Zeitreihenanalyse mittels Singulärspektrumanalyse

Hans-Jörg Starkloff

TU Bergakademie Freiberg (Sachsen)
Institut für Stochastik

Workshop „Mathematik in Forschung und Lehre“
Waschleithe 12.-14.09.2017

Einleitung

- In Praxis: Daten oft in Form einer **Zeitreihe**.
- Analyse z.B. um
  - Gesetzmäßigkeiten zu erkennen;
  - Modelle anzupassen;
  - Vorhersagen zu treffen;
  - besondere Situationen oder Zustände zu erkennen;
  - Prozesse zu steuern.
- Bei Nutzung stochastischer bzw. statistischer Modelle: **statistische Zeitreihenanalyse**
  - Besonderheit für Statistik: in der Regel nur eine Realisierung und abhängige Variablen;
  - enge Verbindung zur Theorie und Statistik der stochastischen Prozesse (Zufallsfunktionen).
- Wichtiges Forschungs- und Lehrgebiet in der Stochastik und deren Anwendungen.

Einige grundlegende Begriffe

- **(Konkrete) Zeitreihe**: ist eine Sammlung von Daten, die in einer zeitlichen Folge beobachtet wurden bzw. die bestimmten geordneten Zeitpunkten zugeordnet werden.
- Hier: **diskrete Zeitreihen**; Werte liegen für diskrete Zeitmomente vor, z.B. $x_{t_1}, x_{t_2}, \ldots, x_{t_n}$.
- Außerdem: **kontinuierliche Zeitreihen**; Beobachtungen werden kontinuierlich in der Zeit gemacht.
- **Univariate Zeitreihe**: für jeden betrachteten Zeitpunkt liegt ein reeller Zahlenwert vor.
- **Multivariate** oder **vektorielle Zeitreihe**: für jeden betrachteten Zeitpunkt wird eine endliche Anzahl von Merkmalen beobachtet (gemessen).
- Auch noch allgemeinere Situationen werden untersucht.

Klassischen ökonomische Zeitreihenanalyse

- In der klassischen Zeitreihenanalyse für ökonomische Zeitreihen zerlegt man eine Zeitreihe $(x_t; t = 1, \ldots, T)$ in verschiedene Komponenten:
  - die **glatte Komponente** $(g_t; t = 1, \ldots, T)$,
    darunter versteht man den Trend, d.h. die Grundrichtung der Zeitreihe, ggf. auch unter Einschluss langfristiger Schwingungen;
  - die **Konjunkturkomponente** $(k_t; t = 1, \ldots, T)$,
    welche kurz- und mittelfristige Konjunkturschwankungen beschreibt;
  - die **Saisonkomponente** $(s_t; t = 1, \ldots, T)$,
    welche Saisonschwankungen beschreibt und
  - die **irreguläre Komponente** $(r_t; t = 1, \ldots, T)$,
    welche irreguläre Schwankungen beschreibt, die durch nicht genauer bestimmbare Störfaktoren hervorgerufen werden; diese Komponente wird auch als **unsystematische Komponente** bezeichnet, wenn die anderen Komponenten systematische Einflüsse modellieren sollen. Bei einer traditionellen stochastischen Modellierung wird die irreguläre Komponente als Realisierung eines stochastischen („Rausch-“)Prozesses angesehen.

Komponentenmodelle

- Bei einem **additiven Komponentenmodell** überlagern sich die Komponenten additiv,

$$x_t = g_t + k_t + s_t + r_t, \qquad t = 1, \ldots, T\,.$$

- Bei einem **multiplikativen Komponentenmodell** werden die Komponenten multipliziert,

$$x_t = g_t \cdot k_t \cdot s_t \cdot r_t, \qquad t = 1, \ldots, T\,,$$

  damit verstärken die Einflüsse einander. Durch Logarithmieren (Voraussetzung: alle Komponenten sind positiv) erhält man wieder eine additive Verknüpfung, was für Berechnungen vorteilhaft sein kann.
- Daneben werden mitunter auch gemischte Verknüpfungsmodelle verwendet, sie erfordern oft spezielle Berechnungsverfahren.

Stochastische Modellierung

- Bei einer stochastischen Modellierung werden die Zahlenwerte einer konkreten (univariaten diskreten) Zeitreihe $x_{t_1}, \ldots, x_{t_n}$ als Realisierungen von Zufallsgrößen $X_{t_1}, \ldots, X_{t_n}$ angesehen. Diese Zufallsgrößen werden im Allgemeinen mit Hilfe eines **stochastischen Prozesses** $(X_t\,;\, t \in \mathbb{T})$ mit einer Zeit-, Parameter- oder Indexmenge $\mathbb{T} \subseteq \mathbb{R}$ modelliert.
- Hier: $\mathbb{T} \subseteq \mathbb{Z}$, die relevanten Zeitmomente sind durchnummeriert und es werden die Nummern genutzt.
- Ein **(zeitdiskreter reeller) stochastischer Prozess** (eine **(reelle) Zufallsfolge** oder eine **(univariate) mathematische Zeitreihe**) ist eine mit Elementen t aus $\mathbb{T} \subseteq \mathbb{Z}$ indizierte Familie $(X_t\,;\, t \in \mathbb{T})$ von Zufallsgrößen, definiert auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbf{P})$ (d.h. $X_t : \Omega \ni \omega \mapsto X_t(\omega)\,, t \in \mathbb{T}$).
- Die Menge $\mathbb{T}$ heißt auch **Zeitmenge**, **Indexmenge** oder **Parametermenge**.

Stationäre Zeitreihenmodelle

- Bei einer stochastischen Modellierung spielen meistens schwach stationäre Zufallsfolgen eine grundlegende Rolle.
- Den stochastischen Prozess $(X_t\,;\, t \in \mathbb{T})$ nennt man **schwach stationärer stochastischer Prozess** (auch **stationär im weiteren Sinne**, ...), wenn für alle Zufallsgrößen $X_t, t \in \mathbb{T}$, Erwartungswerte und Varianzen existieren und für beliebige $s, t, s+h, t+h \in \mathbb{T}$ gilt

$$\begin{aligned} \mathbf{E}[X_t] &= \mathbf{E}[X_{t+h}]\ , \\ \mathbf{Cov}[X_s, X_t] &= \mathbf{Cov}[X_{s+h}, X_{t+h}] =: \gamma_X(s-t)\,. \end{aligned}$$

- Sehr oft wird der grundlegende Zufallseinfluss durch einen Prozess des Weißen Rauschens modelliert. In neueren Arbeiten werden z.B. aber auch fraktale Rauschprozesse zur Modellierung des grundlegenden Zufallseinflusses genutzt.

Weißes Rauschen

- Ein stochastischer Prozess $(X_t\,;\, t \in \mathbb{T} \subseteq \mathbb{Z})$ heißt **(zeitdiskretes) Weißes Rauschen** (**White noise**) falls für $s, t \in \mathbb{T}$ gelten:

$$\mathbf{E}[X_t] = 0\,, \quad \mathbf{Cov}[X_s, X_t] = \begin{cases} \sigma^2 > 0\,, & s = t\,, \\ 0, & s \neq t\,. \end{cases}$$

  Dies wird durch $(X_t) \sim \mathrm{WN}(0, \sigma^2)$ bezeichnet.

- Sind die Zufallsgrößen $X_t, t \in \mathbb{T}$, eines Weißen Rauschens zusätzlich unabhängig und identisch verteilt, dann wird dies durch $(X_t) \sim \mathrm{IID}(0, \sigma^2)$ bezeichnet.

- Sind die Zufallsgrößen $X_t, t \in \mathbb{T}$, eines Weißen Rauschens zusätzlich unabhängig und identisch normalverteilt, dann wird dies durch $(X_t) \sim \mathrm{IIN}(0, \sigma^2)$ bezeichnet („Gauss**sches Weißes Rauschen**“).

Transformationen eines Weißen Rauschens

- Mit Hilfe von Transformationen (Filtern) werden mit Hilfe eines Weißen Rauschens wesentliche Klassen von betrachteten Zufallsfolgen erzeugt.
- Dabei unterscheidet man lineare Transformationen (**lineare stationäre Zeitreihenmodelle**, z.B. ARMA-Modelle oder -Prozesse) und nichtlineare Transformationen (**nichtlineare Zeitreihenmodelle**, z.B. ARCH- oder GARCH-Modelle oder Prozesse).
- Bei den linearen Transformationen werden oft der **Verschiebungs-** oder **Lagoperator** L (auch B von "backshift") bzw. der **(Rückwärts-)Differenzenoperator** Δ (auch ∇) genutzt:

$$\mathrm{L}\,X_t := X_{t-1}\,, \quad \mathrm{L}^2\,X_t := \mathrm{L}\,(\mathrm{L}\,X_t) = X_{t-2}\,, \text{usw.}$$
$$\Delta\,X_t := X_t - X_{t-1} = (1-\mathrm{L})X_t = (\mathrm{I}-\mathrm{L})X_t\,,$$
$$\Delta_2\,X_t := X_t - X_{t-2} = (1-\mathrm{L}^2)X_t\,, \quad \text{usw.}$$

ARMA(p, q)-Prozesse

- Der stochastische Prozess $(X_t\,; t \in \mathbb{Z})$ heißt **autoregressiver Moving-Average-Prozess der Ordnung** (p, q) oder **ARMA(p, q)-Prozess** $(p, q \in \mathbb{N}_0)$, falls
  - er schwach stationär ist **und**
  - mit reellen Zahlen $\phi_p \neq 0, \phi_{p-1}, \ldots, \phi_1$, $\theta_0 \neq 0, \ldots, \theta_q \neq 0$ und einem Weißen Rauschen $(Z_t\,;\, t \in \mathbb{Z})$, d.h. $(Z_t) \sim \mathrm{WN}(0, \sigma^2)$ mit $\sigma^2 > 0$, für beliebige $t \in \mathbb{Z}$ gilt

$$X_t = \phi_1 X_{t-1} + \ldots + \phi_p X_{t-p} + \theta_0 Z_t + \theta_1 Z_{t-1} + \ldots + \theta_q Z_{t-q}\,.$$

- Der stochastische Prozess $(X_t\,; t \in \mathbb{Z})$ heißt **ARMA(p, q)-Prozess mit Mittelwert** $\mu \in \mathbb{R}$, falls $(X_t - \mu\,; t \in \mathbb{Z})$ ein ARMA(p, q)-Prozess ist.

Bemerkungen zu ARMA(p, q)-Prozessen

- Für praktische Anwendungen werden meist zusätzliche Eigenschaften, wie die Kausalität oder die Invertierbarkeit gefordert.
- Diese Eigenschaften können z.B. mit Hilfe der Polynome

$$\Phi(z) := 1 - \phi_1 z - \ldots - \phi_p z^p\,, \qquad z \in \mathbb{C}\,,$$
$$\Theta(z) := \theta_0 + \theta_1 z + \ldots + \theta_q z^q\,, \qquad z \in \mathbb{C}\,,$$

  untersucht werden.
- Die Definitionsbeziehung eines ARMA(p, q)-Prozesses kann als

$$\Phi(L)X_t = \Theta(L)Z_t$$

  geschrieben werden.
- Bei Anwendungen müssen die Ordnungen p und q sowie die Koeffizienten der Polynome bestimmt (geschätzt) werden.
- Daneben kann man die Spektraltheorie stationärer Zufallsfolgen zur theoretischen bzw. statistischen Untersuchung nutzen.

Nichtstationäre Zeitreihenmodelle

- Bei Anwendungen in den Wirtschaftswissenschaften wird oft zwischen differenzenstationären und trendstationären stochastischen Prozessen unterschieden.
- Für $d \in \mathbb{N}_0$ heißt der stochastische Prozess $(X_t\,;\, t \in \mathbb{Z})$ ein **ARIMA(p, d, q)-Prozess**, falls der stochastische Prozess $(Y_t\,;\, t \in \mathbb{Z})$ mit $Y_t = (1-L)^d X_t\,, t \in \mathbb{Z}\,,$ ein kausaler ARMA(p, q)-Prozess ist. Damit genügt ein ARIMA(p, d, q)-Prozess der Differenzengleichung

$$\Phi(L)(1-L)^d X_t = \Phi(L) Y_t = \Theta(L) Z_t\,, \quad (Z_t) \sim \mathrm{WN}(0; \sigma^2)\,.$$

- Der stochastische Prozess $(X_t\,;\, t \in \mathbb{Z})$ heißt trendstationär, falls für eine deterministische Funktion $\mu : \mathbb{Z} \to \mathbb{R}$ der stochastische Prozess $(Y_t\,;\, t \in \mathbb{Z})$ mit

$$Y_t := X_t - \mu(t)\,, \qquad t \in \mathbb{Z}\,,$$

ein schwach stationärer stochastischer Prozess ist.

Transformationen zur Erreichung der Stationarität

- Transformationen zur Erreichung der Stationarität können z.B. in der Bereinigung von
  - deterministischen Trends (beschrieben z.B. durch ein Polynom in der Zeitvariable, welches durch die Methode der kleinsten Quadrate geschätzt werden kann) oder
  - deterministischen saisonalen Komponenten

  bestehen.

- Die Untersuchung der durch Differenzenbildung erhaltenen stationären Zeitreihen führt z.B. auf ARIMA-Prozesse.

- Weitere oft verwendete Transformationen sind das Logarithmieren der Werte oder die Box-Cox-Transformation $Y_t = \frac{(X_t + c)^\lambda - 1}{\lambda}$ für ein $\lambda \neq 0$ (die Konstante c dient dazu, dass alle Werte größer als null sind).

Die Singulärspektrumanalyse

- ist ein Verfahren der Zeitreihenanalyse und Vorhersage;
- kombiniert Elemente der klassischen Zeitreihenanalyse, multivariaten Statistik, multivariaten Geometrie, der Theorie dynamischer Systeme und der Signalverarbeitung;
- setzt kein parametrisches Modell und keine Stationaritätsannahmen voraus, ist demzufolge flexibel anwendbar und konnte schon für viele Anwendungen erfolgreich eingesetzt werden.
- Engl. "Singular spectrum analysis" (SSA), "Caterpillar methodology" (russ. "Gusenitsa").
- Erste Artikel 1986, z.B. Broomhead, King, Extracting qualitative dynamics from experimental data, Physica D.
- Enge Verbindung z.B. zur Hauptkomponentenanalyse und zur Karhunen-Loève-Entwicklung.

Zielstellung der Singulärspektrumanalyse

Die Singulärspektrumanalyse zielt auf die (additive) Zerlegung einer gegebenen Zeitreihe in eine kleine Anzahl interpretierbarer Komponenten, wie einem sich langsam änderndem Trend, einer oszillierenden Komponente und einem „strukturlosen“ Rauschen.

Algorithmus

- **Geg.** konkrete Zeitreihe $\mathbb{X} = \mathbb{X}_N = (x_1, \ldots, x_N)$, $N > 2$, $N \in \mathbb{N}$, $\exists\, i \in \{1, \ldots, N\} : x_i \neq 0$.
- Wahl des Parameters $L \in \mathbb{N}$ $(1 < L < N)$ (**Fensterlänge**); $K := N - L + 1$.
- **1. Stufe** Zerlegung
  - **1. Schritt** Einbettung (Aufstellen der **Trajektorienmatrix** $\mathbf{X}$)
  - **2. Schritt** Singulärwertzerlegung der Trajektorienmatrix
- **2. Stufe** Rekonstruktion
  - **3. Schritt** Gruppierung der Eigentripel
  - **4. Schritt** Diagonale Mittelung
- **Ergebnis** Zerlegung der Ausgangszeitreihe in die Summe von m $(m \in \mathbb{N}, m \leq L)$ rekonstruierten Zeitreihen,

$$x_n = \sum_{k=1}^{m} \widetilde{x}_n^{(k)} \qquad (n = 1, \ldots, N).$$

1. Schritt: Einbettung

- Folge von $K = N - L + 1$ verschobenen Vektoren der Länge L

$$X_i = (x_i, \ldots, x_{i+L-1})^{\mathrm{T}} \qquad (i = 1, \ldots, K)\,.$$

- **L-Trajektorienmatrix** (**Trajektorienmatrix**)

$$\mathbf{X} = [X_1 : \ldots : X_K] = (x_{ij})_{i,j=1}^{L,K} = \begin{pmatrix} x_1 & x_2 & x_3 & \ldots & x_K \\ x_2 & x_3 & x_4 & \ldots & x_{K+1} \\ x_3 & x_4 & x_5 & \ldots & x_{K+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_L & x_{L+1} & x_{L+2} & \ldots & x_N \end{pmatrix}.$$

- $L \times K$-Matrix; (i,j)-tes Element ist $x_{ij} = x_{i+j-1}$ $\Rightarrow$ übereinstimmende Elemente auf den Gegendiagonalen $i + j = \mathrm{const}$ (HANKEL-Matrix).

Bemerkungen zur Einbettung

- Die univariate Zeitreihe $\mathbb{X} = (x_1, \ldots, x_N)$ wird in eine multivariate Zeitreihe $X_1, \ldots, X_K$ mit Vektoren $X_i = (x_i, \ldots, x_{i+L-1}) \in \mathbb{R}^L$ umgewandelt.
- Ist $\mathbf{X}$ die L-Trajektorienmatrix einer univariaten Zeitreihe $\mathbb{X}$, dann ist $\mathbf{X}^{\mathrm{T}}$ die K-Trajektorienmatrix derselben Zeitreihe.
- Für $N, L, K \in \mathbb{N}$ mit $K = N - L + 1$ ist eine $L \times K$-Matrix genau dann die L-Trajektorienmatrix einer univariaten Zeitreihe, wenn sie eine HANKEL-Matrix ist.
- Die Fensterlänge L sollte hinreichend groß sein, damit die verschobenen Vektoren einen wesentlichen Teil des Verhaltens der Ausgangszeitreihe repräsentieren können.
- Man kann immer $L \leq N/2$ wählen; günstige Werte für L hängen von den Daten und dem Ziel der Untersuchung ab, ggf. sollte man die Analyse für verschiedene Werte von L durchführen und die Ergebnisse miteinander vergleichen.

Zerlegungen von Zeitreihen und Trajektorienmatrizen

- Sind univariate Zeitreihen

$$\mathbb{X}^{(1)} = (x_1^{(1)}, \ldots, x_N^{(1)}) \qquad \text{und} \qquad \mathbb{X}^{(2)} = (x_1^{(2)}, \ldots, x_N^{(2)})$$

mit zugehörigen L-Trajektorienmatrizen $\mathbf{X}^{(1)}$ und $\mathbf{X}^{(2)}$ gegeben, dann gilt für L-Trajektorienmatrix $\mathbf{X}$ der Summenzeitreihe

$$\mathbb{X} = \mathbb{X}^{(1)} + \mathbb{X}^{(2)} = (x_1^{(1)} + x_1^{(2)}, \ldots, x_N^{(1)} + x_N^{(2)})$$

die Beziehung $\mathbf{X} = \mathbf{X}^{(1)} + \mathbf{X}^{(2)}$, ebenso für mehr Summanden und umgekehrt.

2. Schritt: Singulärwertzerlegung der Trajektorienmatrix $\mathbf{X}$

- $\mathbf{S} = \mathbf{X}\mathbf{X}^{\mathrm{T}}$ symmetrische, positiv semidefinite $L \times L$-Matrix.
- $\lambda_1 \geq \ldots \geq \lambda_L \geq 0$ Eigenwerte von $\mathbf{S}$, monoton fallend.
- $U_1, \ldots, U_L$ orthonomiertes System der Eigenvektoren der Matrix $\mathbf{S}$ zu diesen Eigenwerten.
- $d := \operatorname{rank} \mathbf{X} = \max\{i \in \mathbb{N} \text{ mit } \lambda_i > 0\}$ (in der Praxis üblicherweise $d = L^* := \min\{L, K\}$).
- $V_i = \mathbf{X}^{\mathrm{T}} U_i / \sqrt{\lambda_i} \quad (i = 1, \ldots, d)$.
- $\Rightarrow$ mit den Matrizen vom Rang 1 $\mathbf{X}_i = \sqrt{\lambda_i} U_i V_i^{\mathrm{T}}$ (**elementare Matrizen**) gilt

$$\mathbf{X} = \mathbf{X}_1 + \ldots + \mathbf{X}_d$$

- $(\sqrt{\lambda_i}, U_i, V_i)$ ***i*-tes Eigentripel der Singulärwertzerlegung**, $\sqrt{\lambda_i}$ ***i*-ter Singulärwert**, U_i ***i*-ter linker Singulärvektor**, V_i ***i*-ter rechter Singulärvektor**, ***i*-ter Faktorvektor**.

Die Matrix $\mathbf{S} = \mathbf{X}\mathbf{X}^{\mathrm{T}}$

- $\mathbf{S} = (s_{ij})_{i,j=1}^{L}$ mit

$$s_{ij} = \sum_{t=0}^{K-1} x_{i+t} x_{j+t} \,,$$

so dass für Differenz der Indizes gilt $j + t - i - t = j - i$.

- Damit wäre für einen zugrundeliegenden zentrierten schwach stationären Prozess $\frac{1}{K} s_{ij}$ ein Schätzwert für den Wert der Kovarianzfunktion $\gamma(j - i)$.

- Eine andere Variante der Singulärspektrumanalyse nutzt statt der Matrix $\mathbf{S}$ die Matrix der Schätzwerte der Kovarianzfunktion mit Einträgen

$$c_{ij} = \frac{1}{N-k} \sum_{t=1}^{N-k} x_t x_{t+k} \qquad \text{mit } i, j = 1, \ldots, L \text{ und } k = |i - j| \,.$$

3. Schritt: Gruppierung der Eigentripel

- Zerlegung der Indexmenge $\{1, \ldots, d\}$ in m disjunkte Teilmengen $I_1, \ldots, I_m$.
- Für $k \in \{1, \ldots m\}$ sei $\mathbf{X}_{I_k} = \sum_{\ell \in I_k} \mathbf{X}_\ell$.
- Die Zerlegung der Indexmenge führt so zur **gruppierten Zerlegung**

$$\mathbf{X} = \mathbf{X}_{I_1} + \ldots + \mathbf{X}_{I_m}$$

der Trajektorienmatrix.

- Im Fall $m = d$ und $I_j = \{j\}, j = 1, \ldots, d$, wird die entsprechende Gruppierung **elementare Gruppierung** genannt.

4. Schritt: Diagonale Mittelung

- Für jeder der Matrizen $\mathbf{X}_{I_j}$ der gruppierten Zerlegung werden die Elemente auf den Gegendiagonalen durch ihren Mittelwert ersetzt, dadurch entstehen HANKEL-Matrizen, denen entsprechende Zeitreihen zugeordnet werden können.
- Man spricht auch von einer HANKELisierung, diese besitzt auch bestimmte Optimalitätseigenschaften.
- **Ergebnis** Zerlegung der Ausgangszeitreihe in die Summe von m $(m \in \mathbb{N}, m \leq L)$ rekonstruierten Zeitreihen,

$$x_n = \sum_{j=1}^{m} \widetilde{x}_n^{(j)} \qquad (n = 1, \ldots, N)\,.$$

Ausgewählte Literatur

- Golyandina,N., Zhigljavsky,A.; Singular Spectrum Analysis for Time series; Springer, 2013.
- Neusser, K.; Zeitreihenanalyse in den Wirtschaftswissenschaften; Vieweg Teubner, 2011 (3. Auflage).
- Brockwell, P.J., Davis, R.A.; Time Series: Theory and Methods; Springer, 2006 (2. Auflage).
- Schlittgen, R.; Angewandte Zeitreihenanalyse mit R; De Gruyter Oldenbourg, 2015 (3. Auflage).
- Golyandina, N., Korobeynikov, A.; Basic Singular Spectrum Analysis and forecasting with R; Computational Statistics and Data Analysis 71 (2014) 934-954.