrt der Satz von Donsker, dass das stochastische Verhalten der so konstruierten zufälligen Funktion asymptotisch durch die Brown'sche Bewegung beschrieben ist!



Wir nehmen an, dass der Trunkenbold einen jungen Hund hat. Auch junge Hunde wandern ziel- und planlos umher: Jede neue interessante Duftspur, die ihre Nase berührt, diktiert ihnen eine neue Richtung. Der Hund vollzieht also auch eine Irrfahrt Y_i mit stationären Schritten V_i . Der Trunkenbold ruft periodisch seinen Hund, damit er bei ihm bleibt. Der Hund hält kurz inne, um zu bellen. Beide taxieren ihren Abstand und korrigieren ihren nächsten Schritt so, dass sie den Abstand wenigstens ein wenig verringern. Dies führt auf die Gleichungen

$$X_n = X_{n-1} + U_n + c(Y_n - X_n) \\ Y_n = Y_{n-1} + V_n + d(X_n - Y_n)$$

für Konstanten c, d . Bild 3 zeigt die ersten zwölf Schritte dieser Irrfahrt mit Fehlerkorrektur. Folgt man den beiden einzeln, dann sieht es so aus, als ob beide ziel- und planlos – einer Irrfahrt folgend – durch die Nacht schreiten. Dennoch entfernen sie sich nicht weit voneinander: Hat man den einen gefunden, ist der andere nicht fern. Der Abstand der Pfade ist stationär und die Reihen X_n und Y_n sind kointegriert.

Verfolgen wir das Paar noch etwas länger: Bild 4 zeigt die Abstände von der Kneipe für 100 Schritte und die zugehörigen Differenzen $X_n - Y_n$. Man erkennt: Über lange Zeiträume können sich kointegrierte Zeitreihen weit vom Ausgangsniveau entfernen. Die stationären Differenzen hingegen schwanken um ein mittleres Niveau.

Sowohl bei technischen Anwendungen als auch in der Ökonomie stellt sich also oft die Frage, ob eine gegebene Zeitreihe Z_n einen stationären Prozess bildet oder sich wie eine Irrfahrt verhält. Hierzu betrachten wir das Modell $Z_n = bZ_{n-1} + E_n$

Wie viele Solarmodelle müssen geprüft werden? Ein altes statistisches Problem, für das Stochastiker der RWTH im Rahmen eines Forschungsprojektes mit dem TÜV Rheinland eine neue Lösung fanden. Foto: Peter Winandy

mit einem stationären Prozess E_n , der um 0 schwankt und einem Parameter b . Für $b=1$ erhalten wir eine Irrfahrt. Im Fall $|b| < 1$ gibt es unter gewissen Bedingungen an die Fehlerterme E_i stationäre Lösungen. Welcher dieser beiden Fälle gilt nun? Da der Parameter b unbekannt ist, bedienen wir uns der Methode der kleinsten Quadrate, um eine Schätzung



für b zu erhalten. Diese Methode – sie geht auf den Göttinger Mathematiker Carl Friedrich Gauß (1777-1855) zurück – minimiert die Summe der Quadrate $(Z_n - bZ_{n-1})^2$. Die Differenz zwischen Schätzwert und hypothetischem Wert 1 verwenden wir als Testfunktion. Es ist nun nahe liegend, von einem stationären Prozess auszugehen, wenn die Testfunktion sehr klein ist. Doch was heißt hier „sehr klein“?

Mathematische Gemeinsamkeiten technischer und ökonomischer Daten:

- Oft liegen Zeitreihen vor: Jede Beobachtung kann einem Zeitpunkt zugeordnet werden. Zu ihrer Modellierung haben sich stochastische Ansätze durchgesetzt.
- Die Beobachtungen stehen online als sequenzieller Datenstrom zur Verfügung: Nach Eintreffen einer neuen Beobachtung können die zur Verfügung stehenden Daten (neu) analysiert werden.
- Viele Modelle in beiden Bereichen weisen starke strukturelle Ähnlichkeiten auf, so dass interessierende Fragen mit demselben mathematischen Instrumentarium gelöst werden können.
- Die Datenströme sind zeitlichen Änderungen unterworfen, die sich in den Verteilungen widerspiegeln. Die Zeitpunkte von Strukturbrüchen sind oft hochinformativ.

...zur mathematischen Analyse
Wie klein muss nun die Testfunktion sein, damit wir die Irrfahrts-Hypothese verwerfen dürfen? Manchmal ist eine weitere Frage der Weg zur Lösung: Wie wahrscheinlich ist es, dass wir fälschlicherweise die eine Möglichkeit zu Gunsten der anderen verwerfen? Das Ablehnen einer Hypothese ist ja nur dann stichhaltig, wenn es bei Gültigkeit der Hypothese nur sehr unwahrscheinlich ist. Zur Berechnung solcher Fehlerwahrscheinlichkeiten benötigt man die Verteilung der Testfunktion. Für unser Problem ist diese nicht exakt berechenbar, unter anderem weil die zugrunde liegenden Prozesse durch Zeitreihenmodelle beschrieben werden müssen. Wenn Mathematiker etwas nicht exakt berechnen können, wollen sie es zumindest möglichst gut approximieren. Die moderne Stochastik stellt mit dem zentralen Grenzwertsatz ein mächtiges Instrument bereit. Das klassische Ergebnis („In großen Stichproben sind Mittelwerte näherungsweise normalverteilt“) reicht jedoch nicht aus. Man braucht solche Grenzwertsätze für allgemeine Klassen von (nicht-) stationären Prozessen. Auch zu berücksichtigen ist, dass bei Detektionsverfahren die Testfunktion an jedem Zeitpunkt berechnet wird. Aachener Stochastiker beschäftigen sich daher mit den zugehörigen zufälligen Funktionen, für die so genannte funktionale Grenzwertsätze zu beweisen sind, um die ge-

Bild 2: Die Notierungen von Aktienpreisen liefern Datenströme, in denen sich ökonomische und rein zufällige Einflüsse überlagern.
Wissenschaftler des Lehrstuhls für

Stochastik beschäftigen sich mit Verfahren zur frühzeitigen Erkennung von Strukturbrüchen.
Quelle:
©Gina Sanders - Fotolia.com

Der Veranstaltungsort der Region.

Ob Kongress, Tagung, Konzert, Ballettaufführung, Ball oder Ausstellung: hier finden Sie stets den richtigen Rahmen.

Top-Technik, variables Raumangebot, Spitzen-gastronomie, Kongress-Service etc. sind hier selbstverständlich!

Das gewisse "Mehr" bei uns: Individueller Service.

Wir informieren Sie!

Eurogress Aachen
Monheimsallee 48
52062 Aachen
Telefon 0241/9131-230
Telefax 0241/9131-200
info@Eurogress-Aachen.de

suchten Approximationen zu erhalten. Für viele Verfahren weiß man hier noch nicht sehr viel, aber es bestehen enge Beziehungen zur Brown'sche Bewegung und hieraus abgeleiteten Ito-Prozessen.

Computersimulationen

Was bringen der Einsatz moderner mathematischer Methoden und Resultate über Verfahren zur Erkennung von Strukturbrüchen konkret für die Anwendungen? Mehr, als man zunächst denkt!

Die Anwendung in Bild 5 zeigt, dass die Konstruktion von Detektionsverfahren durch Verteilungsapproximationen gute Ergebnisse liefert: Die gezeigte Zeitreihe erscheint stationär. Da sie jedoch am Computer simuliert wurde, ist das wahre Modell bekannt: Die ersten 99 Beobachtungen folgen einem stationären Zeitreihenmodell. Ab der 100ten Beobachtung liegt eine Irrfahrt vor. Das Computermodell wurde so spezifiziert, dass es als typisch für real auftretende Zeitreihen gelten kann. Ein am Lehrstuhl für Stochastik entwickeltes Detektionsverfahren – basierend auf einem neuen funktionalen Grenzwertsatz – erkennt den Strukturbruch ab der 161ten Beobachtung!

Der Computer ist auch ein wichtiges Instrument für weitere Analysen: Die Simulation von Modellwelten erlaubt es, wichtige Eigenschaften zu quantifizieren. Das Verfahren wird wiederholt (mitunter millionenfach) auf künstlich erzeugte Datensätze angewendet. Wichtige Größen können so nahezu beliebig genau berechnet werden. Bei realitätsnaher Wahl des Computermodells ergeben sich sehr verlässliche Aussagen über die Anwendbarkeit in der Praxis.

www.stochastik.rwth-aachen.de

Autor:

Univ.-Prof. Dr.rer.nat Ansgar Steland ist Inhaber des Lehrstuhls für Stochastik am Institut für Statistik und Wirtschaftsmathematik.

Bild 3: Der Trunkenbold (blau) und sein Hund (rot) vollziehen beide eine Irrfahrt. Der Fehlerkorrekturmechanismus bewirkt, dass sie sich dennoch nicht weit voneinander entfernen: Ihre Pfade sind kointegriert.

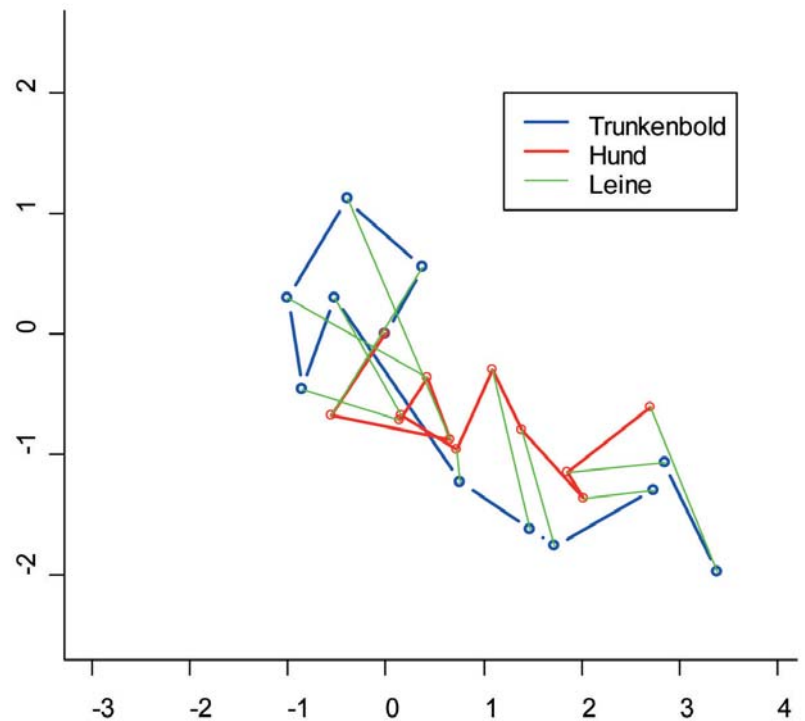


Bild 4: Horizontale Abstände (Trunkenbold: blau, Hund: rot) von der Kneipe für 100 Schritte. Beide Reihen zeigen das Verhalten einer zufälligen Irrfahrt. Die Differenzen (schwarz) bilden jedoch einen stationären Prozess, der gleichmäßig um die Nulllinie fluktuiert.

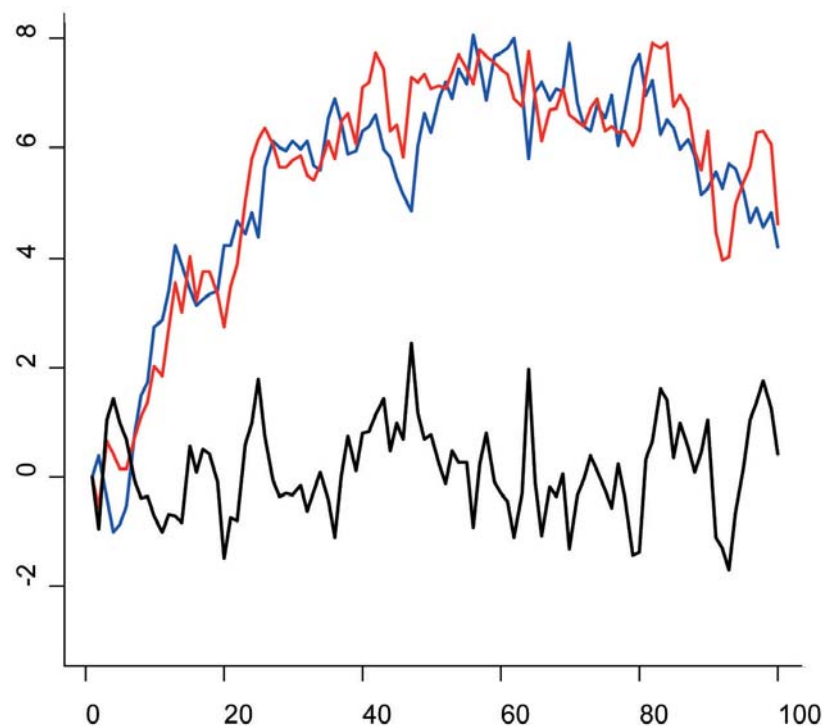


Bild 5: Eine Zeitreihe, die auf den ersten Blick stationär erscheint. Ihr stochastisches Verhalten ändert sich jedoch ab der hundertsten Beobachtung (Strukturbruch) grundsätzlich (rot). Das RWTH-Verfahren deckt dies ab der 161ten Beobachtung auf (grün).

