Lang- und Kurzzeitgedächtnis

Definition: **Gedächtnis** einer Zeitreihe

$$M = \sum_{k=-\infty}^{\infty} \left|\rho(k)\right|$$

Eine Zeitreihe hat **kurzes Gedächtnis**

$$\Leftrightarrow M < \infty$$

Arten von Kausalitätsbeziehungen

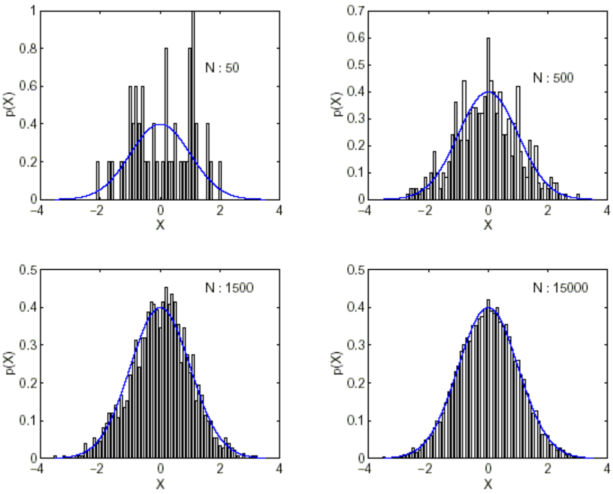
Beziehung	**Eigenschaften der Kreuzkorrelation**
X verursacht Y	$\rho_{xy}(k) \neq 0$ für manche $k > 0$
Y verursacht X	$\rho_{xy}(k) \neq 0$ für manche $k < 0$
Instantane Kausalität	$\rho_{xy}(0) \neq 0$
Rückkopplung	$\rho_{xy}(k) \neq 0$ für manche $k < 0$ und manche $k > 0$
Y verursacht nicht X	$\rho_{xy}(k) = 0$ für alle $k < 0$
X und Y sind unabhängig	$\rho_{xy}(k) = 0$ für alle k

Tests und Trenderkennung

(Oft) gewünschte oder geforderte Eigenschaften von Zeitreihen:

Eigenschaft	**Anmerkungen**	**Test**
Normalität	gegeb. Transformation der Daten	Kolmogorov-Smirnov, Shapiro-Wilks, χ^2-Test
Ergodizität	pragmatisch: Sättigung der Statistik	*für empirische Daten nicht testbar*
Linearität		*lineare Modelle*
Stationarität = Trendfreiheit	• Stationarität des Mittelwertes („Trend" i.e.S.)	Mann-Kendall-Test
	• Stationarität der Varianz (Homoskedastizität)	Bartlett-Test, Levene-Test
	• allgemein: Konstanz aller Momente	Witt-Test

Schätzung der Wahrscheinlichkeitsverteilung



Variation des Stichprobenumfangs bei gleicher Zahl der Klassen
(*Frankenberg 2002*)

Box-Cox Transformation

Ziel: Annäherung der beobachteten (schiefen) Wahrscheinlichkeitsdichte der Werte an eine Normalverteilung

Nebeneffekt: Bestimmte Instationaritäten lassen sich so ebenfalls vermindern.

$$y(t) = \frac{1}{\lambda}(x(t)^{\lambda} + c) \quad \text{für } \lambda \neq 0$$

$$y(t) = \ln(x(t) + c) \quad \text{für } \lambda = 0$$

Wähle c so, dass der Ausdruck in der Klammer immer > 0 wird

Ergodizität

Definition: Ein dynamisches System heisst ***ergodisch***, wenn der tatsächlich verfügbare Phasenraum dicht ausgefüllt wird

=> "Alles was vorkommen kann, kommt auch vor", „Es gibt keine echten Überraschungen", d.h., qualitativ neues Systemverhalten -> gesättigte Statistik

Konsequenzen z.B.:

- der beobachtete Datensatz umfasst die absoluten Extrema
- mit zunehmender Datensatzgröße konvergieren die Mittelwerte gegen den wahren Wert

Relevanz:

- Basis (nahezu) aller Verfahren, insbesondere zur Vorhersage
- In der Praxis aber ein ***unüberprüfbares Postulat***, da für beobachtete endliche Zeitreihen kein automatisches Testverfahren entwickelt werden kann

Theoretischer Test auf Ergodizität

$\lim_{n \to \infty} \sigma(\bar{x}_n) = 0$ -> notwendige und hinreichende Bedingung

$\lim_{n \to \infty} \frac{1}{n} \sum_{k=0}^{n} \rho_{ac}(k) = 0$ -> hinreichende Bedingung

-> nur für $n = \infty$ durchführbar!

Tests und Trenderkennung

(Oft) gewünschte oder geforderte Eigenschaften von Zeitreihen:

Eigenschaft	**Anmerkungen**	**Test**
Normalität	gegeb. Transformation der Daten	Kolmogorov-Smirnov, Shapiro-Wilks, χ^2-Test
Ergodizität	pragmatisch: Sättigung der Statistik	*für empirische Daten nicht testbar*
Linearität		*lineare Modelle*
Stationarität = Trendfreiheit	• Stationarität des Mittelwertes („Trend" i.e.S.) • Stationarität der Varianz (Homoskedastizität) • allgemein: Konstanz aller Momente	Mann-Kendall-Test Bartlett-Test, Levene-Test Witt-Test

Starke Stationarität

Gegeben sei eine Zeitreihe $x(t)$ mit Wahrscheinlichkeitsverteilung $p(x)$

Definition: Eine Zeitreihe heißt ***stark stationär***, wenn für alle k die k-ten bedingten Wahrscheinlichkeiten nicht von der Zeit abhängen:

$$p(x_{t_1+k} | x_{t_1+k-1}, x_{t_1+k-2}, \ldots, x_{t_1}) = p(x_{t_2+k} | x_{t_2+k-1}, x_{t_2+k-2}, \ldots, x_{t_2}) \text{ für beliebige } t_1, t_2$$

Folgerungen:

1. Alle Momente hängen nicht von der Zeit ab
2. Autokorrelation hängt nicht von der Zeit ab.

Problem: Die Überprüfung der starken Stationarität für experimentelle Daten ist i.a. nicht möglich, da unendlich viele Tests erforderlich.

Schwache Stationarität

-> „pragmatischer Kompromiss"

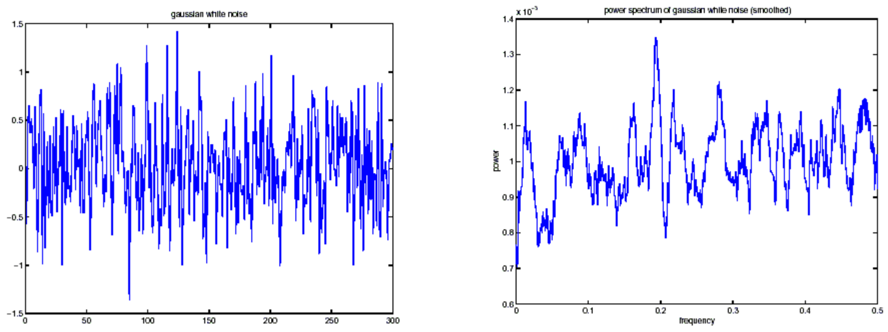
Definition: Eine Zeitreihe heißt *schwach stationär*, wenn ihr ***erstes*** und ***zweites Moment*** nicht von der Zeit abhängen:

$$E(x(t)) = \bar{x} = konst. \quad \neq \bar{x}(t)$$
$$E((x(t) - \bar{x})^2) = \sigma^2 = konst. \quad \neq \sigma^2(t)$$

Beispiel für starke Stationarität

Weißes Rauschen, definiert durch:

- zufällige Daten
- nicht autokorreliert
- Varianz gleichmäßig über alle Frequenzen verteilt



(a) Gaussian white noise process (b) Spectrum of a white noise process

Stationaritätstests

Typen von Instationaritäten
- homogene, explosive, ... Instationarität
- linearer, polynomialer, exponenteller, ..., Trend
- (lange) Periodizitäten
- Heteroskedastizität

Typen von Stationaritätstests
- auf der Werteverteilung basierend
- auf der Fourierzerlegung basierend
- direkte Trenderkennung

Prinzip: "Fenstertechnik„ (*vergl.: Histogramm*)
- Aufteilung des Datensatzes in gleichlange Stücke (***Fenster***)
- Bestimmung statistischer Merkmale in jedem Fenster
- Berechnung der Variabilität des Merkmals von Fenster zu Fenster
- Ermittlung der Signifikanz

Trendanalyse: „Klassischer“ Ansatz

Woldsches Theorem (*Wold 1934*): das additive Komponentenmodell

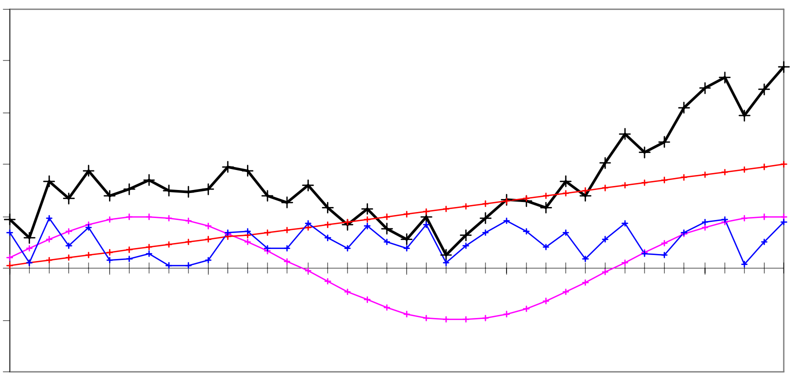
$$X(t) = f(X(t), Y(t)) \quad + S(t) \quad + T(t) \quad + \eta(t)$$

$Y(t)$	externe Faktoren
$S(t)$	saisonale Komponente
$T(t)$	deterministischer Trend
$\eta(t)$	stationäres Rauschen

Globaler monotoner Trend: "im Mittel wächst $X(t)$ an / fällt ab"
=> Trend des Mittelwerts (= 1. Moment der Verteilung)

Trendanalyse: „Klassischer“ Ansatz

$$X(t) = \boxed{f(X(t), Y(t))} + S(t) + T(t) + \eta(t)$$



Test auf Trend des Mittelwertes

Intuitiver Ansatz:

- Bestimmung der bivariaten Pearson-Korrelation zwischen $x=x(t)$ und t und Prüfung auf Signifikanz mittels *t*-Test

- Quantifizierung des Trends: Regression zwischen $x=x(t)$ und t:

$$x = x(t) = a + b \cdot t$$

Limitationen dieses Ansatzes:

- Pearson-Korrelation und lineare Regression erfassen nur ***lineare*** und ***globale*** Zusammenhänge
- *t*-test setzt ***Normalverteilung*** und ***unabhängige*** Stichproben voraus

-> Alternative: Mann-Kendall-Test, Stationaritätstest nach Witt

Mann-Kendall-Test

Ziel: Überprüfung auf Trends

Motivation: viele statistische Verfahren setzen Trendfreiheit voraus

Prinzip: Vergleich der Anzahl der Konkordanzen ($x(t_1) > x(t_2)$ für $t_1 > t_2$) und der Diskordanzen ($x(t_1) < x(t_2)$ für $t_1 > t_2$)

Erweiterung: Seasonal Kendall Test (Trendanalyse getrennt z.B. für einzelne Jahreszeiten)

Beziehungen zu anderen Verfahren: Trendfreiheit ist Voraussetzung für viele andere Verfahren

Der Mann-Kendall Test

Anwendung des Kendall-Tests auf Zeitreihen (d.h., sortiert nach Zeit, ohne doppelte Einträge) => Trendtest:

$$S = \sum_{k=1}^{n-1} \sum_{j=k+1}^{n} \mathrm{sgn}(x(t_j) - x(t_k)) \qquad \mathrm{sgn}(y) = \begin{cases} +1 & y > 0 \\ 0 & y = 0 \\ -1 & y < 0 \end{cases}$$

Für die H_0-Hypothese (= "es gibt keinen Trend") gilt dann:

$$E(S) = \quad \bar{S} = 0$$
$$\mathrm{var}(S) = \quad \sigma^2(S) = \frac{n(n-1)\cdot(2n+5)}{18}$$

=> normalverteilt

=> Ableitung der Testgröße: Abweichung der beobachteten (normierten) S von den laut $\mathbf{H_0}$ erwarteten

Der Mann-Kendall Test

beachte: Korrektur für verbundene Ränge (Ranggleichheit) notwendig

=> statt

$$\mathrm{var}(S) = \sigma^2(S) = \frac{n(n-1)\cdot(2n+5)}{18}$$

$$\mathrm{var}(S) = \sigma^2(S) = \frac{n(n-1)\cdot(2n+5) - \boxed{\sum_{j=1}^{p} R_j(R_j-1)\cdot(2R_j+5)}}{18}$$

R_j = Anzahl der Werte jeweils gleicher Ränge,

p = Anzahl der Gruppen verbundener Ränge

=> Teststatistik: $\tau = \frac{S}{D}$ (*für große n annähernd normalverteilt*),

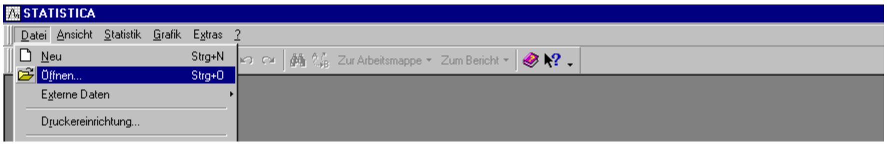
wobei D = maximal mögliche Anzahl der Konkordanzen:

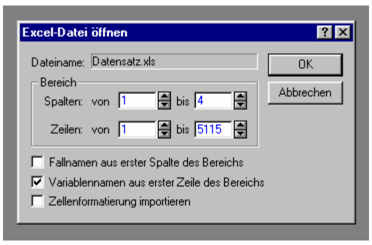
$$D = \sqrt{\frac{1}{2}(n(n-1) - \frac{1}{2}\sum_{j=1}^{p} R_j(R_j-1)} \cdot \sqrt{\frac{1}{2}n(n-1)}$$

Aufgabe

1. Lesen Sie die Daten des Files „Datensatz.xls“ in Statistica ein.
2. Führen Sie eine Autokorrelationsanalyse für die drei Spalten des Datenfiles durch.
3. Bestimmen Sie sämtliche Kreuzkorrelationsfunktionen.
4. Bestimmen Sie die Trends der Mittelwerte jeweils durch Regression der Messwerte mit der Zeitachse.
5. Führen Sie für jeden Parameter eine Trendanalyse mittels des Kendall-Tests durch.

Statistica: Einlesen der Daten





Statistica:Autokorrelation