

# Stochastische Zusammenhänge zweier Zufallsvariablen

Jürgen Grieser

02.07.1997

## Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>1</b>
<b>2</b>	<b>Was bedeutet stochastische Unabhängigkeit?</b>	<b>2</b>
<b>3</b>	<b>Wie erkennt man stochastische Abhängigkeit?</b>	<b>4</b>
<b>4</b>	<b>Wie stark ist ein möglicher stochastischer Zusammenhang?</b>	<b>6</b>
<b>5</b>	<b>Welcher Art ist ein möglicher stochastischer Zusammenhang?</b>	<b>9</b>
<b>6</b>	<b>Beispiele</b>	<b>11</b>
6.1	Erste Experimente . . . . .	11
6.2	Systematische Untersuchungen . . . . .	13
6.3	Weitere ausgewählte Fälle . . . . .	15
	<b>Literatur</b>	<b>18</b>

## 1 Einleitung

Die mehrfache Beobachtung einer Meßgröße führt zu immer wieder unterschiedlichen Ergebnissen. So entstehen Zufallsstichproben und (wenn zeitlich sortiert) auch Zeitreihen. Beobachtet man nun mehrere Meßgrößen, so kann man sich zurecht die Frage stellen, ob man allein aus den Beobachtungsdaten Zusammenhänge zwischen diesen begründet vermuten kann. Dies geht

ausschließlich durch die Betrachtung von Ähnlichkeiten zwischen den Stichproben, die signifikant gesehen werden müssen. Solche statistischen Zusammenhänge sind aber auch nie mehr als auffällige Ähnlichkeiten. Das bedeutet, daß sie dann zwar stochastische Abhängigkeit genannt werden dürfen, was aber nicht bedeutet, daß es zwischen diesen Größen irgendeinen kausalen Zusammenhang geben muß. Diesen Unterschied sollte man immer im Hinterkopf behalten, wenn man signifikante stochastische Abhängigkeiten bewertet.

Im folgenden Abschnitt soll nun das Konzept der stochastischen Abhängigkeit vorgestellt werden. Im dritten Abschnitt wird dann ein Test auf stochastische Abhängigkeit vorgestellt, der darauf beruht, daß man die Hypothese aufstellt, daß die beiden zu untersuchenden Meßgrößen unabhängig sind, und fragt, wie unwahrscheinlich die gefundenen Meßergebnisse unter dieser Annahme sind. Eins minus dieser Unwahrscheinlichkeit ist dann die Wahrscheinlichkeit dafür, daß man sich irrt, wenn man sagt, daß die beiden Variablen voneinander abhängen. Stellt man nun aufgrund dieses Tests fest, daß die beiden Variablen voneinander wahrscheinlich nicht stochastisch unabhängig sind, so weiß man noch nicht wie stark diese Abhängigkeit ist. Ein allgemeines Maß für die Stärke des (stochastischen) Zusammenhangs wird in Abschnitt vier eingeführt. Zusätzlich wird auch hier die Hypothese getestet, daß diese Stärke nur durch Zufall entstanden ist, unter der Annahme, daß kein stochastischer Zusammenhang vorliegt. In Abschnitt fünf wird der Frage nachgegangen, von welcher Art der Zusammenhang ist. Dazu werden einem allgemeinen Maß drei eingeschränkte Maße gegenübergestellt. Dies ist der Pearson-Korrelationskoeffizient, der die Stärke des linearen Zusammenhangs abschätzt sowie der Spearman- und der Kendall-Koeffizient, welche die Stärke des monoton erklärbaren Zusammenhangs messen. Sind nun alle diese Maße gleich groß, so bedeutet dies einen linearen stochastischen Zusammenhang, ist der Pearson-Koeffizient signifikant kleiner als die anderen, und sind diese nicht signifikant unterscheidbar, so liegt ein monotoner stochastischer Zusammenhang vor, falls aber alle Koeffizienten signifikant kleiner sind als das allgemeine Maß des Zusammenhangs, so ist der Zusammenhang weder linear noch monoton. Im sechsten Abschnitt werden einige Beispiele durchgerechnet, welche die Stärke des Verfahrens sowie dessen Schwächen zeigen sollen.

## 2 Was bedeutet stochastische Unabhängigkeit?

Man kann sich leicht vorstellen, daß man von  $n$  Menschen sowohl ihre Körpergröße  $X$ , als auch ihr Gewicht  $Y$  mißt. Weiterhin kann man sich leicht vorstellen, daß diese beiden Größen nicht unabhängig voneinander sind. Die Abhängigkeit gilt aber nur im Mittel, denn man kann zwar erwarten, daß

eine große Person schwerer ist als ein kleine, aber das Umgekehrte ist im Einzelfall immer möglich. In diesem Beispiel, wo man den Zusammenhang sofort einsieht, braucht man natürlich nicht mehr zu testen, ob die Daten einen Zusammenhang suggerieren, sondern kann gleich nach dessen Stärke und nach optimalen Approximationen dieses Zusammenhangs fragen. Anders sieht die Situation aus wenn man als Wissenschaftler Neuland betritt, d.h. nach Zusammenhängen sucht, wo man nicht von vornherein weiß, daß zumindest stochastisch ein Zusammenhang besteht. Zunächst müssen wir klar sehen, was mit stochastischem Zusammenhang gemeint ist. Die wichtigste Einschränkung, die hier gemacht wird, ist, daß die Realisationen von  $X$  und  $Y$  nur paarweise untersucht werden, d.h. daß jeder Realisation von  $X$  genau ein  $Y$  zugeordnet wird und umgekehrt. Bei dem oben gegebenen Beispiel ist das klar: einer Messung ist eine Realisierung von einer Körpergröße und einem Gewicht zugeordnet. Ganz anders ist es aber z.B. bei einer Zeitreihe, die die Realisation des folgenden Prozesses ist:

$$\begin{aligned}x_t &= f(x_{t-i}, y_{t-i}) \\y_t &= g(x_{t-i}, y_{t-i}).\end{aligned}$$

Bei diesem Prozess gibt es einen deterministischen Zusammenhang zwischen  $x_t$ ,  $x_{t-1}$  und  $y_{t-1}$  und einen zwischen  $y_t$  und  $x_{t-1}$  und  $y_{t-1}$ . Damit hängen sowohl  $x_t$  als auch  $y_t$  von der gemeinsamen Vergangenheit ab. Sie hängen damit also von den vorhergehenden Werten der Zeitreihen selbst ab. Demnach ist die Information über den Zusammenhang **vollständig** in den Zeitreihen vorhanden. Bei der Analyse von Paaren der Art  $x_t$  und  $y_t$  muß er aber nicht sichtbar werden. Da der Prozess rekursiv ist, liegt ein Teil der Information über die Realisation von  $X$  zur Zeit  $t$  möglicherweise (das hängt von der konkreten Gestalt von  $f$  und  $g$  ab) in der Realisation von  $X$  und/oder  $Y$  zu viel früheren Zeiten. Die Dynamik könnte konkret so aussehen, daß man in einer endlichen Realisation (Zeitreihe) keine signifikante stochastische Abhängigkeit zu irgendeinem der vorherigen Werte der beiden Variablen finden kann. Man muß dann die Variablen stochastisch unabhängig nennen. Das zeigt, daß stochastische Unabhängigkeit nicht ausschließt, daß die beobachteten Größen sogar völlig deterministisch voneinander abhängen. Nach dieser Warnung nun zur konkreten Definition von stochastischer Unabhängigkeit:

Wir betrachten  $X$  und  $Y$  als Zufallsvariable, da es für uns zunächst zufällig erscheint ob große oder kleine Werte realisiert werden. Die Frage ist nun, ob die Wahrscheinlichkeit dafür, daß für die Variable  $X$  der Zahlenwert  $x$  realisiert wird, davon abhängt, daß für das zugeordnete  $Y$  der Wert  $y$  realisiert wird. Diese bedingte Wahrscheinlichkeit [1] nennen wir  $p(X = x|Y = y)$  oder kürzer  $p(x|y)$ . Falls die Realisation von  $X$  nicht von der Realisation von  $Y$  abhängt, muß gelten:

$$p(X = x|Y = y) = p(X = x) = p(x) \tag{1}$$

und umgekehrt auch

$$p(Y = y|X = x) = p(Y = y) = p(y). \quad (2)$$

Dabei stellen die Terme ganz rechts wieder nur verkürzte Schreibweisen dar. Die nächste wichtige Größe ist die Wahrscheinlichkeit dafür, daß das Verbundereignis  $X = x$  und  $Y = y$  eintritt. Diese Verbundwahrscheinlichkeit nennen wir  $p(X = x, Y = y)$ , oder kurz  $p(x, y)$ . Man kann sich nun durch kurzes Überlegen klar machen, daß bei stochastischer Unabhängigkeit, d.h. wenn die Gleichungen (1) und (2) gelten, die Verbundwahrscheinlichkeit  $p(x, y)$  gleich dem Produkt der Einzelwahrscheinlichkeiten  $p(x)$  und  $p(y)$  sein muß. Kennt man also die Verbundwahrscheinlichkeit und die Einzelwahrscheinlichkeiten, so kann man die stochastische Unabhängigkeit sofort erkennen. Nun ist es aber so, daß man diese im allgemeinen nicht kennt, sondern schätzen muß. Selbst wenn man sie wüßte, gäbe es noch das Problem, daß eine endliche Realisierung immer auch durch Zufall mal ein sehr seltenes Ereignis sein kann. Der im nächsten Abschnitt vorgestellte Test, berechnet nun gerade, wie unwahrscheinlich das geschätzte  $p(x, y)$  unter der Annahme  $p(x, y) = p(x) \cdot p(y)$  ist.

$X$  und  $Y$  können dabei sehr verschiedene Arten von Zufallsvariablen sein. Z.B. können die Variablen nominal skaliert sein, wie es bei  $X = \text{Farbe des Apfels}$  und  $Y = \text{Geschmack des Apfels}$  der Fall wäre. Sie müssen nur in disjunkte Klassen eingeteilt sein, d.h. ein Apfel kann nicht gleichzeitig grün und rot sein. Die Variablen können auch ordinal skaliert sein, wie es zum Beispiel die Wittereinteilung in *sehr schlecht* über *mittel* bis *sehr gut* ist. In diesem Fall ist eine Klasseneinteilung vorgegeben. Hat man metrische Variablen, z.B. Körpergröße in *cm* oder Temperaturen in  $^{\circ}C$ , so muß man diese selbst in Ereignisklassen einteilen und daraus die Wahrscheinlichkeit für das Eintreten eines Ereignisses einer bestimmten Klasse schätzen.

Zum Schluß dieses Abschnittes soll nicht unerwähnt bleiben, daß man das Konzept der stochastischen Abhängigkeit bei Zeitreihen auch selbstbezüglich und über Kreuz anwenden kann. Man erhält dann stochastische Auto-Abhängigkeit bzw. stochastische Kreuzabhängigkeit.

### 3 Wie erkennt man stochastische Abhängigkeit?

Um stochastische Abhängigkeit zu erkennen, muß man zunächst die Hypothese stochastischer Unabhängigkeit aufstellen und dann testen, ob man diese Hypothese auf hohem Niveau ablehnen kann [2]. Dazu definiert man zuerst die Nullhypothese, nämlich daß die zwei Zufallsvariablen  $X$  und  $Y$  stochastisch unabhängig sind. Wenn dem so ist, muß für die bedingten Wahr-

scheinlichkeiten  $p(x|y)$  und  $p(y|x)$  gelten:

$$\begin{aligned} p(x|y) &= p(x) \\ p(y|x) &= p(y) \end{aligned}$$

also auch

$$p(x, y) = p(x) p(y). \quad (3)$$

Diese Wahrscheinlichkeitsdichten sind nicht bekannt, können aber geschätzt werden. Dies geschieht durch Klasseneinteilung und die Kontingenztafel. Teilen wir also zunächst allgemein die Variable  $X$  in  $M$  Klassen und die Variable  $Y$  in  $R$  Klassen ein. Diese Einteilung sollte so gewählt sein, daß in jeder Klasse mindestens 5 Werte liegen [5]. Die Klassen müssen aber auf jeden Fall disjunkt sein. Die Wahrscheinlichkeiten aus Gleichung (3) können nun über die absoluten Häufigkeiten  $H$  dieser zweidimensionalen Klasseneinteilung bestimmt werden. Es wird also geschätzt:

$$\left. \begin{aligned} p(x \in x_m) &\approx \frac{H_m}{n} \\ p(y \in y_r) &\approx \frac{H_r}{n} \\ p(x \in x_m, y \in y_r) &\approx \frac{H_{m,r}}{n} \end{aligned} \right\}, \text{ für } m = 1, \dots, M \text{ und } r = 1, \dots, R.$$

Falls die Nullhypothese (d.h. Glg. (3)) richtig ist, dann ist die Antwort auf die Frage wieviel der insgesamt  $n$  Realisierungen in der Klasse  $m, r$  liegen, durch ein Bernoulli-Experiment gegeben.  $z_{m,r}(k, n)$  ist dann die Zufallsvariable, die angibt, wie wahrscheinlich (unter der Bedingung der Nullhypothese)  $k$  von  $n$  Realisierungen in der Klasse  $m, r$  liegen. Diese Variable ist binomial-verteilt mit dem Erwartungswert  $E_{m,r}$ :

$$\begin{aligned} E_{m,r} &= n p(x \in x_m, y \in y_r) \\ &= n p(x \in x_m) p(y \in y_r) = \frac{H_m H_r}{n}, \text{ für } m = 1, \dots, M \text{ und } r = 1, \dots, R. \end{aligned}$$

Für die zugehörige Varianz  $D_{m,r}^2$  gilt dann:

$$\begin{aligned} D_{m,r}^2 &= n p(x \in x_m, y \in y_r) (1 - p(x \in x_m, y \in y_r)) \\ &= \frac{H_m H_r}{n} \left( 1 - \frac{H_m H_r}{n} \right), \text{ für } m = 1, \dots, M \text{ und } r = 1, \dots, R. \\ &\approx \frac{H_m H_r}{n}, \text{ für } m = 1, \dots, M \text{ und } r = 1, \dots, R. \end{aligned}$$

Die normierte Variable von  $z_{m,r}(k, n)$  ist dann für große  $n$  näherungsweise normalverteilt, d.h.

$$\frac{z_{m,r}(k, n) - E_{m,r}}{D_{m,r}} = N(0, 1).$$

Damit ist die Summe der Quadrate dieser Normierten von  $z_{m,r}(k, n)$  eine  $\chi^2$ -verteilte Variable:

$$\chi^2 = n \sum_{m=1}^M \sum_{r=1}^R \frac{\left( z_{m,r}(k, n) - \frac{H_m H_r}{n} \right)^2}{H_m H_r} \quad (4)$$

Diese  $\chi^2$ -verteilte Zufallsvariable hat die folgenden Freiheitsgrade. Da es  $M \cdot R$  Klassen gibt ergibt dies schon mal  $M \cdot R - 1$  Freiheitsgrade. Die Anzahl der Freiheiten verringert sich jedoch um  $M - 1 + R - 1$ , da die Wahrscheinlichkeiten  $p(x \in x_m)$  und  $p(y \in y_r)$  jeweils bis auf einen Wert (der sich ergibt, weil die Summe 1 sein muß) geschätzt werden mußten. Damit ist der Freiheitsgrad der  $\chi^2$ -Verteilung  $\Phi$  gegeben durch:

$$\Phi = M \cdot R - 1 - (M + R - 2) = (M - 1)(R - 1)$$

Nun kann man mit Hilfe von Gleichung (4) testen, wie wahrscheinlich die vorliegende Realisation unter der Annahme der Nullhypothese ist. Dazu muß man nur für  $z_{m,r}(k, n)$  in Gleichung (4) die absolute beobachtete Häufigkeit  $H_{m,r}$  einsetzen. Wenn der damit berechnete Wert  $\chi_{ber}^2$  einen zu einem bestimmten Signifikanzniveau vorgegebenen Wert überschreitet, dann ist es auf diesem Niveau unwahrscheinlich, daß die Nullhypothese wahr ist.

Das Programm *zus.for* (s. Anhang C) geht noch einen Schritt weiter und berechnet zu  $\chi_{ber}^2$  den genauen Wert des Signifikanzniveaus.

## 4 Wie stark ist ein möglicher stochastischer Zusammenhang?

Um die Stärke eines stochastischen Zusammenhangs zu messen, stehen verschiedene Maße zur Verfügung. So werden in den *Numerical Recipes* [4] ein Continigenzkoeffizient und Cramer's  $V$  angegeben. Sie sind beide zwischen null und eins skaliert, aber können nicht objektiv gewertet werden. Auf der anderen Seite gibt es bedingte Maße, wie den Pearson Korrelationskoeffizient [5], dessen Quadrat angibt, welcher Anteil der Varianz der Variablen durch einen linearen Zusammenhang erfaßt werden kann. Spearman und Kendall geben ähnliche Koeffizienten an, die für monotone Zusammenhänge gelten [4].

Im Folgenden soll nun ein Maß aus der Informationstheorie [3] verwendet werden, daß die Stärke eines Zusammenhangs angibt, ohne auf Bedingungen wie Linearität oder Monotonie angewiesen zu sein [6].

Ausgangspunkt ist dabei die in einer Zufallsvariablen enthaltene Information  $H(X)$ , die gegeben ist durch:

$$H(X) = - \int_{-\infty}^{\infty} p(x) \log_2 p(x) dx. \quad (5)$$

Hat man nun eine Stichprobe bzw. eine Zeitreihe (oder eine Variable), diskret skaliert ist, so geht das Integral in die Summe über die relativen Klas-

senhäufigkeiten über:

$$H(X) = - \sum_i p(x) \log_2 p(x).$$

Hat man es nun mit zwei Stichproben zu tun, die keine gleichartige Information enthalten, so gilt  $p(x, y) = p(x) \cdot p(y)$  und die Information beider Stichproben zusammengenommen ist die Summe der Einzelinformationen. Andernfalls, d.h. falls die Stichproben  $X$  und  $Y$  gleichartige Information enthalten, ist die gemeinsame Information  $H(X, Y)$  geringer als die Summe der Einzelinformationen. Der Betrag um den die gemeinsame Information geringer ist, heißt Transinformation  $I(X, Y)$  zwischen  $X$  und  $Y$ . Demnach gilt:

$$H(X, Y) = H(X) + H(Y) - I(X, Y) \quad (6)$$

Theoretisch können die Informationen alle Werte zwischen null und  $+\infty$  annehmen. Es werden nun verschiedene Normierungen verwendet, um daraus relative Maße zu machen. In der Literatur findet man die folgenden Normierungen gewöhnlich alle unter dem Namen Redundanz  $R(X, Y)$ :

- Wenn man sich dafür interessiert, wie hoch der Anteil der Transinformation an der Information der Stichprobe  $X$  ist, definiert man sinnvoll:

$$R(X, Y) = \frac{I(X, Y)}{H(X)}.$$

- Wenn man sich dafür interessiert, wie hoch der Anteil der Transinformation an der Information der Stichprobe  $Y$  ist, definiert man sinnvoll:

$$R(Y, X) = \frac{I(X, Y)}{H(Y)}.$$

- Interessiert man sich für die maximale Redundanz, so kann man definieren:

$$R_{max}(X, Y) = \frac{I(X, Y)}{\min(H(X), H(Y))}.$$

- Interessiert man sich für die mittlere Redundanz, so setzt man die mittlere Information der Stichproben ein:

$$\overline{R(X, Y)} = 2 \frac{I(X, Y)}{H(X) + H(Y)}.$$

Zwar sind alle diese Maße zwischen null und eins beschränkt, aber nicht geeignet um sie mit anderen Maßen (insbesondere dem Pearson-Korrelationskoeffizient) zu vergleichen. Desweiteren sagt eine Redundanz von z.B. 10% allein noch nicht viel aus.

Es soll nun der Zusammenhang zwischen der Information und den Pearson-Korrelationskoeffizienten gezeigt werden, der es erlaubt, letztendlich auf ein vergleichbares Maß zu kommen. Zunächst kann aus den Gleichungen (5) und (6) die folgende Gleichung für die Transinformation hergeleitet werden:

$$I(x, y) = \int_x \int_y p(x, y) \log_2 \frac{p(x, y)}{p(x)p(y)} dx dy \quad (7)$$

Um nun den Zusammenhang zum Pearson-Korrelationskoeffizienten zu erhalten, wird für  $p(x, y)$  die zweidimensionale Gauß-Verteilung eingesetzt:

$$p(x, y) = \frac{1}{2\pi s_x s_y \sqrt{1-\rho^2}} \exp \left\{ \frac{-1}{2(1-\rho^2)} \left( \frac{(x-\bar{x})^2}{s_x^2} - \frac{2\rho(x-\bar{x})(y-\bar{y})}{s_x s_y} + \frac{(y-\bar{y})^2}{s_y^2} \right) \right\} \quad (8)$$

mit

$$\begin{aligned} p(x) &= \int_y p(x, y) dy \\ p(y) &= \int_x p(x, y) dx \end{aligned} \quad (9)$$

Die Integration von Gleichung (7) unter Berücksichtigung von Gleichung (8) und (9) führt auf folgenden Zusammenhang:

$$I(x, y) = -\frac{1}{2} \ln(1 - \rho^2)$$

Stellt man diesen Zusammenhang nach  $\rho$  um, so erhält man einen Kontingenzkoeffizienten, der aus der Information abgeleitet ist, und im Fall eines linearen Zusammenhangs gleich dem Pearson-Korrelationskoeffizienten ist:

$$\rho_{I(x,y)} = \sqrt{1 - \exp(-2I(x, y))}.$$

Diese Gleichung bildet die Transinformation, die theoretisch Werte zwischen null und unendlich annehmen kann auf das Intervall null bis eins ab. In der Praxis ist die Transinformation aber durch den kleineren Werte von  $H(X)$  und  $H(Y)$  beschränkt. Deshalb wird  $\rho_{I(x,y)}$  noch weiter normiert, in dem es auf diesen maximal möglichen Wert bezogen wird. Man erhält dann für die informationstheoretische Kontingenz  $R_I$  folgenden Zusammenhang:

$$R_I = \frac{\rho_{I(x,y)}}{\rho_{max}} = \frac{\sqrt{1 - \exp(-2I(x, y))}}{\sqrt{1 - \exp(-2 \min(H(X), H(Y)))}}.$$

Nun kann man zu  $R_I$  auch die Signifikanz ausrechnen. Dazu stellt man erneut die Nullhypothese auf, daß  $X$  und  $Y$  unabhängig sind, d.h., daß  $R_I = 0$  ist. Wenn  $X$  in  $M$  und  $Y$  in  $R$  Klassen eingeteilt ist, dann ist der Erwartungswert der Transinformation nur unter Berücksichtigung der Nullhypothese nur von den Freiheitsgraden dieser Klasseneinteilung abhängig. Man erhält den Erwartungswert  $E(I(X, Y))$ :

$$E(I(X, Y)) = \frac{1}{2}[(M - 1)(R - 1)]$$

Nun ist  $\frac{1}{2}I(X, Y)$  genau  $\chi^2$ -verteilt mit  $\Phi = (M - 1)(R - 1)$  Freiheitsgraden. Dann gilt:

$$\chi_{\Phi, \alpha}^2 = 2nI(X, Y).$$

Mit dieser Gleichung läßt sich zu jedem berechneten Wert von  $I(X, Y)$  und deshalb auch zu jedem Wert von  $R_I$  die zugehörige Signifikanz berechnen.

Bei der Berechnung der informationstheoretischen Maße ist zu beachten, daß die Klasseneinteilung möglichst so gewählt werden sollte, daß die Klassenhäufigkeiten in etwa gleichverteilt sind [6]. Dies ist im Gegensatz zu der Einteilung, die man zweckmäßigerweise bei dem  $\chi^2$ -Unabhängigkeitstest macht.

## 5 Welcher Art ist ein möglicher stochastischer Zusammenhang?

Nachdem man nun testen kann, ob und wie stark ein Zusammenhang zwischen zwei Zufallsvariablen ist, stellt sich als nächste Frage, von welcher Art er ist. Dazu werden drei mögliche Arten des Zusammenhangs unterschieden:

1. linear
2. monoton
3. nicht monoton.

Falls der Zusammenhang rein linear ist, sollte der Pearson-Korrelationskoeffizient gleich der informationstheoretischen Kontingenz sein. Theoretisch ist dies möglich, praktisch jedoch quasi unmöglich. Nur wenn in der Praxis die informationstheoretische Kontingenz signifikant größer ist als der Pearson-Korrelationskoeffizient kann man mit einer bestimmten Irrtumswahrscheinlichkeit sagen, daß der Zusammenhang nicht linear ist. Die Signifikanz kann man über die Konfidenzintervalle des Korrelationskoeffizienten berechnen. (Die Konfidenzintervalle sollten mit der Fisher-Transformierten berechnet werden [5].)

Um monotone aber nicht lineare Zusammenhänge zu sehen, haben Spearman und Kendall Methoden entwickelt [4]. Spearmans Methode basiert darauf Ranglistenplätze zu korrelieren (Pearson-Korrelation der Ranglistenplätze). Das heißt, falls großen Werten in  $X$  immer auch große Werte in  $Y$  zugeordnet sind, und kleinen Werten in  $X$  immer auch kleine Werte in  $Y$ , dann ist der Spearmankoeffizient eins, sonst geringer. Falls größeren  $X$ -Werten kleinere  $Y$ -Werte zugeordnet sind, ist der Koeffizient negativ. Auch für den Spearman-Koeffizienten  $r_s$  kann die Signifikanz getestet werden. Und zwar ist die Größe  $t_s$  mit

$$t_s = r_s \sqrt{\frac{n-2}{1-r_s^2}}$$

t-verteilt mit  $n-2$  Freiheitsgraden.

Obwohl die Spearman-Korrelation unabhängig von der Verteilung der Variablen ist, gibt es doch das Problem, das die hohen Rangplätze in die Korrelation wesentlich stärker eingehen als die niedrigen. Um dieses Problem zu umgehen, hat Kendall einen Koeffizienten  $\tau$  eingeführt. Dieser Koeffizient berücksichtigt nur die Summe der Vorzeichen des Produkts zwischen  $x_i - x_{i+j}$  und  $y_i - y_{i+j}$ . Es gehen also nur Rangplatzunterschiede ein, ohne Berücksichtigung deren Größe. Kendall's  $\tau$  kann zwischen minus eins und eins normiert werden. Kendall hat auch gezeigt, daß unter der Annahme keines monotonen Zusammenhangs der Wert  $\tau = 0$  mit der Standardabweichung

$$s_\tau = \frac{4n+10}{9n(n-1)}$$

zu erwarten ist und daß dann  $\tau$  annähernd normalverteilt ist.

Falls die Variablen nur ordinal skaliert sind, ist der Kendall-Koeffizient dem Spearman-Koeffizient überlegen. Ansonsten neigt er wegen des Informationsverlustes (als Folge der ausschließlichen Berücksichtigung der Vorzeichen) dazu, einen Zusammenhang zu unterschätzen.

Falls also aus dem Unabhängigkeitstest folgt, daß sehr wahrscheinlich ein Zusammenhang vorliegt, dessen Größe man aus der informationstheoretischen Kontingenz bekommt, und alle drei Korrelationskoeffizienten ununterscheidbar davon sind, handelt es sich um einen linearen Zusammenhang. Falls der Pearsonkoeffizient signifikant kleiner ist als die informationstheoretische Kontingenz, nicht aber der Spearman, bzw. Kendall-Koeffizient, kann man von einem monotonen Zusammenhang ausgehen. Sind aber alle Korrelationskoeffizienten kleiner als die informationstheoretische Kontingenz folgt daraus, daß ein nicht-linearer und nicht monotoner Zusammenhang vorliegt. Dann hilft meiner Meinung nach nur ein Blick auf das Scatterdiagramm weiter, auf dem man vielleicht die Form des Zusammenhangs sehen kann.

## 6 Beispiele

In diesem Abschnitt wird die Stärke der vorgestellten Strategie an einigen ausgewählten synthetischen Beispielen getestet. Dabei bedeuten die Buchstaben  $\eta$  und  $\varepsilon$  immer unkorreliertes Rauschen. Es werden Zeitreihen der Länge 300 miteinander verglichen. Ausschnitte der Zeitreihen und Scatterdiagramme befinden sich im Anhang. Zunächst einige Vorexperimente.

### 6.1 Erste Experimente

- Zeitreihe:

$$\begin{aligned} X_t &= \eta_t \\ Y_t &= \eta_t^2 \end{aligned}$$

$X_t$  und  $Y_t$  sind weder linear noch monoton voneinander abhängig, d.h. im Idealfall müsste der  $\chi^2$ -Test den Zusammenhang sehen, die Kontingenz  $R_I$  müsste eins sein und damit signifikant und die anderen Maße müssten unsignifikant klein geschätzt werden. Es wurden folgende Werte geschätzt:

Koeffizient	Wert	Signifikanz
$\chi^2$ -Abhängigkeit		1.0
Kontingenz	.92	1.0
Pearson	-.04	.52
Spearman	-.03	.34
Kendall	-.00	.09

Wie die Tabelle zeigt, wird ein ausgeprägter signifikanter Zusammenhang gefunden, der weder linear noch monoton ist. (Das Streudiagramm läßt diesen klar erkennen.)

- Was ist, wenn der Zusammenhang verrauscht ist?

$$\begin{aligned} X_t &= \eta_t \\ Y_t &= \eta_t^2 + .4\varepsilon_t \end{aligned}$$

Koeffizient	Wert	Signifikanz
$\chi^2$ -Abhängigkeit		1.0
Kontingenz	.81	1.0
Pearson	-.05	.66
Spearman	-.06	.67
Kendall	-.05	.80

Trotz des verrauschten Zusammenhangs kommt das gleiche klar interpretierbare Bild heraus. Es gibt sicher einen Zusammenhang, der sicher nicht linear und nicht monoton ist.

- Nun ein Beispiel einer monotonen Abhängigkeit:

$$\begin{aligned} X_t &= \eta_t \\ Y_t &= \eta_t^3 \end{aligned}$$

Koeffizient	Wert	Signifikanz
$\chi^2$ -Abhängigkeit		1.0
Kontingenz	.82	1.0
Pearson	.82	1.0
Spearman	1.0	1.0
Kendall	1.0	1.0

Da der Zusammenhang deterministisch monoton ist sind die Koeffizienten, die einen solchen Zusammenhang suchen, gleich eins. Sowohl der Pearsonkoeffizient als auch die Kontingenz sind kleiner. Aber alle Maße detektieren einen stochastischen Zusammenhang (der in diesem, wie im ersten Beispiel deterministisch ist).

- Das Gleiche verwechselt:

$$\begin{aligned} X_t &= \eta_t \\ Y_t &= \eta_t^3 + .4 \varepsilon_t \end{aligned}$$

Koeffizient	Wert	Signifikanz
$\chi^2$ -Abhängigkeit		1.0
Kontingenz	.83	1.0
Pearson	.81	1.0
Spearman	.85	1.0
Kendall	.71	1.0

Auch hier wird eindeutig ein Zusammenhang gesehen. Nun ist einer der Koeffizienten monotoner Zusammenhänge größer als der Pearsonkoeffizient, der andere kleiner. Die Kontingenz ist nicht signifikant größer als der Pearsonkoeffizient, was ein Blick auf die Konfidenzintervalle des Pearsonkoeffizienten zeigt (s. Anhang), wohl aber der Spearmankoeffizient. Demnach wird eine signifikante monotone Abhängigkeit gefunden. Stärkere als monotone Nichtlinearitäten sind richtiger Weise nicht zu erkennen.

- Da die Abbildung  $y = x^3$  sehr sensibel auf den Betrag der seltenen Werte des Gauß-Rauschens reagiert, soll noch eine weitere monotone Abbildung betrachtet werden. Die folgende Abbildung unterdrückt die Flügel der Gauß-Verteilung, anstatt sie, wie dies  $y = x^3$  tut, zu vergrößern.

$$\begin{aligned} X_t &= \eta_t \\ Y_t &= \tanh(2 \eta_t) + .4 \varepsilon_t \end{aligned}$$

Koeffizient	Wert	Signifikanz
$\chi^2$ -Abhängigkeit		1.0
Kontingenz	.86	1.0
Pearson	.81	1.0
Spearman	.87	1.0
Kendall	.66	1.0

Wieder wird der Zusammenhang klar gesehen. Die Kontingenz ist hoch signifikant größer als der Pearsonkoeffizient, der Spearmankoeffizient sogar höchst signifikant. Daß deutet wieder stark auf den monotonen Zusammenhang hin, den der Kendallkoeffizient, der auf jeden Fall rauschanfälliger ist, nicht so stark sieht.

- Nun ein AR(1)-Prozeß als lineares Beispiel:

$$X_t = .5 X_{t-1} + \eta_t$$

Koeffizient	Wert	Signifikanz
$\chi^2$ -Abhängigkeit		1.0
Kontingenz	.53	1.0
Pearson	.50	1.0
Spearman	.47	1.0
Kendall	.33	1.0

Wieder wird ein Zusammenhang eindeutig gefunden. Weder der Kendallkoeffizient noch die Kontingenz unterscheiden sich jedoch signifikant vom Pearsonkoeffizient (dessen Konfidenzintervalle sind im Anhang).

Nachdem diese ersten Experimente erfolgreich waren, werden nun ein linearer und ein nichtlinearer autoregressiver Prozess systematischer untersucht.

## 6.2 Systematische Untersuchungen

In diesem Abschnitt wird die Autoabhängigkeit eines linearen AR(1)-Prozesses und der chaotischen Feigenbaumdynamik bis zu zehn Zeitschritten systematisch untersucht. Dabei wird beim AR(1)-Prozeß ein großer Autoregressionskoeffizient von .8 verwendet, um einen weiten Bereich von Autokorrelationen bei zehn Verzögerungen abzudecken.

1. Der AR(1)-Prozeß

$$X_t = .8 X_{t-1} + \eta_t$$

hat die theoretische Autokorrelationsfunktion

$$r_\tau = .8^\tau$$

Tabelle 1: Verschiedene Autokorrelationsmaße beim AR(1)-Prozeß

$\tau$	Pearson theor.	Kontingenz	Pearson beob.	Spearman	Kendall
1	.800	.856	.818	.802	.607
2	.640	.745	.681	.679	.489
3	.512	.636	.569	.568	.400
4	.410	.514	.480	.474	.324
5	.328	.454	.379	.387	.258
6	.262	.376	.323	.336	.225
7	.210	.329	.259	.269	.181
8	.168	.291	.222	.234	.159
9	.134	.316	.188	.209	.144
10	.107	.321	.133	.152	.103

In Tabelle 1 sind nun alle Zusammenhangsmaße gegen die Zeitverschiebung bis zu zehn Zeitschritten eingetragen. Die Zusammenhänge sind alle mindestens signifikant. Für Zeitverschiebungen bis drei und ab neun ist die Kontingenz signifikant größer als der Pearson-Koeffizient und damit zu groß. Dies hat zur Folge, daß man einen nichtlinearen Zusammenhang vermuten könnte, der nicht vorhanden ist. Auch der Pearson-Koeffizient überschätzt die Autokorrelation in diesem Beispiel, jedoch liegt der wahre Autokorrelationskoeffizient in allen zehn Fällen innerhalb des geschätzten 90 % Konfidenzintervalls. Der Spearman-Koeffizient ist etwa gleich dem Pearson-Koeffizient, der Kendall-Koeffizient wie üblich etwas kleiner.

## 2. Die Feigenbaumdynamik

Die Feigenbaumdynamik folgt der Abbildung

$$X_t = a X_{t-1}(1 - X_{t-1}).$$

Sie ist damit eine quadratische autoregressive Abbildung. Tabelle 2 gibt die Autokorrelationsmaße für die ersten zehn Zeitverschiebungen dieses nichtlinearen Prozesses für den Fall  $a = 4$  (voll entwickeltes Chaos) an. Der  $\chi^2$ -Test sieht bis zur Zeitverschiebung  $\tau = 4$  einen höchst signifikanten Zusammenhang. Ab  $\tau = 7$  findet er keinen signifikanten Zusammenhang mehr. Der Kontingenzkoeffizient wird mit zunehmendem  $\tau$  immer kleiner, bleibt aber bis  $\tau = 9$  hoch signifikant, während alle anderen Maße nicht signifikant von null unterscheidbar sind. Hier wird also die nichtlineare Struktur richtig und hoch signifikant erkannt.

Tabelle 2: Verschiedene Autokorrelationsmaße beim Feigenbaumprozeß

$\tau$	Kontingenz	Pearson	Spearman	Kendall
1	.929	-.010	-.012	-.012
2	.670	.075	.046	.021
3	.432	-.084	-.083	-.058
4	.365	.095	.096	.071
5	.280	.050	.059	.035
6	.342	-.031	-.033	-.021
7	.263	-.069	-.077	-.051
8	.247	.061	.062	.043
9	.246	-.066	-.060	-.040
10	.220	-.019	.000	.000

### 6.3 Weitere ausgewählte Fälle

Im letzten Abschnitt der Anwendungen werden noch einmal verschiedene Zeitreihenpaare erzeugt und analysiert. Dabei wurden alle Zahlenwerte zwischen null und eins beschränkt. Folgende 10 Zeitreihenpaare der Länge 300 werden untersucht:

1.

$$\begin{aligned} X_t &= \eta_t \\ Y_t &= .7 X_t + .3 \varepsilon_t \end{aligned}$$

2.

$$\begin{aligned} X_t &= .5 \eta_t \\ &+ \frac{1}{6} [\cos(2\pi \frac{t}{11.02}) + 1] \frac{1}{2} \\ &+ \frac{1}{6} [\cos(2\pi \frac{t}{97.887}) + 1] \frac{1}{2} \\ &+ \frac{1}{6} [\cos(2\pi \frac{t}{60.5}) + 1] \frac{1}{2} \\ Y_t &= .5 X_t + .25 \varepsilon_t + \frac{1}{4} [\cos(2\pi \frac{t}{80}) + 1] \frac{1}{2} \end{aligned}$$

3.

$$\begin{aligned} X_t &= \varepsilon_t \\ Y_t &= 4 Y_{t-1} (1 - Y_{t-1}) \end{aligned}$$

4.

$$\begin{aligned} X_t &= 4 X_{t-1} (1 - X_{t-1}) \\ Y_t &= 4 Y_{t-1} (1 - Y_{t-1}) \end{aligned}$$

5.

$$\begin{aligned} X_t &= .5 \eta_t \\ &+ \frac{1}{6} [\cos(2\pi \frac{t}{11.02}) + 1] \frac{1}{2} \\ &+ \frac{1}{6} [\cos(2\pi \frac{t}{97.887}) + 1] \frac{1}{2} \\ &+ \frac{1}{6} [\cos(2\pi \frac{t}{60.5}) + 1] \frac{1}{2} \\ Y_t &= .5 \eta_t \\ &+ \frac{4}{30} [\cos(2\pi \frac{t}{11.02} + \frac{\pi}{4}) + 1] \frac{1}{2} \\ &+ \frac{1}{6} [\cos(2\pi \frac{t}{97.887} + \frac{\pi}{3}) + 1] \frac{1}{2} \\ &+ \frac{1}{5} [\cos(2\pi \frac{t}{60.5} + \pi) + 1] \frac{1}{2} \end{aligned}$$

6.

$$\begin{aligned} X_t &= 3.7 X_{t-1} * (1 - X_{t-1}) + .06 (Y_{t-1} - X_{t-1}) \\ Y_t &= 3.7 Y_{t-1} * (1 - Y_{t-1}) + .06 (X_{t-1} - Y_{t-1}) \end{aligned}$$

7.

$$\begin{aligned} X_t &= \eta_t \\ Y_t &= .5 X_t + .5 \varepsilon_t \end{aligned}$$

8.

$$\begin{aligned} X_t &= \eta_t \\ Y_t &= Y_{t-1} * (1 - Y_{t-1}) (3.4699456 + .1 \cdot X_t) \end{aligned}$$

9.

$$\begin{aligned} X_t &= [\cos(2\pi \frac{t}{62.765}) + 1] \frac{1}{2} \\ Y_t &= Y_{t-1} * (1 - Y_{t-1}) (3.4699456 + .1 \cdot X_t) \end{aligned}$$

10.

$$\begin{aligned} X_t &= X_{t-1} * (1 - X_{t-1}) (3.4699456 + .1 \cdot \eta_t) \\ Y_t &= Y_{t-1} * (1 - Y_{t-1}) 3.5699456 \end{aligned}$$

Ergebnisse der Analyse der zehn ausgewählten Zeitreihenpaare (Anhang A enthält die ausführlichen protokolle. In Anhang B sind Zeitreihenausschnitte und Streudiagramme):

1. Der lineare Zusammenhang wird eindeutig erkannt.
2. Der lineare Zusammenhang wird eindeutig erkannt.
3. Es wird kein signifikanter Zusammenhang erkannt.
4. Es wird höchst signifikant (99.89%) ein schwacher und mit 99.9 % nichtlinearer Zusammenhang erkannt (Kontingenz .3). Da aber die beiden (zwar gleichen) Dynamiken mit unterschiedlichen Anfangswerten gestartet wurden, wird man vergeblich einen funktionalen Zusammenhang suchen. Die Gemeinsamkeit liegt darin, daß beide Zeitreihen auf dem selben Attraktor liegen.
5. Es wird kein signifikanter Zusammenhang erkannt, denn er ist zu schwach um ihn mit diesen Methoden bei der gewählten Zeitreihenlänge zu erkennen.
6. Es wird höchst signifikant (99.89%) ein schwacher und mit 99.9 % nichtlinearer Zusammenhang erkannt (Kontingenz .3). Es handelt sich um Hyperchaos, das ohne die Untersuchung von Autokorrelationsmaßen nicht weiter identifizierbar sein wird.
7. Der lineare Zusammenhang wird eindeutig erkannt.
8. Es wird kein signifikanter Zusammenhang erkannt. Die erste Reihe enthält eine Zufallsschwankung, die den Kontrollparameter der Dynamik der zweiten Reihe stört. Diese Störung ist zu schwach, um von dem Verfahren gesehen zu werden.
9. Der Zusammenhang wird eindeutig erkannt und ist höchst signifikant nichtlinear. Die langsamen Veränderungen des Kontrollparameters der Dynamik der zweiten Reihe durch die harmonische Funktion der ersten Reihe wird klar gesehen.
10. Die Ähnlichkeit zwischen der zufällig gestörten Dynamik der ersten Reihe und der ungestörten Dynamik der zweiten Reihe wird von allen Maßen klar gesehen. Die Kontingenz ist höchst signifikant größer als die anderen Maße. Daran erkennt man die nichtlineare Abhängigkeit der Dynamik vom Wert des Kontrollparameters.

## Literatur

- [1] K. Bosch. *Elementare Einführung in die angewandte Statistik*. Vieweg, Bern, Braunschweig, 1987.
- [2] K. Bosch. *Elementare Einführung in die Wahrscheinlichkeitsrechnung*. Vieweg, Bern, Braunschweig, 1987.
- [3] H. Haken. *Synergetik*. Springer, Berlin, 1990.
- [4] W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery. *Numerical Recipes*. Cambridge University Press, Cambridge, 1992.
- [5] C.-D. Schönwiese. *Praktische Statistik*. Gebrüder Bornträger, Berlin, Stuttgart, 1985.
- [6] H. Weingärtner. Korrelation und Information. *Meteorol. Rundschau*, 38:1–8, 1985.