

Skript zur Vorlesung :

**Mathematische Methoden zur Analyse von
Zeitreihen komplexer Systeme**

Vorlesung WS 06/07

Prof. Dr. Jens Timmer

Inhaltsverzeichnis

1	Einleitung	7
2	Ein bißchen Statistik	11
2.1	Verteilungen	12
2.1.1	Zufallsvariablen	12
2.1.2	Momente	12
2.1.3	Beispiele für Verteilungen	13
2.1.4	Zentraler Grenzwertsatz	18
2.1.5	Schätzen von Parametern von Verteilungen	19
2.2	Hypothesen Tests	23
2.2.1	Parametrische Tests	23
2.2.2	Nichtparametrische Tests	31
3	Grundlegende Begriffe	33
3.1	Ergodizität	33
3.2	Stationarität	34
3.3	Mixing properties	35
4	Exemplarische Prozesse und Datensätze	39
4.1	Deterministische Systeme	39
4.1.1	Van der Pol Oscillator	39
4.1.2	Rössler System	41
4.1.3	Integration deterministischer Systeme	42
4.2	Stochastische Systeme	43
4.2.1	Weißes Rauschen	43
4.2.2	Autoregressive Prozesse	44
4.2.3	Stochastischer van der Pol Oscillator	50
4.3	Exemplarische Datensätze	52

4.3.1	Sunspots	52
4.3.2	Canadische Luchse	53
4.3.3	Southern Oscillation Index	53
5	Spektralanalyse	55
5.1	Autospektrum	55
5.1.1	Theorie im Unendlichen	55
5.1.2	Schätzung im Endlichen	74
5.1.3	Schätzung für periodisch korrelierte Prozesse	98
5.1.4	Nachweis von Fractional Browian Motion oder FGN	99
5.1.5	Zum Merken	99
5.2	Kreuzspektralanalyse	100
5.2.1	Analytisches	100
5.2.2	Schätzerei	107
5.2.3	Graphische Modelle	113
5.3	Bispektrum	116
5.3.1	Bispektrum	118
5.3.2	Kreuzbispektrum	119
5.3.3	Wavelets	119
6	Parameterschätzung in dynamischen Systemen	121
6.1	Parameterschätzung	121
6.1.1	Maximum Likelihood Schätzer	121
6.1.2	Methods of Moments	127
6.1.3	Ausflug Bayesianismus	127
6.2	ARMA Modelle	129
6.2.1	AR Modelle	129
6.2.2	MA Modelle	130
6.2.3	Innovation-Algorithmus	132
6.2.4	Maximum Likelihood Schätzer	133
6.2.5	ARMA[p,q] Modelle	135
6.2.6	Diskret vs. Kontinuierlich III	135
6.3	Nichtlineare Differenzgleichungen	135
6.3.1	Diskret vs. Kontinuierlich IV	137
6.3.2	Diskret vs. Kontinuierlich V	137
6.4	EM Algorithmus	138
6.5	Zustandsraummodell	143
6.5.1	Gründe fürs ZRM	143

6.5.2	Eichfixierung	145
6.5.3	Diskret vs. Kontinuierlich Part VI	146
6.5.4	Kalman und Smoothing Filter	147
6.5.5	Parameterschätzung im Zustandsraummodell	153
6.6	Markov Prozesse	161
6.6.1	Motivation, Defition, Eigenschaften	161
6.6.2	Parameterschätzung in (nicht aggregierten) Markov Modellen	165
6.6.3	Identifizierbarkeit	166
6.6.4	Likelihood Ratio Tests	175
6.7	Hidden Markov Modell	184
6.7.1	EM Algorithmus im HMM, Baum-Welsh-Algorithmus .	185
6.7.2	Viterbi Algorithmus	190
6.7.3	HMM Erweiterungen	192
6.8	Stochastische Differentialgleichungen	195
6.8.1	Diskret vs. kontinuierlich Teil VIII	199
6.8.2	Maximum likelihood Schätzung von Parametern in SDEs	199
6.8.3	Quasi Maximum Likelihood Schätzer	200
6.8.4	Diskret vs. Kontinuierlich Part IX	203
6.9	Deterministische Systeme	204
6.9.1	Parametrische Methoden	204
6.9.2	Nichtparametrische Methoden	221
7	Punktprozesse	229
7.1	Shot noise	233
8	Modellselektion	239
8.1	Akaike Information Criterion (AIC)	241
8.2	Bayes'sches Information Criterion (BIC)	242
8.3	Minimum Description Length (MDL)	243
8.4	Mallows' C_p	243
8.5	Final Prediction Error (FPE)	244
9	Prozesse der Finanzmathematik	247
10	Neuronale Netze	249
10.1	Ein bißchen Neurobiologie	249
10.2	Hopfield	250

10.3 Back-Propagation	250
10.4 Clustering	250
10.5 Kohonen	251
10.6 Reinforcement learning	251
11 Jobs, careers, the future and all that	253

Kapitel 1

Einleitung

Org-Krams:

- Welche Fakultät ?
- Übung ganz wichtig! Wer kann nicht hacken ? MATLAB
- Kommunikation über homepage
- Script auf homepage unter und math_meth.pdf. Ist eine Baustelle.
- Fragen bei Unklarheiten !
- Mittwoch wird flexibel gehandhabt
- Abteilungshomepage und Vorlesungshomepage
- Inhaltsverzeichnis
- "Wo bleibt der Zwang ?", Scheinkriterien

ZEICHNUNG: Die 3 Richtungen der Physik

Das allgemeine Problem :

Gegeben ein Modell, ist es höchstens technisch schwierig Daten zu generieren.

Dieses ist das Direkte Problem.

Gegeben Daten, ist es i.d.R. schwierig, das Modell zu finden.

Dieses ist das Inverse Problem.

Model \implies Daten
 aber
 Daten $\not\Rightarrow$ Modell

Daher: Hin und her zwischen Direktem und Inversem Problem.
 Hier Schwerpunkt auf dem Inversen Problem.
 Daten von dynamischen Systemen heißen Zeitreihen

FOLIE Beispiele von Zeitreihen

Was immer man tut, man hat irgendwelche Modellvorstellungen im Hinterkopf (aus Vorwissen oder Dogmatismus). Darum ist es sinnvoll, sich mögliche Modellklassen zu überlegen:

	deterministisch	stochastisch
linear	$\dot{\vec{x}} = \mathbf{A}\vec{x}$	$\dot{\vec{x}} = \mathbf{A}\vec{x} + \vec{\epsilon}$
nicht linear	$\dot{\vec{x}} = \vec{f}(\vec{x}, \vec{p})$	$\dot{\vec{x}} = \vec{f}(\vec{x}, \vec{p}, \vec{\epsilon})$

ϵ modelliert zahlreiche Einflüsse auf das System.

Weitere Prozeßklassen:

- Partielle Differentialgleichungen $\dot{x} = f(\nabla x, x, p)$
- Delay-Differentialgleichungen $\dot{x}(t) = f(x(t), x(t - \tau), p)$
- Markov-Prozesse mit diskretem Zustandsraum
 $p(x_i(t)) = \sum_j p(x_i(t)|x_j(t - 1))p(x_j(t - 1)), i, j = 1, 2, \dots, M$
- Punktprozesse

$$x(t) = \sum_j \Theta(t - T_j) g(t - T_j) \quad (1.1)$$

mit "Shot-Profil" $g(\cdot)$ z.B.:

$$g(t - T_j) = M e^{-(t - T_j)/\tau}$$

und T_j z.B. aus Poisson-Prozeß:

$$\lim_{\Delta t \rightarrow 0} \text{prob} (\text{Event in } (t, t + \Delta t)) = \rho \Delta t, \quad (1.2)$$

In der Regel x multivariat (mehrdimensional) und zeitkontinuierlich, aber die Beobachtung y univariat (eindimensional), zeitdiskret und mit Beobachtungsrauschen η versehen:

Somit:

$$\begin{aligned} \dot{x} &= f(x) && \text{Dynamische Gleichung oder Systemgleichung} \\ y(t_i) &= g(\vec{x}(t_i)) + \eta(t_i) && \text{Beobachtungsgleichung} \end{aligned}$$

Ziel: Lerne aus $y(t_i)$ etwas über $\dot{x} = f(x)$

Beispiele :

	deterministisch	stochastisch
linear	Sonne/Erde	gesunder Tremor [284]
nicht linear	chaotische Schaltkreise [247, 295]	kranke Tremores [287]

Beispiele der anderen Prozeßklassen:

- Chemische Prozesse auf Oberflächen [94, 245] (partielle DGL)
- Blutbildung [179] (Delay-DGL)
- Ionenkanäle [195] (Markov)
- Röntgenvariabilität Schwarzer Löcher [296], Herzschlag [152] (Punktprozesse)

Charakteristische Größen:

	deterministisch	stochastisch
linear	Perioden	Spektren
nicht linear	Dimensionen	??

Für die weiteren Prozeßklassen charakteristisch:

- Diffusionskonstante

- Delay time τ
- Verweilzeitenverteilung
- Poisson-Eigenschaft, Shot-Profile

Mögliche Ziele der Zeitreihenanalyse:

- Vorhersage des Zeitverlaufs, Börse
- Charakterisieren/Klassifizieren, z.B. bei medizinischen Daten zur Diagnoseunterstützung
- Modellieren $\dot{x} = f(x, p)$
 - Bestimme Parameter p ,
 - "Verstehen" der Dynamik
 - Hypothesen testen: Ist es $f_1(x, p)$ oder $f_2(x, p)$?

Zeitkontinuierliche vs. zeitdiskrete Dynamik

(Physikalisch) natürlich ist meistens zeitkontinuierliche Dynamik

$$\dot{\vec{x}} = \vec{f}(\vec{x}, \vec{\epsilon})$$

Ausnahme : Generationen von Viechern, z.B. Anzahl von Wölfen die *jährlich* gezählt werden.

Viele (mathematische) Modelle und Methoden sind zeitdiskret

$$x(i + 1) = g(x(i), \epsilon(i))$$

Wichtig: Die Zeitdiskretheit der Datenerhebung ist KEIN Argument für zeitdiskrete Modelle oder Methoden (wird running gag).

Kapitel 2

Ein bißchen Statistik

Literatur:

- Hartung, *Statistik* [108] Ein Klassiker, sehr detailliert
- Sachs, *Applied Statistics* [249] Kompendium, angewandt
- D.R. Cox, D.V. Hinkley *Theoretical Statistics* [51] lesbares Theoriebuch
- J. Honerkamp *Stochastic Dynamical Systems* [125] Kap. 1-3.
Kondensierte Darstellung der für Physiker relevanten Grundlagen der Statistik
- S. Kotz and N.L. Johnson *Breakthroughs in Statistics 1 & 2* [169]
Sammlung der 40 wichtigsten Statistikpaper mit Kommentaren, nett

Grundsätzlich zur Statistik :

- Kein Fall von Mathematik
- Ein paar Sachen muß man richtig verstehen.
- Vieles muß man kennen.
- Vieles muß man einfach nur nachschlagen können.
- Ein bisschen jetzt, der Rest im Laufe der Verhandlung

2.1 Verteilungen

2.1.1 Zufallsvariablen

Zufallsvariable X :

- Etwas, das eine Wahrscheinlichkeitsdichte $p_X(x)$ hat
- Wahrscheinlichkeit, eine Realisierung x in $(x, x + dx)$ zu beobachten, ist $p_X(x)dx$
- $p_X(x) \geq 0$, $\int p_X(x) dx = 1$

ZEICHNUNG dazu

- Physikalisch entsteht Zufall entweder
 - durch Quantenmechanik (eher selten in makroskopischen komplexen Systemen)
 - durch Chaos (auch selten)
 - oder durch viele Einflüsse á la Brownian Motion.
- Es gibt auch diskrete Verteilungen $p(x_i)$, denke an Würfel

Im folgenden, wenn Bezug klar: $p_X(x) = p(x)$

2.1.2 Momente

- $\langle (x)^k \rangle = \int x^k p(x) dx$

- 1. Moment Mittelwert

$$\mu_1 = \bar{x} = \mu = \langle x \rangle = \int xp(x) dx$$

- 2. Moment

$$\langle x^2 \rangle = \int x^2 p(x) dx$$

Varianz: $\sigma^2 = \langle (x - \bar{x})^2 \rangle = \mu_2 - \mu_1^2$.

Standardabweichung: σ .

Varianzen σ^2 , nicht Standardabweichungen σ sind additiv. Dies gilt bei unabhängigen Ereignissen.

- 3. Moment

$$\langle x^3 \rangle = \mu_3 = \int x^3 p(x) dx$$

Schiefe: $\kappa = \langle (x - \mu)^3 \rangle$, Maß für Asymmetrie.

- 4. Moment

Kurtosis (Bauchigkeit): $\gamma = \langle (x - \bar{x})^4 \rangle / \sigma^4 - 3$

2.1.3 Beispiele für Verteilungen

- Gaußverteilung oder Normalverteilung:

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\bar{x})^2}{2\sigma^2}}$$

Bezeichnung: $N(\mu, \sigma^2)$

Standard-Gaußverteilung $\bar{x} = 0$, $\sigma^2 = 1$: $N(0, 1)$

In $\pm 1\sigma$ liegt 68 % der Masse

In $\pm 1.96\sigma$ liegt 95 % der Masse

Momente von $N(0, 1)$:

$$\langle (x)^k \rangle = \begin{cases} 0 & \text{für k ungerade} \\ 1 \times 3 \dots (k-1) & \text{für k gerade} \end{cases}$$

Damit klar, warum bei Kurtosis "-3"

Zentraler Grenzwertsatz:

Summe von ZVs mit endlichen Momenten strebt gegen Gaußverteilung

Bedeutung nicht zu unterschätzen.

- Gleichverteilung $U(a, b)$:

$$p(x) = \begin{cases} 1/(b-a) & \text{für } a \leq x \leq b \\ 0 & \text{sonst} \end{cases}$$

Erzeugung erklären.

- Alle ZV-Erzeugung basiert gleichverteilten Zufallsvariablen
- Problem:
Wie auf einem (deterministischem) Computer "zufällige" Zahlen erzeugen.
- Diskussion: Erkennen von Zufall. Statistische Hypothese "5.6 ist zufällig" ist nicht abzulehnen.
ZEICHNUNG: Sinus Reihe
Zufall = Nicht Vorhersagbar, de facto Definition
- Lösung :
Chaotische dynamische Systeme zeigen Eigenschaften, die im Endlichen mitunter von echtem Zufall nicht zu unterscheiden sind.
- (Pseudo-)Zufallszahlen-generator: Poincaré - Schnitt durch ein hochdimensionales deterministisches chaotisches System.

$$x(t+1) = f(x(t))$$

ZEICHNUNG für ran0.f

- ähnliche Werte von $x(t)$ habe sehr verschiedene Werte $x(t+1)$.

- Exponential-Verteilung

$$p(x) = \frac{1}{\tau} e^{-x/\tau}$$

Es gilt:

$$\begin{aligned} \mu &= \tau \\ \sigma^2 &= \tau^2 \end{aligned}$$

Ergibt sich bei “konstanter Ausfallrate”

Erzeugung von exponentiell verteilten Zufallsvariablen per Transformationsmethode:

$$1 = \int_I dx p_X(x) = \int_{I'} dy \left| \frac{dx}{dy} \right| p_X(x(y)) = \int_{I'} dy p_Y(y)$$

- Sei x gleichverteilt
- Bilde $y(x)$
- Für Verteilung von y gilt

$$p_Y(y) = p_X(x) \left| \frac{dx}{dy} \right| = \left| \frac{dx}{dy} \right|$$

- Wähle $y(x) = -\log x$
- Gibt:

$$\tilde{p}(y) = \left| \frac{dx}{dy} \right| = e^{-y}$$

ZEICHNUNG dazu

Bemerkung Box-Müller

- χ_r^2 Verteilung mit r Freiheitsgraden:

$$”\chi_r^2 = \sum_{i=1}^r (N(0, 1))^2”$$

$$Y \sim \chi_r^2, \quad X_i \sim N(0, 1) = p_G(x_i)$$

$$p(y) = \int dx_1 \dots dx_r \delta(y - (x_1^2 + \dots + x_r^2)) \prod_{i=1}^r p_G(x_i)$$

$$\chi_r^2(x) = \frac{x^{r/2-1} e^{-x/2}}{2^{r/2} \Gamma(r/2)}$$

Es gilt:

$$\begin{aligned}\langle \chi_r^2 \rangle &= r \\ \text{Var}(\chi_r^2) &= 2r,\end{aligned}$$

d.h. Erwartungswert und Varianz sind keine unabhängigen Parameter.

BILD χ_r^2 Verteilungen, Apfel-Bemerkung

χ^2 Verteilungen sind additiv:

$$”\chi_{r_1}^2 + \chi_{r_2}^2 = \chi_{r_1+r_2}^2”$$

Aus dem zentralen Grenzwertsatz (CLT) folgt:

$$\lim_{r \rightarrow \infty} \chi_r^2 = N(r, 2r)$$

Bemerkungen :

- $\chi_2^2 = \frac{1}{2}e^{-x/2}$ ist Exponential-Verteilung mit $\tau = 2$.
- χ_2^2 wird wichtig in Kapitel 5 Spektralanalyse.

- t-Verteilung

$$t(r, x) = \frac{N(0, 1)}{\sqrt{\chi_r^2/r}} = \frac{1}{\sqrt{r}} \frac{1}{B(1/2, r/2)} \left(1 + \frac{x^2}{r}\right)^{-\frac{1}{2}(r+1)}, \quad B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$

Bemerkung: Siehe t-Test: Test auf Gleichheit von Mittelwerten zweier Gaußverteilungen.

$$\lim_{r \rightarrow \infty} t(r, x) = N(0, 1)$$

- F-Verteilung

$$F(r_1, r_2, x) = \frac{\chi_{r_1}^2 / r_1}{\chi_{r_2}^2 / r_2} = \dots$$

Bemerkung: Siehe F-Test: Test auf Gleichheit von Varianzen zweier Gaußverteilungen.

- Cauchy(-Lorenz) Verteilung:

$$p_{Cauchy}(x, \mu, \gamma) = \frac{1}{\pi} \frac{\gamma^2}{(x - \mu)^2 + \gamma^2}$$

$$\mu_k = \int x^k p(x) = \infty$$

Momente existieren nicht!

μ ist Lokalisationsparameter, aber kein Mittelwert.

Cauchy-Verteilung spielt bei den Zuwächsen der Aktienkurse eine Rolle.

Erzeugung per Transformationsmethode:

$$p_{Cauchy}(x, 0, 1) = \tan(\pi(U[-1/2, 1/2]))$$

Es gilt:

$$"Cauchy(x, 0, 1) = \frac{N(0, 1)}{N(0, 1)}"$$

das ist auch einfachste Erzeugungsweise.

Bezug zur t-Verteilung:

$$t(1, x) = p_{Cauchy}(x, 0, 1)$$

Mit t -Verteilung kann man also zwischen Cauchy (keine Momente) und Gauß (alle Momente) hin- und herfahren.

Bemerkung: Aus der Physik bekannt als Breit-Wigner-Verteilung.

- Binomial Verteilung

$$B(n, p, k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \dots, n$$

Zwei mögliche Ereignisse: x_1, x_2 ; $p = \text{prob}(x_1)$

Bei n -maliger Durchführung des Experimentes ist $B(n, p, k)$ die Wahrscheinlichkeit, x_1 k -mal zu realisieren.

- Poisson-Verteilung

$$P(k, \lambda) = \frac{e^{-\lambda} \lambda^k}{k!}$$

Bemerkung: Diese Verteilung wird für Poisson-Punktprozesse wichtig. Ferner ist sie die Grenzverteilung der Binomialverteilung:

$$\lim_{n \rightarrow \infty} B(n, k, p) = P(k, \lambda) \text{ wobei } \lim_{n \rightarrow \infty} np = \lambda$$

2.1.4 Zentraler Grenzwertsatz

Existieren die ersten beiden Momente, so strebt die Summe von unabhängigen, identisch verteilten (iid) Zufallsvariablen gegen eine Gaußverteilung.

Darum sind die beiden Größen μ und σ so wichtig.

- Konvergenzrate hängt i.w. von der Schiefe ab.
- Gilt nicht für Cauchy-Verteilung.
- Aber auch für Verteilungen mit nicht-existenten Momenten gibt es Grenzwertsätze.

2.1.5 Schätzen von Parametern von Verteilungen

- Mittelwert der Gaußverteilung $N(\mu, \sigma^2)$:
Gegeben N Realisierungen x_i aus $N(\mu, \sigma^2)$.
Der Schätzer

$$\hat{\mu} = \frac{1}{N} \sum_{i=1}^N x_i$$

ist eine gaußverteilte Zufallsvariable mit

$$\begin{aligned} \langle \hat{\mu} \rangle &= \mu \\ \langle \text{Var}(\hat{\mu}) \rangle &= \frac{1}{N} \sigma^2 \end{aligned}$$

Mit zunehmender Datenanzahl kann man den Mittelwert immer genauer wissen.

Vertrauensintervall: Mit 95%iger Wahrscheinlichkeit liegt der wahre Wert im Intervall

$$\left[\hat{\mu} - 1.96 \sqrt{\frac{1}{N} \sigma^2}, \hat{\mu} + 1.96 \sqrt{\frac{1}{N} \sigma^2} \right]$$

Anschaulich begründen.

- Der Schätzer für die Varianz

$$\hat{\sigma}^2 = \frac{1}{N-1} \sum_i (x_i - \hat{\mu})^2$$

ist χ_{N-1}^2 verteilt.

Bemerkung: Die Anzahl der Freiheitsgrade ist nur $N - 1$, da μ geschätzt werden musste (1 Freiheitsgrad).

Sei jeweils $X \sim N(\mu, \sigma^2)$

– Schätzer für den Mittelwert μ

$$\hat{\mu} = \frac{1}{N} \sum_{i=1}^N x_i$$

$$\langle \hat{\mu} \rangle = \frac{1}{N} \sum_{i=1}^N \langle X_i \rangle = \langle X \rangle = \mu$$

Schätzer ist unverzerrt oder unbiased. Liegt im Mittel richtig.

Varianz des Schätzers

$$\text{Var}(\hat{\mu}) = \frac{1}{N^2} \sum_{i=1}^N \text{Var}(X_i) = \frac{1}{N} \text{Var}(X)$$

Zusammengefaßt: $\hat{\mu}$ ist Gauß'sche Zufallsvariable mit

$$* \langle \hat{\mu} \rangle = \mu$$

$$* \text{Var}(\hat{\mu}) = \frac{1}{N} \text{Var}(X)$$

$$* \sigma(\hat{\mu}) = \sqrt{\frac{1}{N}} \sigma(X)$$

Damit folgt: $\pm\sigma$ (=68%) Vertrauensintervall für (wahres) μ :

$$[\hat{\mu} - \sigma(\hat{\mu}), \hat{\mu} + \sigma(\hat{\mu})]$$

oder

$$\left[\hat{\mu} - \sqrt{\frac{1}{N}} \sigma(X), \hat{\mu} + \sqrt{\frac{1}{N}} \sigma(X) \right]$$

Vertrauensintervall wird mit $\sqrt{\frac{1}{N}}$ kleiner
Schätzer ist konsistent.

Konsistent: Für $N \rightarrow \infty$ wird alles gut

– Drei Schätzer S_k^2 , $k = 1, 2, 3$ für die Varianz

* Sei Mittelwert unbekannt

First try:

$$S_1^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \hat{\mu})^2$$

Betrachte einen Summanden, addiere geschickt 0

$$\begin{aligned} \langle (X_i - \hat{\mu})^2 \rangle &= \langle ((X_i - \langle X \rangle) - (\hat{\mu} - \langle X \rangle))^2 \rangle \\ &= \text{Var}(X) - 2 \langle (X_i - \langle X \rangle)(\hat{\mu} - \langle X \rangle) \rangle \\ &\quad + \text{Var}(\hat{\mu}) \end{aligned}$$

Von oben: $\text{Var}(\hat{\mu}) = \frac{1}{N} \text{Var}(X)$

und

$$\begin{aligned} \langle (X_i - \langle X \rangle)(\hat{\mu} - \langle X \rangle) \rangle &= \frac{1}{N} \sum_{j=1}^N \langle (X_i - \langle X \rangle)(X_j - \langle X \rangle) \rangle \\ &= \frac{1}{N} \langle (X_i - \langle X \rangle)^2 \rangle \\ &= \frac{1}{N} \text{Var}(X) \end{aligned}$$

Alles zusammen

$$\begin{aligned} \langle (X_i - \hat{\mu})^2 \rangle &= \text{Var}(X) - 2 \frac{1}{N} \text{Var}(X) + \frac{1}{N} \text{Var}(X) \\ &= (1 - 1/N) \text{Var}(X) \\ &= \frac{N-1}{N} \text{Var}(X) \end{aligned}$$

Somit

$$\langle S_1^2 \rangle = \frac{1}{N} \frac{N-1}{N} \sum_{i=1}^N \text{Var}(X) = \frac{N-1}{N} \text{Var}(X) = \text{Var}(X) - \frac{1}{N} \text{Var}(X)$$

Ergo: Schätzer S_1^2 hat einen

$$\text{bias}(s_1^2) = \frac{1}{N} \text{Var}(X)$$

* Second try

$$S_2^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \hat{\mu})^2$$

Selbe Rechnung wie oben

$$\langle S_2^2 \rangle = \text{Var}(X)$$

Begründung:

- Die Berechnung des Mittelwertes kostet einen Freiheitsgrad.
- x_1, \dots, x_N unterliegen Nebenbedingung:

$$\sum_{i=1}^N x_i - \hat{\mu} = 0$$

* Sei Mittelwert μ bekannt

$$S_3^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2$$

Selbe Rechnung wie oben

$$\langle S_3^2 \rangle = \text{Var}(X)$$

- Vertrauensintervall für p der Binomial Verteilung

Mit $m = \#x_1$, ist der Schätzer

$$\hat{p} = \frac{m}{n}$$

asymptotisch ($np(1-p) > 10$) (Gleichung erläutern, Schiefe muß klein werden, am Rand langsamer) gaußverteilt (CLT) :

$$\hat{p} \sim N\left(p, \frac{1}{n} p(1-p)\right)$$

95 % Konfidenzintervall:

$$\left[\frac{m}{n} - 1.96 \sqrt{\frac{1}{n} p(1-p)}, \frac{m}{n} + 1.96 \sqrt{\frac{1}{n} p(1-p)} \right]$$

Für ($np(1-p) < 10$): schlage Pearson-Clopper-Werte nach.

2.2 Hypothesen Tests

2.2.1 Parametrische Tests

Häufig laufen Fragestellungen auf statistische Tests hinaus.
Alles weitere am Beispiel des t -Tests.

Das Procedere

- Formuliere Nullhypothese H_0 :

Hier:

Die Mittelwerte μ_1, μ_2 zweier Gauß-Verteilungen (mit gleicher Varianz σ^2) sind gleich.

- Berechne (analytisch/simulativ) Verteilung einer Testgröße unter der Nullhypothese.

Hier:

- Berechne empirische Mittelwerte $\hat{\mu}_1, \hat{\mu}_2$ für jeweils N_1 Messwerte x_i^1 und N_2 Messwerte x_i^2 :

$$\hat{\mu}_1 = \frac{1}{N_1} \sum_{i=1}^N x_i^1, \quad \hat{\mu}_2 = \frac{1}{N_2} \sum_{i=1}^N x_i^2$$

entsprechend empirische Varianzen $\hat{\sigma}_1^2, \hat{\sigma}_2^2$, da σ^2 in der Regel nicht bekannt ist:

$$\hat{\sigma}_k^2 = \frac{1}{N_k - 1} \sum_{i=1}^{N_k} (x_i^k - \hat{\mu}_k)^2$$

- Berechne das gewichtete Mittel:

$$\hat{S}^2 = \frac{1}{N_1 + N_2 - 2} [(N_1 - 1) \hat{\sigma}_1^2 + (N_2 - 1) \hat{\sigma}_2^2]$$

und den Standardfehler:

$$S = \hat{S} \left(\frac{1}{N_1} + \frac{1}{N_2} \right)$$

- Dann ist

$$t = (\mu_1 - \mu_2)/S$$

(bei Gültigkeit von H_0) t -verteilt mit $N_1 + N_2 - 2$ Freiheitsgraden.

- Erinnere:

$$\text{CLT: } \lim_{N \rightarrow \infty} t(r, x) = N(0, 1)$$

- Anschaulich $N_1 = N_2 = N$

$$(\hat{\mu}_1 - \hat{\mu}_2)/\hat{S} \sim N(0, 1/(2N))$$

- Verteilung von t unter Alternative $N(\mu_1 - \mu_2, 1/(2N))$

- **Dilemma I** des Testens:

- Test läuft auf Frage hinaus:

Gehört ein Punkt zu der Verteilung ?

Problem:

Das ist nicht zu verneinen.

Im Prinzip kann unter H_0 jeder Wert der Test-Statistik (beim t -Test: t) vorkommen.

p-Wert Wahrscheinlichkeit für Wert größer t :

$$p = 1 - \int_{-\infty}^t p(x) dx$$

Unter H_0 : p -Wert der Teststatistik ist gleichverteilt auf $[0,1]$.

Bemerkung: Aus jedem Test kommt eine bestimmte Testgröße als Ergebnis heraus. Beim t -Test ist dies wiederum t . Diese Testgröße kann jeden beliebigen Wert einer bestimmten Wertemenge (z.B. die reellen Zahlen, die positiven reellen Zahlen, ...) annehmen. Zu jedem dieser Werte gehört eine bestimmte Wahrscheinlichkeit p , der p -Wert der Teststatistik. Simuliert man viele solcher Tests, so ergibt sich als Resultat eine Gleichverteilung der p -Werte auf $[0,1]$, d.h. es gibt keinen bevorzugten p -Wert.

- Ausweg: Verwerfe die Nullhypothese H_0 für extreme Ereignisse durch Festlegen eines Signifikanz-Niveaus α . Typische Werte für α : 0.05 oder 0.01 (Statistik ist kein Fall von Mathematik!)

Einseitigen und zweiseitigen Test diskutieren

ZEICHNUNG α Bereich, Fehler 1. und 2. Art

- Dabei können 2 Arten von Fehlern auftauchen:
Fehler 1. Art: H_0 wird abgelehnt, obwohl sie wahr ist.
 falsch positiv
Fehler 2. Art: H_0 wird nicht abgelehnt, obwohl sie falsch ist.
 falsch negativ
- Fehler 2. Art kosten ein gutes paper
 Fehler 1. Art kosten die Karriere

- Power of the test: Häufigkeit der Ablehnung eines Tests, wenn H_0 falsch.

ZEICHNUNG Power of the test

- Häufigkeit Fehler 1. Art $< \alpha$: Test ist konservativ.
- Häufigkeit Fehler 1. Art $> \alpha$: Test ist Schrott.

Übung :

Die Power des t -Tests

Dilemma II: Dichotomie von Kakutani [147]

- Wenn H_0 nicht wahr ist, gilt eine Alternative H_1 mit Verteilung $p_{H_1}(x)$
- Nun gilt:

$$\lim_{N \rightarrow \infty} \int p_{H_0}(X) p_{H_1}(X) dX = \begin{cases} 0 \\ \text{or} \\ 1 \end{cases}$$

Bemerkung: Für grosses N werden Verteilungen $p_{H_0}(X), p_{H_1}(X)$ immer enger. Falls $H_0 \neq H_1$ gibt es irgendwann keinen gemeinsamen Träger mehr.

Idee des Beweises:

O.B.d.A.: $E(p_{H_0}(X)) = 0$

Wenn Test überhaupt Power hat, dann

$$\lim_{N \rightarrow \infty} E(p_{H_1}(X))/\text{Var}(p_{H_1}(X)) \rightarrow \infty$$

- *Wenn ein Test überhaupt Power hat, wird H_0 mit zunehmender Datenzahl immer abgelehnt*
- ”All null hypotheses are wrong” (Fischer, 1925) ”... but some are useful!”

Dilemma III: Statistische Signifikanz vs. Inhaltliche Relevanz

Wiederhole p -Wert

- Patienten mit Puls 180 ± 10 Schläge/min
- Ein Medikament verringert Puls auf 170 ± 10 Schläge/min
- Führe t -test mit n Messungen durch:

$$\begin{array}{ll} n=5 & : \text{ n.s.} \\ n=10 & : p = 0.03 \\ n=100 & : p < 10^{-7} \\ n=1000 & : p < 10^{-20} \end{array}$$

- Drehe es um: Fallzahl Kalkulation
 - Was ist klinisch relevante Herabsetzung des Pulses ?
 - Wie viele Messungen/Patienten n braucht man, um Nullhypothese H_0 : Es gibt keinen Effekt. abzulehnen ?
 - Wenn H_0 dann abgelehnt: Medikament ist sinnvoll.
 - Wenn H_0 nicht abgelehnt basierend auf n Messungen/Patienten: Gemessen in relevante inhaltlicher Skala hat das Medikament keinen Effekt.

Bemerkung: Beliebige kleine Verletzungen führen zu signifikanten Unterschieden falls hinreichend viele Daten verfügbar sind. Man muss sich im Vorfeld überlegen, welche Verletzung relevant ist. Daraus lässt sich dann bestimmen, wieviele Daten notwendig sind, um eine sinnvolle Verletzung statistisch signifikant nachzuweisen.

Dilemma IV: Multiples testen

- Setting: An m Größen soll ge- t -testet werden, ob sich zwei Spezies unterscheiden.
- H_0 : Es gibt keine Unterschiede
- Procedere: Mache m t -Tests, jeweils zu Signifikanzniveau α .
- Wahrscheinlichkeit $\tilde{\alpha}$, H_0 abzulehnen:

$$\tilde{\alpha} = 1 - (1 - \alpha)^m \quad (\ddagger)$$

- Beispiel: $\alpha = 0.01$

$$\begin{aligned} m = 10 &\implies \tilde{\alpha} = 0.1 \\ m = 100 &\implies \tilde{\alpha} = 0.63 \\ m = 1000 &\implies \tilde{\alpha} = 0.99996 \end{aligned}$$

Lösung 1:

- Bonferroni - Korrektur:
- Löse (†) nach α auf:

$$\alpha = 1 - (1 - \tilde{\alpha})^{1/m} \approx \frac{\tilde{\alpha}}{m}$$

- Berechne für gewünschtes (globales) $\tilde{\alpha}$ das notwendige α für die Einzel-Tests.
- Problem :
 - α wird sehr klein,
 - Test sehr konservativ, keine Power
 \implies viele Fehler 2. Art
- Variation: Bonferroni-Holm: Korrigiere in jedem Schritt j mit j/m .

Lösung 2:

- Ein Experiment (m) zur Hypothesen Generierung,
- ergibt $m' \ll m$ Kandidaten.
 ein paar echt positive, ein paar falsch positive
- ein zweites Experiment zum Testen mit m' .

Variation über dieses Thema:

- AIDS Test
- Erst (billiger) sensitiver Test, der unspezifisch ist
- Falls positiv, dann mehrfach (teurer) Test, der spezifisch aber nicht so sensitiv ist.

Lösung 3:

- Benutze Binomial-Verteilung $B(n, p, k)$ mit $p = \alpha$, um Anzahl der erwarteten falsch-positiven abzuschätzen:

$$\langle \#(\text{Falsch positive}) |_{H_0} \rangle = \sum_{k=1}^m kB(m, \alpha, k)$$

- Wenn's viel mehr sind, ist da ein Unterschied.

Oder Bootstrap-Methoden : [20, 323]

Spezialfall: ANOVA

- Betrachte: Experiment untersucht mehrere Bedingungen in der selben Hinsicht.
- Beispiel Placebo, $\text{Med}_1, \dots, \text{Med}_M$ in Bezug auf # Rote Blutkörperchen
- ANalysis Of VAriance (ANOVA) ist die Alternative zu $\frac{M(M-1)}{2}$ t -Tests.

Herleitung:

- H_0 : Kein Effekt.
- M Bedingungen, jeweils N Beobachtungen: x_{ij}
- Empirischen Mittelwert pro Bedingung

$$\bar{x}_i = \frac{1}{N} \sum_{j=1}^N x_{ij}$$

Mittelwert über alles:

$$\bar{x}_{..} = \frac{1}{M} \sum_{i=1}^M \bar{x}_i$$

Varianz aller Daten, genannt SS_{total} , SS für sum of squares, ist:

$$\begin{aligned} SS_{\text{total}} &= \sum_{i=1}^M \sum_{j=1}^N (x_{ij} - \bar{x}_{..})^2 \\ &= \sum_{i=1}^M N(\bar{x}_i - \bar{x}_{..})^2 + \sum_{i=1}^M \sum_{j=1}^N (x_{ij} - \bar{x}_i)^2 \end{aligned}$$

- Erste Summand: Varianz der Gruppenmittel bezüglich des Gesamtmittels (SS_{between}) mit $M - 1$ Freiheitsgraden.

Zweiter Summand: Varianz in den einzelnen Gruppen (SS_{within}) und hat $(N - 1)M$ Freiheitsgrade.

- Unter Gültigkeit der Nullhypothese sollten ihr Quotient somit einer $F(M - 1, (N - 1)M)$ - Verteilung folgen.
- ZEICHUNG zur Veranschaulichung

Im Falle der Ablehnung:

- Ist ANOVA signifikant, stellt sich Frage: Wer war's ?
- A posteriori Test (Tukey-Kramer oder Scheffé) gibt kritische Differenz, die die Mittelwerte überschreiten müssen, um als signifikant verschieden angesehen werden zu können.
- Berücksichtig, daß die Daten für die ANOVA schon einmal statistisch "benutzt" worden sind.
- *A posteriori* Test hat immer kleinere Power als ein einzelner t - Test z.B. zwischen den grössten Mittelwertsdifferenzen. Es ist daher wichtig¹, sich *vor* einem Experiment zu überlegen, wie die minimal zu testende Hypothese lautet. Man läuft sonst Gefahr, vorliegende Effekte statistisch nicht nachweisen zu können.

Die ANOVA läßt sich auf den Fall unterschiedlicher Varianzen und Stichprovenumfänge und vor allem mehrerer Einflußgrößen, der sogenannten Faktoren, verallgemeinern.

Bemerkung: ANOVA in Kurzform: Vergleich der Varianz von Gruppenmittel bezüglich Gesamtmittel mit Varianz in einzelnen Gruppen (F-Test). So kann man sich $\frac{M(M-1)}{2}$ t -Tests sparen.

¹aber leider nicht üblich

2.2.2 Nichtparametrische Tests

Bisher:

- Tests gingen von Gaußverteilten Stichproben aus: Parametrische Tests
- Fällt Verteilung sehr viel langsamer als Gaußverteilung ab, verliert t -Test seine Power, siehe Übung
- Alternative: Nichtparametrische Tests
- Statt Mittelwerts- dann Lokationsvergleich.

$$p_1(x) = p_2(x + \Delta) \quad H_0 : \Delta = 0$$

Hier:

- t -Test geht nach Wilcoxon Rangsummentest (oder U -Test von Mann-Whitney)
- Berechne die Ränge R_1, \dots, R_{N_1} der ersten Stichprobe bezüglich der Gesamtstichprobe.
- Nullhypothese: Ränge sind gleichverteilt.
- Sei $N_1 = N_2 = N$, so gilt:

$$\left\langle \sum_{i=1}^N R_i \right\rangle = N^2 + 0.5N, \quad \text{Var} \left(\sum_{i=1}^N R_i \right) = \frac{1}{6}N^3 + \frac{1}{12}N^2$$

- Mit Zentralem Grenzwertsatz folgt:

$$\left(\sum_{i=1}^N R_i - (N^2 + 0.5N) \right) / \sqrt{\frac{1}{6}N^3 + \frac{1}{12}N^2} \sim N(0, 1)$$

- Nichtparametrische ANOVA: Kruskal/Wallis.

Effizienz

- Sind Daten Gaußverteilt, erkennt parametrischer t - Test eine Mittelwertdifferenz mit weniger Daten als Wilcoxon - Test
- Geringere Power der nichtparametrischen im Vergleich zu den parametrischen Tests bei Gültigkeit der parametrischen Annahmen wird durch die sogenannte Effizienz angegeben.

$$Eff = \frac{Power(NP - test)}{Power(P - test)}$$

gegeben die Gültigkeit der parametrischen Verteilung.

- Wilcoxon-Test hat gegenüber t - Test eine Effizienz von 0.95, d.h. der t - Test hat für Gaußverteilte Daten mit 95% der Datenanzahl die selbe Power wie der Wilcoxon-Test.
- Da falsche Verteilungsannahmen zu einem Verlust der Power parametrischer Tests führen, nichtparametrische Tests aber eine Effizienz < 1 haben, folgt

Dilemma V : Power vs. Effizienz

- Da Unterschied gering und nichtparametrischen Tests robuster, werden sie heute in der Regel bevorzugt.

Gepaarte Tests

- Werden in den Stichproben Daten von den selben Individuen erhoben, so hat dieses eine Auswirkung auf die Varianz:

ZEICHNUNG: gepaarte Werte

die berücksichtigt werden muß.

- Man spricht von gepaarten Tests.
- Wie im Falle des gepaarten t - Tests läßt sich auch für die ANOVA eine Meßwiederholung, i.e. die Abhängigkeit der Stichproben zu den verschiedenen Versuchsbedingungen, berücksichtigen. Dies geschieht z.B. durch die Greenhouse - Geisser - Korrektur.

Mehr Statistik später, wenn's dran iss.

Kapitel 3

Grundlegende Begriffe

3.1 Ergodizität

- Prozesse

$$\dot{x} = f(x, \epsilon)$$

induzieren eine invariante Dichte $\mu(x)$ im Phasenraum.

- Für dissipative (nicht - Hamilton'sche) ist diese i.d.R. attraktiv, i.e. es gibt Transienten.
- Erwartungswerte interessierender Größen müssen bezüglich $\mu(x)$ gebildet werden:

$$\langle G \rangle = \int dx \mu(x) G(x)$$

- Ergodisch: System kommt allen erlaubten Systemzuständen ($\mu(x) \neq 0$) beliebig nahe.
- Oder: Jede Invariante Menge ist entweder die leere oder die gesamte Menge.
- Für ergodische Systeme gilt "Zeitmittel=Scharmittel" und damit

$$\langle G \rangle = \int dx \mu(x) G(x) = \frac{1}{T} \int dt G(x(t)) \quad \left(= \frac{1}{N} \sum_i^N G(x(t_i)) \right)$$

D.h.: man braucht $\mu(x)$ nicht.

Bemerkung: Falls das System ergodisch ist, also allen erlaubten Systemzuständen beliebig nahe kommt, können Erwartungswerte statt über die Dichte direkt als Zeitmittel berechnet werden.

3.2 Stationarität

- Gegeben Prozess x
- Betrachte:

$$p(x(t), x(t - \tau_1), \dots, x(t - \tau_m))$$

- Gilt

$$p(x(t), x(t - \tau_1), \dots, x(t - \tau_m)) = f(\tau_1, \dots, \tau_m) \quad \forall (\tau_1, \dots, \tau_m)$$

dann ist der Prozeß (streng) stationär.

- Betrachte Autokovarianz-Funktion (Annahme $\langle x(t) \rangle = 0$) :

$$ACF(t, \tau) = \langle x(t)x(t - \tau) \rangle$$

- Gilt

$$ACF(t, \tau) = ACF(\tau)$$

heißt der Prozeß schwach stationär.

*Beachte: Stationarität \neq Zeitunabhängige Dynamik
(Zeitunabhängige Dynamik bedeutet hier, daß der funktionelle Zusammenhang zeitunabhängig ist.)*

Beispiele für nichtstationäre Prozesse mit zeitunabhängiger Dynamik:

- Brownian Motion

$$x(t) = x(t-1) + \epsilon(t), \quad ACF(t, 0) \propto t$$

- Periodisch korrelierte Prozesse, z.B.:

$$\begin{aligned} \ddot{x} &= -(\omega_0 + \omega)^2 x \\ \ddot{\omega} &= -\Omega^2 \omega \\ ACF(t, \tau) &= ACF(t + 2\pi/\Omega, \tau) \end{aligned}$$

Es gibt ∞ viele Arten, Prozesse instationär zu machen.

(Pathologisches) Beispiel für schwach, aber nicht streng stationär

$$\begin{aligned} x(i) &= 1 - 4x^2(i-1) \quad \text{für } i < 0 \text{ Logistische Gleichung} \\ x(i) &\sim N(0, 1) \quad \text{für } i \geq 0 \text{ Weisses Rauschen} \end{aligned}$$

3.3 Mixing properties

Zwei Fundamentalklassen von Prozessen

- Nicht-vergessende: i.w. strengperiodische
 - Grenzzyklen (dissipative)
 - Integrale Systeme (konservative)
- Vergessliche
 - Stochastische
 - Chaotische

Literatur: [98]

- φ -mischend (uniformly mixing)

Z_i streng stationärer Prozeß, F_n^m σ -Algebra generiert durch $\{Z_i, n \leq i \leq m\}$

Z_i ist φ mischend, wenn für

$$\varphi_k = \sup_n \sup_{\substack{A \in F_{-\infty}^n \\ P(A) > 0, \\ B \in F_{n+k}^\infty}} |P(B|A) - P(B)|$$

gilt

$$\lim_{k \rightarrow \infty} \varphi_k = 0 \quad .$$

Bemerkung: Betrachte zwei beliebige Ereignisse im zeitlichen Abstand k und wähle solche, bei denen das spätere Ereignis am meisten vom früheren Ereignis weiss. Wenn für $k \rightarrow \infty$ diese Kenntnis trotzdem Null wird, ist der Prozess φ -mischend.

- ρ mixing, asymptotisch unkorreliert

Z_i mit endlicher Varianz (o.B.d.A. $\langle Z \rangle = 0$) ist ρ -mischend, wenn für

$$\rho_k = \sup_n \sup_{\substack{X \in L_2(F_{-\infty}^n) \\ Y \in L_2(F_{n+k}^\infty)}} \frac{\langle XY \rangle}{(\text{Var}(X)\text{Var}(Y))^{1/2}}$$

gilt

$$\lim_{k \rightarrow \infty} \rho_k = 0$$

$L_2(F_n^m)$: Menge der ZV X , messbar auf F_n^m , $0 < E(X^2) < \infty$

Bemerkung: Anschaulich ist die Grösse ρ_k ähnlich zur *ACF*. D.h. ρ -mischend heisst die *ACF* zerfällt.

- Z_i ist α (strong) mixing, wenn für

$$\alpha_k = \sup_n \sup_{\substack{A \in F_{-\infty}^n \\ B \in F_{n+k}^\infty}} |P(B \cap A) - P(A)P(B)|$$

gilt:

$$\lim_{k \rightarrow \infty} \alpha_k = 0 \quad .$$

Bedeutung:

- Mischende Prozesse, hinreichend grob gesamlet ergeben iid Zufallsvariable (so funktionieren Zufallszahlengeneratoren).
- Mixing Konzept ursprünglich für stochastische Prozesse entwickelt, läßt sich aber auf deterministische übertragen [170].
- Erlaubt statistische Zugänge

Gegeben Daten, laufen Fragen nach Stationariät, Ergodizität, Mischungseigenschaften (der zu Grunde liegenden Prozesse) auf statistische Tests hinaus (inclusive Dilemmata I-V)

Kapitel 4

Exemplarische Prozesse und Datensätze

4.1 Deterministische Systeme

4.1.1 Van der Pol Oscillator

Literatur: [307]

$$\ddot{x} = \mu(1 - x^2)\dot{x} - x, \quad \mu > 0. \quad (4.1)$$

$$\begin{aligned} \dot{x}_1 &= x_2 \\ \dot{x}_2 &= \mu(1 - x_1^2)x_2 - x_1 \end{aligned}$$

mit x_1 Ort und x_2 Geschwindigkeit.

BILDER van der Pol

van der Pol ist nicht mischend.

Bemerkung: Für $\mu = 0$ ergibt sich der harmonische Oszillator mit Frequenz 1. Ziel von van der Pol war, einen Oszillator zu konstruieren, der gegen einen Grenzyklus geht. Der Grenzyklus ist über den Parameter μ zu steuern.

Stabilitätsuntersuchungen:

- Fixpunkt:

$$\vec{f}(\vec{x}^*) = 0$$

Für van der Pol folgt: $x_1^* = x_2^* = 0$

- Stabilität des Fixpunktes: Linearisiere Dynamik um Fixpunkt

Mit

$$x_1 = x^* + \tilde{x}, \quad x_2 = x^* + \tilde{x}$$

$$\begin{aligned} (x^* + \tilde{x})' &= \dot{\tilde{x}} = f(x^* + \tilde{x}, y^* + \tilde{y}) \approx f(x^*, y^*) + \frac{\partial f(x^*, y^*)}{\partial x} \tilde{x} + \frac{\partial f(x^*, y^*)}{\partial y} \tilde{y} \\ (y^* + \tilde{y})' &= \dot{\tilde{y}} = g(x^* + \tilde{x}, y^* + \tilde{y}) \approx g(x^*, y^*) + \frac{\partial g(x^*, y^*)}{\partial x} \tilde{x} + \frac{\partial g(x^*, y^*)}{\partial y} \tilde{y} \end{aligned}$$

Ergibt mit $\vec{x} = \begin{pmatrix} \tilde{x} \\ \tilde{y} \end{pmatrix}$

$$\dot{\vec{x}} = \begin{pmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} \\ \frac{\partial g}{\partial x} & \frac{\partial g}{\partial y} \end{pmatrix} \vec{x} = A\vec{x}$$

Lösung:

$$\vec{x}(t) = \vec{x}(0)e^{At}$$

Die Eigenwerte $\lambda_{1,2}$ von A legen das qualitative Verhalten fest:

$$\tilde{x}(t) = a_1 e^{Re(\lambda_1)t} \cos(Im(\lambda_1)t + \phi_1) + a_2 e^{Re(\lambda_2)t} \cos(Im(\lambda_2)t + \phi_2)$$

Allgemein:

$$\exists \operatorname{Re} \lambda > 0 \quad \text{instabil} \quad \forall \operatorname{Re} \lambda < 0 \quad \text{stabil}$$

- Eigenwerte reell: exponentiell
- Eigenwerte complex: Spiral

Für van der Pol:

$$\left. \frac{\partial \vec{f}}{\partial \vec{x}} \right|_{x^*} = \begin{pmatrix} 0 & 1 \\ -1 & \mu \end{pmatrix}$$

$$\det \begin{pmatrix} -\lambda & 1 \\ -1 & \mu - \lambda \end{pmatrix} = \lambda^2 - \lambda\mu + 1$$

$$\lambda_{1,2} = \frac{\mu}{2} \pm \sqrt{\frac{\mu^2}{4} - 1}$$

- Realteile positiv
- $\mu < 2$ imaginär, $\mu \geq 2$ reell

ZEICHNUNG dazu

4.1.2 Rössler System

Literatur: [244]

$$\begin{aligned} \dot{x} &= -y - z \\ \dot{y} &= x + ay \\ \dot{z} &= b + (x - c)z \end{aligned}$$

mit $a, b, c > 0$. typische Parameter $a = b = 0.1$, c ist Kontrollparameter, $c = 14$ gibt Chaos.

BILD Rössler-System

Sehr langsam mischend [216].

Bemerkung: Betrachtet man zunächst die xy-Ebene alleine mit $a = 0$ ergibt sich erneut der harmonische Oszillator. Dieses Oszillieren in der xy-Ebene bleibt auch bei Betrachtung aller Komponenten mit beliebigen Parametern im wesentlichen erhalten. Über die z-Komponente kommt eine zusätzlich Variabilität in das System, die über den Parameter c gesteuert werden kann.

4.1.3 Integration deterministischer Systeme

Integration deterministischer Systems basiert auf einer Taylorentwicklung

$$\dot{y} = f(y)$$

$$y_{t+h} = y_t + \dot{y}_t h + \frac{1}{2} \ddot{y}_t h^2 + \frac{1}{6} y_t^{(3)} h^3 + O(h^4) \quad (\dagger)$$

\dot{y}_t gegeben durch $f(y_t)$, aber $y_t^{(n)}$ möchte man nicht ausrechnen.

Abbruch nach erster Ordnung: Euler-Verfahren:

$$y_{t+h} = y_t + f(y_t)h + O(h^2)$$

”Erster Ordnungs Verfahren”

- Idee: Höhere Ordnung durch geschickte Funktionsauswertungen.
- Ziel:

$$y_{t+h} = y_t + f(y_t)h + \frac{1}{2} f'(y_t) f(y_t) h^2 \quad (*)$$

aber: ohne erste Ableitung $f'(y_t)$.

- Lösung: Betrachte

$$\begin{aligned} \text{Ansatz } y_{t+h} &= y_t + f(y_t + \frac{1}{2} k_1) h \\ \text{mit } k_1 &= f(y_t) h \\ \Leftrightarrow y_{t+h} &= y_t + f(y_t + \frac{1}{2} f(y_t) h) h \\ &= y_t + f(y_t) h + f'(y_t) (\frac{1}{2} f(y_t) h) h \\ &= y_t + f(y_t) h + \frac{1}{2} f'(y_t) f(y_t) h^2 \quad (*) \end{aligned}$$

- Mit (*) cancelled sich der 2. Ordnungs-Term in (†) und man erhält ein Verfahren 2. Ordnung (Midpoint Method).

- Also:

$$y_{t+h} = y_t + f(y_t + \frac{1}{2}f(y_t)h)h$$

ist ein Verfahren 2. Ordnung ohne Ableitung von f .

- Dieses läßt sich weiterspinnen, üblich Runge-Kutta 4. Ordnung.
- Auf Grund der hohen Ordnung gilt in der Regel: Integrations-schrittweite = Samplingschrittweite
- Vernünftige Samplingschrittweite erklären

Übung :

Integration von Rössler- und van der Pol- System

4.2 Stochastische Systeme

Literatur für stochastische Prozesse im allgemeinen:

- A. Papoulis: Probability, random variables and stochastic processes [214]
- C.W. Gardiner: Handbook of Stochastic Methods [85]
- W.A. Gardner: Introduction to Random Processes with Application to Signals and Systems [86]
- N.G. van Kampen: Stochastic Processes in Physics and Chemistry [308]

4.2.1 Weißes Rauschen

$$x(i) = \sigma\epsilon(i), \quad \epsilon(i) \sim N(0, 1)$$

$$ACF_{WN}(\tau) = \langle x(i) x(i - \tau) \rangle = \sigma^2 \delta(\tau)$$

modelliert zahlreiche Einflüsse.

4.2.2 Autoregressive Prozesse

Literatur: [332, 36, 227]

4.2.2.1 AR[1] Prozess

Drei Zugänge:

- Verallgemeinerung der linearen Regression (Mathematiker): Erkläre möglichst viel Varianz von $y(i)$ durch $x(i)$

$$y(i) = ax(i) + \sigma\epsilon(i)$$

ist oft (grade bei wenigen, verrauschten Daten) keine schlecht Näherung, darum versuche für Prozesse:

$$x(i) = ax(i-1) + \sigma\epsilon(i)$$

Bemerkung: Was man aus vorherigen Schritt nicht erklären kann, “steckt” man in das Rauschen.

- Diskrete lineare Dynamik (Yule [332])

Eine lineare Dynamik 1. Ordnung $\dot{x}(t) = -\frac{1}{\tau}x(t)$ führt auf $x(t) = Ae^{-t/\tau}$. Additionstheorem:

$$x(t) = Ae^{-t/\tau} = Ae^{-(t-1+1)/\tau} = Ae^{-1/\tau} e^{-(t-1)/\tau} = e^{-1/\tau} x(t-1) = ax(t-1)$$

Um es realistisch zu machen, störe die Zeitentwicklung in jedem Schritt ein bißchen. Dies führt zu:

$$x(i) = ax(i-1) + \sigma\epsilon(i), \quad 0 < a < 1, \quad \epsilon(i) \sim N(0, 1)$$

- Diskretisierte lineare Dynamik (Physiker)

Betrachte stochastischen Relaxator

$$\dot{x} = -\alpha x + \sigma' \epsilon$$

Für kleines Δt aufintegriert (für $\sqrt{\Delta t}$ siehe unten):

$$\begin{aligned}x(t) &= x(t - \Delta t) - \Delta t \alpha x(t - \Delta t) + \sqrt{\Delta t} \sigma \epsilon \\ &= (1 - \Delta t \alpha)x(t - \Delta t) + \sqrt{\Delta t} \sigma \epsilon \\ \Rightarrow x(i) &= ax(i - 1) + \sigma' \epsilon(i)\end{aligned}$$

mit $\Delta t = 1$ und t wird identifiziert mit i .

Der relaxierende Prozeß wird vom Rauschen immer wieder aus der Gleichgewichtslage um 0 weggetrieben.

Der Prozeß heißt Autoregressiver Prozeß erster Ordnung, AR[1].

FOLIE AR1

- Formal: Bedingte Wahrscheinlichkeit $p(x(i)|x(i - 1))$ für (die ZV) $x(i)$ gegeben die Zahl (Realisierung) $x(i - 1)$:

$$p(x(i)|x(i - 1)) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x(i) - ax(i-1))^2}{2\sigma^2}}$$

- Für seine Varianz folgt mit:

$$\langle x^2(i) \rangle = \langle (ax(i - 1) + \sigma\epsilon(i))^2 \rangle$$

Auf Grund von Stationarität und Orthogonalität:

$$\langle x^2(i) \rangle = \frac{\sigma^2}{1 - a^2}$$

- Da

$$\langle x^2(i) \rangle > \sigma^2$$

Autoregression nimmt Energie von $\epsilon(i)$ auf.

- Seine Autokorrelationsfunktion $ACF(\tau)$ lautet (vorrechnen):

$$ACF(\tau) = \frac{\langle x(t)x(t - \tau) \rangle}{\langle x^2(t) \rangle} = a^\tau \quad \tau \geq 0$$

Bedenke: $a < 1$, also zerfällt die *ACF* exponentiell.

Merke:

*Es gilt allgemein für $AR[p]$ Prozesse, daß die *ACF* die Prozeßparameter eindeutig festlegt.*

Bemerkungen:

- Name "Autoregressiv" ist historisch bedingt in Unterscheidung von "moving average", siehe Kapitel 4.2.2.4.
- Alle dynamischen Prozesse sind autoregressiv.
- "Autoregressiv" meint "Linear autoregressiv mit Gaußischem Rauschen"

4.2.2.2 AR[2] Prozess

$$x(i) = a_1x(i-1) + a_2x(i-2) + \epsilon(i)$$

ergibt mit:

$$\begin{aligned} a_1 &= 2 \cos(2\pi/T)e^{-1/\tau} \\ a_2 &= -e^{-2/\tau} \end{aligned}$$

einen stochastisch getriebenen, gedämpften Oszillator der Periode T und Relaxationszeit τ . Zeitdiskrete Version von $\ddot{x} + \gamma\dot{x} + \omega_0x = 0$

FOLIEN AR[2]

- Mischungseigenschaften AR[p] Prozesse:

$$\phi_k = A\beta^k, \quad 0 < \beta < 1$$

exponentiell ϕ -mischend

- Lineare Gaußsche Prozesse sind multivariat gaußverteilt, Buckel in der Mitte. Beispiel für Loch in der Mitte ist der van der Pol, der nichtlinear ist.

FOLIE dazu

- (Gaußsche) AR-Prozesse sind zeitumkehrinvariant [320], da die $ACF(\tau)$ die Parameter eindeutig festlegt und da die ACF symmetrisch ist:

$$ACF(\tau) = \langle x(t)x(t+\tau) \rangle = \langle x(t+\tau)x(t) \rangle = \langle x(t)x(t-\tau) \rangle = ACF(-\tau)$$

$$"x(t) = ax(t+1) + \epsilon(t)"$$

FOLIE AR[2] PROZESS umdrehen

Übung:

Für welche Parameter a_1, a_2 sind AR[2] Prozesse stationär ?

4.2.2.3 Partielle ACF

Seien Z_i Zufallsvariablen (ZV).

- Betrachte:

Best linear predictor:

$$P_{1,Z_1,\dots,Z_n}(X) = \sum_{i=0}^n \alpha_i Z_i, \quad Z_0 = 1$$

mit α_i aus:

$$\langle X, Z_j \rangle = \sum_{i=0}^n \alpha_i \langle Z_i Z_j \rangle$$

- Die PACF ist definiert:

$$PACF(1) = Corr(x(i), x(i-1)) = ACF(1), \quad \text{für } \tau = 1$$

$$PACF(\tau) = Corr(x(t) - P_{x_{t-1}, \dots, x_{t-\tau+1}}(x(t)), x(t+\tau) - P_{x_{t-1}, \dots, x_{t-\tau+1}}(x(t+\tau)))$$

$PACF(\tau)$: Korrelation zwischen $x(t)$ und $x(t-\tau)$, wenn man die Information der $x(t-1), \dots, x(t-\tau+1)$ herausprojiziert.

- Betrachte AR[1] Prozeß:

$$x(t) = ax(t-1) + \epsilon(t)$$

$\tau = 1$:

$$PACF(1) = a$$

$\tau > 1$:

$$P_{x_{t-1}, \dots, x_{t-\tau+1}}(x(t)) = ax(t-1)$$

und

$$P_{x_{t-1}, \dots, x_{t-\tau+1}}(x(t-\tau)) = ax(t-\tau+1)$$

Damit für $\tau > 1$:

$$\begin{aligned} PACF(\tau) &= Corr(x(t) - ax(t-1), x(t-\tau) - ax(t-\tau+1)) \\ &= Corr(\epsilon(t), \epsilon(t-\tau)) \\ &= 0 \end{aligned}$$

- Allgemein für AR[p]-Prozesse:

- $PACF(\tau) \neq 0$ für $\tau \leq p$
- $PACF(\tau) = 0$ für $\tau > p$
- Ein Testproblem

(läßt sich auf nichtlineare Prozesse verallgemeinern)

Bemerkung: Über diesen Weg kann sich die Ordnung des AR -Prozesses bestimmen lassen. Wird an den AR -Prozess ein Modell mit der richtigen Ordnung angefitet, so ist der Rest unkorreliertes Gaußsches Rauschen.

4.2.2.4 Moving average (MA[q]) Prozesse

$$x(i) = \sum_{j=0}^q b_j \epsilon(i-j)$$

Werden wichtig beim Kapitel Filterung
Autokovarianz:

$$ACF(\tau) = \begin{cases} \sigma^2 \sum_{j=0}^q b_j b_{j+k} & \tau \leq q \\ 0 & \tau > q \end{cases}$$

mit $b_j = 0$ für $j > q$.

MA[q]-Prozesse:

- $ACF(\tau) = 0$ für $\tau > q$
- $PACF(\tau) \neq 0 \forall \tau$
- Ein Testproblem

Bemerkung: Über diesen Weg kann sich die Ordnung des *MA*-Prozesses bestimmen lassen. Dabei wird die entscheidende Information über die *ACF* vermittelt.

4.2.2.5 ARMA-Prozesse

Kombination von AR(p) und MA(q) liefert ARMA(p,q)-Prozesse:

$$x(i) = \sum_{j=1}^p a_j x(i-j) + \sum_{j=1}^q b_j \epsilon(i-j) + \epsilon(i)$$

Hat man einen gesampelten kontinuierlichen AR[p] Prozess, so ergibt sich im diskreten ein Zustandsraummodell, siehe Kap. 6.5.

Bemerkung: Für ARMA-Prozesse ist die Bestimmung der Ordnung über ACF/PACF nicht möglich.

4.2.2.6 Invertierung von AR Prozessen

$$x(i) = ax(i-1) + \epsilon(i)$$

Betrachte Backshift-Operator:

$$B(x(i)) = x(i-1)$$

Damit wird aus AR[1] Prozeß :

$$\begin{aligned} (1 - aB)x(i) &= \epsilon(i) \\ x(i) &= \frac{1}{1 - aB} \epsilon(i) \\ &= \sum_{j=1}^{\infty} (aB)^j \epsilon(i) \end{aligned}$$

ein MA(∞) Prozeß.

Dieses geht unter schwachen Annahmen \forall AR[p] Prozesse und heißt **Invertierung**.

Bemerkung: Die Eigenwerte eines stationären AR[p]-Prozesses sind < 1 . Deswegen ist $||aB|| < 1$. Damit funktioniert die Invertierung, d.h. die geometrische Reihe konvergiert.

Analog lassen sich MA[q] Prozesse als AR[∞] Prozesse schreiben.

Aber: Es gibt vernünftige MA[q] Prozesse, die sich nicht invertieren lassen.

Allgemein gültige Folgerung: Hat man die falsche Modellklasse, braucht man Prozesse ∞ Ordnung zur Modellierung.

Falsches Modell \implies Ordnung = ∞

4.2.3 Stochastischer van der Pol Oscillator

$$\ddot{x} = \mu(1 - x^2)\dot{x} - x + \epsilon, \quad \mu > 0. \quad (4.2)$$

$$\begin{aligned} \dot{x}_1 &= x_2 \\ \dot{x}_2 &= \mu(1 - x_1^2)x_2 - x_1 + \epsilon \quad , \end{aligned}$$

4.2.3.1 Kleiner Exkurs: Stochastische Integrale

$$\begin{aligned}\dot{x} &= -\alpha x + \sigma \epsilon \\ x_{t+h} &= x_t + \int_t^{t+h} -\alpha x(t') dt' + \sigma \int_t^{t+h} \epsilon(t') dt'\end{aligned}$$

Aber was ist ein Integral über $\epsilon_{t'}$??

- Betrachte:

$$\int_t^{t+h} \epsilon_{t'} dt'$$

Macht weder im Riemann- noch im Lebesgue-Sinn Sinn.

- Beobachtung:
Ergebnis des Integrals ist Brownian Motion (BM)
- BM in diskreter Zeit ($\Delta t = 1$) ist:

$$x(t+1) = x(t) + \sigma \epsilon(t) \quad x(0) = 0, \quad \epsilon \sim N(0, 1)$$

- Da Varianzen additiv, gilt

$$\langle x^2(t) \rangle = \sigma^2 t, \quad \langle x(t) \rangle = 0, \quad \text{sowie so}$$

$x(t)$ ist Zufallsvariable mit Mittelwert 0 und Standardabweichung $\sigma\sqrt{t}$

- DEFINIERE:

$$\int_t^{t+h} \epsilon_{t'} dt' = \sqrt{h} \epsilon_t$$

- Bemerkung: Die Mathematiker drehen es herum.
 1. Definiere zeitkontinuierliche Brownian Motion
 2. Definiere "ε" als Zuwächse der Brownian Motion

Ito und Stratonovich-Verfahren (s. Kapitel 6.8):

- für additives Rauschen identisch
- für multiplikatives Rauschen unterschiedlich

Damit Euler-Verfahren für $\dot{x} = a(x) + b(x)\epsilon$ (nach Ito)

$$x_{t+h} = x_t + a(x_t)h + b(x_t)\epsilon_t\sqrt{h}$$

Höhere Ordnung i.a. sehr schwierig, da komplizierte stochastische Integrale, siehe [125].

Exkurs Ende

Auf Grund des Rauschterms ϵ können stochastische Differentialgleichungen nicht mit Verfahren hoher Ordnung integriert werden [162, 163].

Euler-Verfahren für stochastischen van der Pol:

$$\begin{aligned}x_1(t + \delta t) &= x_1(t) + \delta t x_2(t) \\x_2(t + \delta t) &= x_2(t) + \delta t (\mu (1 - x_1^2(t)) x_2(t) - x_1(t)) + \sqrt{\delta t} \epsilon(t) \quad .\end{aligned}$$

Auf Grund der geringen Ordnung des Verfahrens muß i.a. der Integrations-schritt δt viel kleiner gewählt werden als der Samplingschritt Δt [286].

FOLIE SVDP

(Nichtpathologische) stochastische Prozesse sind immer mischend.

Übung :

Realisierung von AR[p] und stochastischem van der Pol- System

4.3 Exemplarische Datensätze

4.3.1 Sunspots

- Die Magneto-Hydrodynamik der Sonne produziert einen 22 jährigen Zyklus der Sonnenaktivität, dessen Extrema zu fleckenhaften Verdunklungen der Sonnenoberfläche in etwa elf-jährlichem Rhythmus führen [272].

- Die Sonnenflecken Zeitreihe besteht aus jährlichen Mittelwerten, in die professionelle und hobbymäßige Beobachtungen eingehen. Beobachtung seit ca. 1700 durchgehend.
- Charakteristische Eigenschaft: Asymmetrie up and down, aber siehe [217], Vorverarbeitung wichtig !
- Leseempfehlung: G.U. Yule: On a method of investigating periodicities in disturbed series, with special reference to Wolfer's sunspot numbers [332] die Entwicklung der AR Modelle.

3 FOLIEN Sunspots

4.3.2 Canadische Luchse

Anzahl der in den Jahren 1820-1936 in Kanada gefangenen Luchse.
FOLIE Canadische Luchse

Auf Grund von Tierschutz seit 1936 keine Fortsetzung der Reihe

4.3.3 Southern Oscillation Index

Im Südpazifik gibt es eine Klimaschaukel, die mit einer Periode von 3-6 Jahren für extremes Klima von Australien bis Amerika sorgt. El Nino.

FOLIEN SOI

Problem aller Datensätze: Zu kurz, um sich zu wehren

Es kann kaum eine Hypothese verworfen werden (keine Power).

Vorgeschlagene Modelle für SOI

- Linear stochastischer Oscillator [48]
- Chaos [116] $D_2 = 1.7$:-) [304]
- Self organized criticality, Long range dependency [8]
- Antipersistenz [11, 194]
- Delay Differential Gleichung [298]

Für Sonnenfleckendaten: Alles, wo gibt.

Alle diese Prozesse sind nicht (streng/schwach) stationär. Aber erinnere Dilemma III : Für welche Frage macht welches Ausmaß der Verletzung wieviel aus ? Siehe z.B. [285].

Kapitel 5

Spektralanalyse

Literatur:

- M.B. Priestley *Spectral Analysis and Time Series* [227].
Der mathematische Klassiker No.1
- P.J. Brockwell, R.A. Davis *Time Series: Theory and Methods* [36].
Der mathematische Klassiker No.2
- R.H. Shumway and D.S. Stoffer: *Time Series Analysis and Its Application* [267]
- D.S.G. Pollock *A Handbook of Time-Series Analysis, Signal Processing and Dynamics* [223]
- J. Honerkamp *Stochastic Dynamical Systems* [125] Kap. 13.3
Kondensierte Darstellung für Physiker

5.1 Autospektrum

!! ALLES IMMER NUR BIS AUF NORMIERUNG !!

5.1.1 Theorie im Unendlichen

Erinnere Harmonischen Oszillator:

$$\ddot{x} = -\omega_0 x, \quad x(t) = A \cos(\omega_0 t)$$

Spektrum :

$$S(\omega) = \left| \int dt e^{-i\omega t} x(t) \right|^2 = A^2 \delta(\omega - \omega_0)$$

Betrachte stationären zeitdiskreten Prozeß $x(t)$, $t \in Z$

Sei

$$f(\omega) = \lim_{N \rightarrow \infty} \frac{1}{2\sqrt{\pi N}} \sum_{t=-N}^N e^{-i\omega t} x(t)$$

seine Fouriertransformierte.

Dann ist das Spektrum

$$S(\omega) = \langle |f(\omega)|^2 \rangle$$

Wegen

$$\begin{aligned} \langle |f(\omega)|^2 \rangle &= \langle f(\omega) f^*(\omega) \rangle \\ &= \lim_{N \rightarrow \infty} \frac{1}{4\pi N} \left\langle \sum_{t_1, t_2 = -N}^N x(t_1) x(t_2) e^{-i\omega t_1} e^{i\omega t_2} \right\rangle \\ &= \lim_{N \rightarrow \infty} \frac{1}{4\pi N} \sum_{\tau, t_1} \langle x(t_1) x(t_1 - \tau) \rangle e^{-i\omega \tau}, \quad \tau = t_1 - t_2 \\ &= \lim_{N \rightarrow \infty} \frac{1}{4\pi N} \sum_{t_1} \sum_{\tau} \langle x(t_1) x(t_1 - \tau) \rangle e^{-i\omega \tau} \\ &= \frac{1}{4\pi N} 2N \sum_{\tau} ACF(\tau) e^{-i\omega \tau} \end{aligned}$$

gilt auch

$$S(\omega) = \frac{1}{2\pi} \sum ACF(\tau) e^{-i\omega \tau},$$

was die Definition der Mathematiker ist.

Die Fouriertransformation ist unitär (alle Eigenwerte = 1, d.h. keine unterschiedliche Gewichtung der Fourierkomponenten), d.h.:

$$\int S(\omega) d\omega = \text{Var}(x(t))$$

Spektrum ist "Varianz pro Frequenz" Darstellung des Prozesses.

5.1.1.1 Spektrum von weißem Rauschen

- Weißes Rauschen:

$$x(t) = \epsilon(t), \quad \epsilon(t) \sim N(0, \sigma^2)$$

- Fouriertrafo:

$$f(\omega) = \frac{1}{2\sqrt{\pi}} \sum e^{-i\omega t} \epsilon(t) = \frac{1}{2\sqrt{\pi}} \eta, \quad \eta \sim N_C(0, \sigma^2)$$

mit unabhängigem Real- und Imaginärteil.

- Beachte: $f(\omega)$ ist Zufallsvariable
- Spektrum:

$$S(\omega) = \langle |f(\omega)|^2 \rangle = \frac{\sigma^2}{2\pi}$$

Bemerkung: Betrachte $\int S(\omega) d\omega = \text{Var}(x(t))$ für weißes Rauschen:
 $\int S(\omega) d\omega = \int \frac{\sigma^2}{2\pi} d\omega = \sigma^2 = \text{Var}(x(t))$.

5.1.1.2 Spektren von AR[p] Prozessen

Definiere Backshift-Operator:

$$B(x(t)) = x(t-1)$$

Sei

$$f(\omega) = \frac{1}{\sqrt{2\pi N}} \sum_{t=1}^N e^{-i\omega t} x(t)$$

(Normierung fehlt ab jetzt) dann ist

$$\sum_{t=1}^N e^{-i\omega t} B(x(t)) = \sum_{t=1}^N e^{-i\omega t} x(t-1) = e^{-i\omega} \sum_{t=1}^N e^{-i\omega t} x(t) = e^{-i\omega} f(\omega)$$

Allgemein durch Induktion:

$$\sum_{t=1}^N e^{-i\omega t} B^d(x(t)) = e^{-id\omega} f(\omega)$$

AR[p] Prozeß:

$$x(t) - \sum_{j=1}^p a_j B^j(x(t)) = \epsilon(t)$$

Fouriertrafo:

$$f(\omega)(1 - \sum_{j=1}^p a_j e^{-ij\omega}) = \tilde{\epsilon}$$

Spektrum:

$$S_{AR}(\omega) = \langle |f(\omega)|^2 \rangle = \frac{1}{2\pi} \frac{\sigma^2}{|1 - \sum_{j=1}^p a_j e^{-ij\omega}|^2}$$

Polynom im Nenner führt zu Peaks im Spektrum des AR-Prozesses

Wichtig:

Spektrum von AR[p]-Prozess ist glatt.

ZEICHUNG AR[1], AR[2] Spektren

Glattheit gilt auch

- für alle nichtlinearen stochastischen Prozesse.
- für alle chaotischen Prozesse, bis auf eventuell einzelne Frequenzen.

Allgemein: Das Spektrum ist immer glatt, wenn die ACF zerfällt.

- Gilt, wenn der Prozess vergeblich, also mischend ist. In der Regel sind alle stochastischen und chaotischen Prozesse mischend.

- “Beweis”:

Wenn ACF zerfällt, z.B.

$$ACF(\tau) = \cos(\tau\omega_0) e^{-\tau^2/c}$$

dann geht Produkt im Zeitraum über in Faltung im Frequenzraum:

$$S(\omega) = \delta(\omega - \omega_0) * e^{-\omega^2/c} ,$$

da die Fouriertransformierte einer Gauß-Funktion wieder eine Gauß-Funktion ist, hier also $FT(e^{-t^2/c}) = e^{-\omega^2/c}$ bis auf Normierung.

- Ein ähnliches Argument gilt für jedes Zerfallsverhalten.

$$FT(e^{-t/\tau}) = \sqrt{\frac{2}{\pi}} \frac{1/\tau}{1/\tau^2 + \omega^2}$$

Alle Frequenzen sind sofort bevölkert, das Spektrum also glatt.

Kompaktschreibweise:

AR-Prozeß:

$$\phi(B)x(t) = \epsilon(t)$$

mit

$$\phi(z) = 1 - a_1z - a_2z^2 - \dots - a_pz^p$$

Dann lautet das Spektrum

$$S_{AR}(\omega) = \frac{\sigma^2}{2\pi} \frac{1}{|\phi(e^{-i\omega})|^2}$$

5.1.1.3 Spektrum Moving Average Prozeß

$$X(t) = \sum_{j=0}^q b_j \epsilon(t-j) = \Theta(B)\epsilon(t)$$

mit

$$\Theta(z) = 1 + b_1z + b_2z^2 + \dots + b_qz^q \quad (\text{O.B.d.A. : } b_0 = 1)$$

Analog zu oben:

$$f(\omega) = \sum_{t=1}^N e^{-i\omega t} x(t) = \sum_{t=1}^N e^{-i\omega t} \Theta(B)\epsilon(t)$$

folgt:

$$S_{MA}(\omega) = \frac{\sigma^2 |\Theta(e^{-i\omega})|^2}{2\pi \cdot 1}$$

Polynom im Zähler: MA macht dips im Spektrum

Zusammen:

$$S_{ARMA}(\omega) = \frac{\sigma^2 |\Theta(e^{-i\omega})|^2}{2\pi |\phi(e^{-i\omega})|^2}$$

Die Funktion:

$$H = \frac{\Theta(e^{-i\omega})}{\phi(e^{-i\omega})}$$

heißt Transfer-Funktion des Systems. Sie beschreibt, was aus WN im speziellen oder sonstigem Input i.a. wird.

Bemerkung: Es gilt: $S = H^* \sigma^2 H$ für alle Prozesse, die sich mit einer Transferfunktion darstellen lassen.

Wertet man

$$|H| = \frac{|\Theta(e^{-i\omega})|}{|\phi(e^{-i\omega})|}$$

nicht nur auf dem Einheitskreis $e^{i\omega}$ sondern auf der ganzen komplexen Ebene aus, spricht man von der z -Transformation. Diese macht Pole und Nullstellen explizit sichtbar. Nähe zum Einheitskreis bestimmt "Durchschlagskraft" (erinnere "Resonanzen" in der QM).

ZEICHNUNG Z-TRAFO

5.1.1.4 Diskret vs. Kontinuierlich Part I

Das Spektrum eines AR[1] Prozesses war:

$$S(\omega) = \langle |f(\omega)|^2 \rangle = \frac{\sigma^2}{|1 - a e^{-i\omega}|^2} = \frac{\sigma^2}{1 + a^2 - 2a \cos(\omega)}. \quad (5.1)$$

Für den kontinuierlichen Fall

$$\dot{x} = -\alpha x + \epsilon, \quad (5.2)$$

bestimmt der Parameter α die Relaxationszeit τ durch:

$$\tau = \frac{1}{\alpha}. \quad (5.3)$$

Das Spektrum ergibt sich aus:

$$\int e^{-i\omega t} \dot{x} dt = -\alpha \int e^{-i\omega t} x dt + \int e^{-i\omega t} \epsilon dt, \quad (5.4)$$

Partielle Integration und Sortieren führt zu:

$$f(\omega) = \frac{\int e^{-i\omega t} \epsilon(t) dt}{\alpha - i\omega}. \quad (5.5)$$

und

$$S(\omega) \propto \frac{\sigma^2}{\alpha^2 + \omega^2}. \quad (5.6)$$

Der Zusammenhang folgt aus der 1.Ordnungs-Näherung des diskreten Modells an das kontinuierliche.

$$a = 1 - \Delta t \alpha. \quad (5.7)$$

Nähert man den Nenner in Gl. (5.1) in 2. Ordnung und nutzt Gl. (5.7), ergibt sich

$$1 + a^2 - 2a \cos(\omega) \approx \alpha^2 + \omega^2 - \alpha\omega^2 - \frac{1}{6}(1 - \alpha)\omega^4 + \dots. \quad (5.8)$$

FOLIE AR1 KONTI-DIS

Die Differenz zeigt sich im Hochfrequenten.

The difference between the spectra of discrete and continuous time processes shows up whenever a sampled continuous time process is modeled by a discrete time model and spectra are calculated from the fitted parameters.

5.1.1.5 Cramér Darstellung

Die Fourierdarstellung

$$f(\omega) = \sum_{t=-\infty}^{\infty} e^{-i\omega t} x(t)$$

läßt sich auch umdrehen:

$$x(t) = \sum_{\omega=-\infty}^{\infty} e^{i\omega t} f(\omega)$$

Dies heißt Cramér-Darstellung eines Prozesses.

- Nun ist $f(\omega)$ das zu Grunde liegende und $x(t)$ das abgeleitete.
Statt $f(\omega)$ wird i.d.R. die Bezeichnung $dZ(\omega)$ ("increment process") gewählt.
- Fragen:
 1. Welche Eigenschaften muß ZV $dZ(\omega)$ haben ?
 2. Welche Eigenschaften von $dZ(\omega)$ liefern welche von $x(t)$?

Ad 1. Soll der Prozess $x(t)$ stationär sein, so gilt mit $t_2 = t_1 - \tau$:

$$\begin{aligned} \langle f(\omega_1) f^*(\omega_2) \rangle &= \left\langle \sum_{t_1, t_2} x(t_1) x(t_2) e^{-i\omega_1 t_1} e^{i\omega_2 t_2} \right\rangle \\ &= \sum_{\tau, t_1} \langle x(t_1) x(t_1 - \tau) \rangle e^{-i\omega_2 \tau} e^{-i(\omega_2 - \omega_1) t_1} \\ &= \sum_{\tau} ACF(\tau) e^{-i\omega_2 \tau} \delta(\omega_2 - \omega_1) \\ &= S(\omega_1) \delta(\omega_2 - \omega_1) \end{aligned}$$

also

$$\langle dZ(\omega_1) dZ^*(\omega_2) \rangle = S(\omega_1) \delta(\omega_2 - \omega_1)$$

$dZ(\omega)$ muß orthogonal sein.

Drum heißt er auch "orthogonaler Inkrement Prozeß".

Inkrement-Prozess weil:

$$F(\omega) = \int_{-\pi}^{\omega} \langle dZ(\omega) dZ^*(\omega) \rangle$$

$$F(\omega + d\omega) = F(\omega) + \langle dZ(\omega) dZ^*(\omega) \rangle$$

Ad 2. Die ersten nichttrivialen Abhängigkeiten bestehen zwischen den Frequenzen $(\omega_1, \omega_2, \omega_3 = \omega_1 + \omega_2)$ und werden durch das Bispektrum (ω_1, ω_2)

$$BiSpe(\omega_1, \omega_2) = \sum c(\tau_1, \tau_2) e^{-i(\omega_1 \tau_1 + \omega_2 \tau_2)}$$

mit

$$c(\tau_1, \tau_2) = \langle x(t)x(t + \tau_1)x(t + \tau_2) \rangle_t$$

beschrieben.

Analog:

$$BiSpe(\omega_1, \omega_2) = \langle f(\omega_1)f(\omega_2)f^*(\omega_1 + \omega_2) \rangle$$

- Für lineare Prozesse ist das Bispektrum 0 (Gaußscher Fall) oder konstant (nichtGaußscher Fall). Für lineare Gaußsche Prozesse sind alle höheren Korrelationen der $dZ(\omega)$ s gleich 0. Dieses definiert lineare Prozesse. Superpositionsprinzip.
- Nichttriviales Bispektrum bei quadratischen Nichtlinearitäten. Weiteres, siehe unten.

Bemerkung:

Deterministische Oszillatoren haben perfekt korrelierte Fourierkomponenten (Fourierreihe). Anfangswerte werden nicht vergessen, gehören zum Prozeß, sie sind nicht stationär.

Satz:

Jeder nichtpathologische mischende (deterministische oder stochastische) Prozeß läßt sich durch

$$x(t) = \sum_{\omega=-\infty}^{\infty} e^{i\omega t} f(\omega)$$

durch spezifische Wahl der Korrelationen höherer Ordnung von $dZ(\omega)$ realisieren. Zu jedem Spektrum lässt sich also ein linearer Prozess finden mit dem gleichen Spektrum. Jeder Prozess ist als Summe von harmonischen Prozessen darstellbar, d.h. harmonisierbar.

Periodisch korrelierte Prozesse [133]

- Andere Bezeichnungen: Cyclostationär, periodisch stationär
- In der Rechnung zur Unabhängigkeit der Fourierkoeffizienten ging die vorausgesetzte Stationarität dadurch ein, daß die $ACF(t, \tau)$ als $ACF(\tau)$ geschrieben werden konnte.
- Betrachte periodisch korrelierte Prozesse, siehe Seite 35 mit

$$ACF(t, \tau) = ACF(t + T, \tau), \quad T = 2\pi/\Omega$$

- Beispiel:

Gibt es einen Jahresgang in der Dynamik des Wetters ? [26]

$$\langle f(\omega_1) f^*(\omega_2) \rangle = S(\omega_1, \omega_2) \delta(\omega_2 - \omega_1 + \frac{2\pi k}{T}) \quad S(\omega_1, \omega_2) \in \mathbb{C}$$

- Die periodische Zeitabhängigkeit der ACF bewirkt Korrelationen der Fourierkomponenten

$$(\omega_1, \omega_2), \text{ mit } \omega_1 = \omega_2 + \frac{2\pi k}{T}, \quad k \in \mathbb{Z},$$

also genau auf Nebendiagonalen (daher "periodisch korreliert").

5.1.1.6 Wiener-Khinchin Theorem

(Physiker Version)

- Erinnere:

$$S(\omega) = \sum_{\tau} e^{-i\omega\tau} ACF(\tau)$$

Gegeben $ACF(\tau)$ (eines stationären) Prozesses, wie sieht das Spektrum aus.

- Wann aber ist $ACF(\tau)$ die ACF eines stationären Prozesses?

Betrachte:

$$F(\omega) = \int_{-\pi}^{\omega} d\omega S(\omega) = \int_{-\pi}^{\omega} \langle dZ(\omega)dZ^*(\omega) \rangle$$

die spektrale Verteilungsfunktion.

- Wiener-Khinchin Theorem:

Wenn $F(\omega)$ beschränkt ist, dann ist

$$ACF(\tau) = \int_{-\pi}^{\pi} e^{i\omega\tau} dF(\omega) \left(= \int_{-\pi}^{\pi} e^{i\omega\tau} S(\omega) d\omega \right)$$

die ACF eines stationären Prozesses.

(Achtung: Steht in vielen Physiker-Büchern falsch)

Korollar:

Divergiert $S(\omega) \propto 1/\omega^\alpha$, $\alpha \geq 1$ ist der Prozess instationär.

5.1.1.7 Zerfallsverhalten von Spektren

[270, 269]

Nur für mischende Systeme interessant

5.1.1.7.1 Gewöhnliche Differentialgleichungen

Da mischend = chaotisch

- Ist $x(t)$ differenzierbar, seine FT $f(\omega)$, so ist die Fouriertrafo von $\dot{x} = i\omega f(\omega)$, da

$$\int dt e^{-i\omega t} \dot{x}(t) = i\omega \int dt e^{-i\omega t} x(t) = i\omega f(\omega)$$

- Da \dot{x} existiert, muß $f(\omega)$ schneller zerfallen als ω^{-1} damit die Inverse von $i\omega f(\omega)$ existiert.

- Für C^1 Funktionen muß das Spektrum also schneller als ω^{-2} zerfallen, für C^∞ schneller als jede Power von ω^{-1} .
- Betrachte

$$\int dt e^{-i\omega t} x(t)$$

Setze $x(t)$ analytisch ins Komplexe fort

Residuensatz:

$$= \sum_k C_k e^{i\sigma_k \omega} \omega^{\rho_k} e^{-\tau_k \omega}$$

mit C_k, ρ_k singularitäts-spezifische Konstante ρ üblicher Weise = -0.5 .
 σ_k, τ_k Real- und Imaginärteil der Singularität.

Für große ω trägt nur Singularität mit kleinstem τ_k bei.

(N.B.: $\tau \propto \lambda_{max}$, wobei λ_{max} der größte Lyapunov-Exponent ist.)

Also:

$$S_{det} \propto \omega^{2\rho} e^{-2\tau\omega}$$

i.w. exponentiell.

5.1.1.7.2 Stochastische Differentialgleichungen - Langevin Gleichungen

- Hat man:

$$\dot{x} = f(x) + g(x)\epsilon$$

so ist für große ω das Spektrum für \dot{x} flach, da $f(x)$ niederfrequente Dynamik produziert.

Bemerkung: Diesen Typ von Gleichungen bezeichnet man als Langevin Gleichung.

- Eine Integration gibt im Spektrum einen Faktor ω^{-2}
- Spektren stochastischer Systeme m -ter Ordnung zerfallen also mit

$$S_{stoch} \propto \omega^{-2m}$$

5.1.1.7.3 Beobachtungsrauschen Hat man Beobachtungsrauschen

$$\begin{aligned}\dot{x} &= f(x, \epsilon) \\ y &= x + \eta, \quad \eta \sim N(0, \sigma^2)\end{aligned}$$

so gilt

$$S(\omega) = \frac{\sigma^2}{2\pi}, \quad \text{für große } \omega$$

Macht obige Betrachtungen häufig nicht anwendbar.

5.1.1.7.4 Nichtlineare Oszillatoren

Nichtlineare Oszillatoren sehen i.w. so aus:

$$\ddot{x} = \mu f(x, \dot{x}) - \omega_0^2 x \quad ,$$

erinnere van der Pol:

$$\ddot{x} = \mu(1 - x^2)\dot{x} - \omega_0^2 x \quad .$$

- Naive Störungstheorie geht nicht durch, führt aber zum qualitativ richtigen Ergebnis :-)
- Beispiel van der Pol:
 $x_0 = \cos \omega_0 t$ eingesetzt, ergibt einen Term $\cos^2 \omega_0 t \sin \omega_0 t$ also i.w. $\cos^3 \omega_0 t$.
- Das heißt, die erste höhere Harmonische ist beim 3-fachen der Grundfrequenz, die weiteren beim $2n - 1$ -fachen.
- Beispiel Rössler:
 Bilineare Nichtlinearität: Höhere Harmonische bei allen Vielfachen der Grundfrequenz.
- Spezielle Art der Nichtlinearität sieht das Spektrum nicht.
- Formal richtige Störungstheoriemethoden [207]: Lindstedt-Poincaré, Harmonische Balance, Methode der Multiplen Skalen

5.1.1.8 Fractional Brownian Motion

[183, 252]

Betrachte Prozeß mit der Eigenschaft:

$$p(x(at_1), \dots, x(at_m)) \stackrel{d}{=} p(a^H x(t_1), \dots, a^H x(t_m)), \quad 0 < H < 1$$

Remarks :

- Änderung der zeitlichen Skala entspricht Änderung der räumlichen Skala: Selbstähnlich
- H nach Hurst [134]

•

$$"x(t) \stackrel{d}{=} t^H x(1)", \quad t > 1$$

•

$$x(0) = 0, \quad \text{wegen } x(0) = x(0t) = 0^H x(t) : -$$

•

$$\langle x^2(t) \rangle = t^{2H} \langle x^2(1) \rangle, \quad 0 < H < 1 \text{ darum: FBM, BM: } H=0.5$$

- Ihre ACF lautet mit Hilfe der Grafen von Binomi

$$\begin{aligned} ACF_{FBM}(t_1, t_2) &= \langle x(t_1)x(t_2) \rangle \\ &= \frac{1}{2} \left(\langle x^2(t_1) \rangle + \langle x^2(t_2) \rangle - \langle (x(t_1) - x(t_2))^2 \rangle \right) \\ &= \frac{1}{2} \left(\langle x^2(t_1) \rangle + \langle x^2(t_2) \rangle - \langle x^2(t_1 - t_2) \rangle \right) \\ &= \frac{1}{2} \left(t_1^{2H} + t_2^{2H} - |t_1 - t_2|^{2H} \right) \langle x^2(1) \rangle \end{aligned}$$

Betrachte Inkremente:

$$y(t) = x(t+1) - x(t), \quad y(0) = x(1)$$

sind stationär, heißen Fractional Gaussian Noise

$$\begin{aligned}
ACF_{FGN}(\tau) &= \langle y(0)y(\tau) \rangle \\
&= \langle x(1)(x(\tau+1) - x(\tau)) \rangle \\
&= ACF_{FBM}(1, \tau+1) - ACF_{FBM}(1, \tau) \\
&= \frac{1}{2} \left((\tau+1)^{2H} - 2\tau^{2H} + (\tau-1)^{2H} \right)
\end{aligned}$$

$$\begin{aligned}
0 < H < \frac{1}{2} & \quad ACF_{FGN}(\tau) < 0 \text{ für } \tau \neq 0 \\
\frac{1}{2} < H < 1 & \quad ACF_{FGN}(\tau) > 0 \\
H = \frac{1}{2} & \quad WN
\end{aligned}$$

Es gilt:

$$\begin{aligned}
ACF_{FGN}(\tau) &\propto \left((\tau+1)^{2H} - 2\tau^{2H} + (\tau-1)^{2H} \right) \\
&\approx \Delta\tau^{2H} \\
&= 2H(2H-1)\tau^{(2H-2)}
\end{aligned}$$

$ACF_{FGN}(\tau)$ zerfällt nur algebraisch, also nur mit einer Potenz und nicht exponentiell wie z.B. AR[p] Prozesse.

$0 < H < \frac{1}{2}$:

•

$$\sum_{\tau=-\infty}^{\infty} |ACF_{FGN}(\tau)| < \infty, \quad \sum ACF_{FGN}(\tau) = 0$$

letzteres heißt:

$$ACF_{FGN}(0) = -2 \sum_{\tau=1}^{\infty} ACF_{FGN}(\tau)$$

Harte unnatürliche Bedingung, kleinste Änderungen ändern Struktur

- Heißt "negative dependence".

Wichtiger: $\frac{1}{2} < H < 1$:

- $ACF_{FGN}(\tau) \rightarrow 0$ für $\tau \rightarrow \infty$, aber

$$\sum_{\tau=-\infty}^{\infty} ACF_{FGN}(\tau) = \infty$$

Heißt: "Long range dependence" im Gegensatz zu den handelsüblich exponentiell zerfallenden ACFs.

Bei short range dependent:

$$\sum_{\tau=-\infty}^{\infty} ACF_{srd}(\tau) < \infty$$

Diskussion: Lange exponentielle Zerfallszeiten vs. long range dependent

- Auswirkung: Zentraler Grenzwertsatz gilt nicht mehr. Man muß

$$\frac{1}{N^H} \sum_{i=1}^N y(i)$$

betrachten, um einen vernünftigen Limes zu bekommen.

Aus ACF folgt Spektrum:

$$\begin{aligned} S(\omega) &\propto \int \tau^{2H-2} e^{-i\tau\omega} d\tau, \quad x = \tau\omega, \quad \tau = \frac{x}{\omega}, \quad d\tau = \frac{1}{\omega} dx \\ &= \int \left(\frac{x}{\omega}\right)^{2H-2} e^{-ix} \frac{1}{\omega} dx \\ &= \omega^{1-2H} \int x^{2H-2} e^{-ix} dx \\ &= \text{const } \omega^{1-2H} \end{aligned}$$

ZEICHNUNG Spektren $0 < H < 0.5$, $0.5 < H < 1$

$0 < H < 0.5$ muß nach 0 gehen wg. $S(0) = \sum ACF(\tau)$

$0.5 < H < 1$: Spektrum divergiert (integrable, Wiener-Khinchin) für $\omega \rightarrow 0$

- \implies Es gibt beliebig lange Perioden = Joseph-Effekt: 7 gute, 7 schlechte Jahre [184]
- Nimmt man statt Gaußverteilten ZV, welche mit fat tails folgt: Noah-Effekt: Extreme Ereignisse.

Für Genießer: FGN Fixpunkt eines Renormalisierungsgruppenspielchens

Fürs inverse Problem, see [185]

Andere Formulierung

5.1.1.8.1 ARIMA und FARIMA Prozesse

- ARIMA: AR integrated MA Prozesse

$$\phi(B)(1 - B)x(t) = \theta(B)\epsilon(t)$$

Einfacher Fall: Brownian Motion

$$(1 - B)x(t) = \epsilon(t) \implies x(t) = x(t - 1) + \epsilon(t)$$

Bemerkung: Integration stammt aus Abhängigkeit von $x(t)$ von $x(t - 1)$ mit Vorfaktor $1 -$ Prozess sammelt also gesamte Vergangenheit auf.

- FARIMA (Fractional ARIMA), $d > -1$, $d \in R$

$$\phi(B)(1 - B)^d x(t) = \theta(B)\epsilon(t)$$

$$\nabla^d = (1 - B)^d = \sum_{i=0}^{\infty} \pi_i B^i$$

- Wenn $d \in (-0.5, 0.5)$, dann äquivalent zu FGN mit $d = H - \frac{1}{2}$
- Stellen ein explizites (lineares) Beispiel für FGN dar
- Vorteil der Formulierung: Man kann noch kurzreichweitige ARMA-Prozesse einbauen.

5.1.1.8.2 1/f Rauschen (Nichtbeinhalteter) Grenzfall zwischen FGN und FBM

Integral des Spektrums divergiert für $\omega \rightarrow 0$ und $\omega \rightarrow \infty$, aber nur logarithmisch, d.h. man kann cut-offs beliebig weit rausschieben.

Frequently claimed:

- Strom durch Widerstände [144, 259]
- Ozeanische Strömungen
- Lautstärke eines Radiosenders
- Frequenz von Quarz-Kristall-Oszillatoren
- Transistoren
- Erdbeben aktivität
- traffic flow
- Conference "Unsolved problems in noise"

Reviews: [321] [156]

ACF-Zerfall: logarithmisch, irre langsam.

Zwei physikalische Modelle

- Self organized criticality SOC [14]

System mit instabilem Fixpunkt, zu dem es langsam wieder hinläuft

Beispiel: Plattentektonik & Erdbeben, Argument dagegen PRL Jun-Aug 04

2D-Sandpile Modell:

$z(x, y) \in N$, K eine Schwelle, fixed, entspricht Steigung des Sandhaufens.

Wenn $z(\cdot) > K$, dann:

$$\begin{aligned} z(x, y) &\rightarrow z(x, y) - 4 \\ z(x \pm 1, y) &\rightarrow z(x \pm 1, y) + 1 \\ z(x, y \pm 1) &\rightarrow z(x, y \pm 1) + 1 \end{aligned}$$

”Riesele Sand drauf” und schaue, was vom Teller fällt.

Übung: Darstellung der 1 ?

FOLIE aus Bak paper

1/f bei:

- Clustergrösse
- Inter-Lawinen Abstand
- Shotnoise Prozeß, exponentieller Shot, Zerfallszeit AR modulierte [154, 155, 253]
- Überlagerung von AR[1] Prozessen mit verschiedenen Zeitskalen. Kandidat für angebliches 1/f Rauschen im Herzschlag [110] durch Einfluss von Herzschlag, Atmen, Tag-Nacht-Rhythmus.
- Andere Möglichkeiten [233]
- Kritik: [330] Siehe auch comments on ”Do Earthquakes Exhibit Self-Organized Criticality?” R. Woodard et al. Phys. Rev. Lett. 93, 249801 (2004)
- Siehe Kritik-paper zu El Nino Analyse [194]
- ”Programmierfehler” in originalpaper, siehe [100] und Ref. therein

Bemerkung: $1/f^\alpha$ heißt auch einfach $1/f$ noise.

Merke:

”Spektralanalyse = lineare Analyse”

Da das Spektrum die Fouriertransformierte der Autokovarianzfunktion ist und die ACF die Parameter eines linearen Prozesses eindeutig festlegt, ist die Spektralanalyse eine ”lineare Analyse”, deren Aussagekraft allerdings nicht auf lineare Prozesse beschränkt ist.

$S(\omega) \iff ACF(\tau) \iff$ Linearer Prozess (Struktur und Parameter)

$S(\omega) \iff ACF(\tau) \Leftarrow$ Nichtlinearer Prozess (Struktur und Parameter)

Für nichtlineare Prozesse: $ACFs(\tau_1, \dots, \tau_m)$ und Spektren $S(\omega_1, \dots, \omega_m)$ höherer Ordnung relevant.

Aber:

Was man alles im Spektrum sieht:

- Oszillationen
- Oszillator oder Relaxator(en)
- Höhere Harmonische bei nichtlinearen Schwingern und Grad der Nichtlinearität
- Varianz des Beobachtungsrauschen im Hochfrequenten
- Stochastisch vs. chaotisch im Hochfrequenten [270, 269]
- $1/f^\alpha$ Rauschen
- Instationaritäten (Beispiel: Run down von Ionenkanal)

Nur für wenige nichtlineare Prozesse läßt sich das Spektrum analytisch berechnen.

Übung:

Berechne das Spektrum der logistischen Abbildung bei $r = 4$

$$x(i+1) = 4x(i)(1-x(i))$$

5.1.2 Schätzung im Endlichen

5.1.2.1 Nyquist Frequenz

- Die höchste auflösbare Frequenz eines mit Δt gesampelten (abgetasteten) Prozesses ist

$$\omega_{Ny} = \frac{\pi}{\Delta t}, \text{ oder } f_{Ny} = \frac{1}{2\Delta t}$$

ZEICHNUNG

- Für eine Zeitreihe $x(t)$ mit N Datenpunkte und einer Samplingrate von $1/\Delta t$ ist eine Auswertung der Fouriertrafo

$$f(\omega) = \frac{1}{\sqrt{N}} \sum_{k=1}^N e^{-i\omega k \Delta t} x(k \Delta t)$$

sinnvoll an den natürlichen Fourierfrequenzen ω_k :

$$\omega_k = \frac{2\pi k}{N\Delta t} = \frac{2\pi}{T} k, \quad k = -\frac{N}{2}, \dots, 0, \frac{N}{2},$$

falls $N\Delta t = T$ die zeitliche Länge der Zeitreihe ist.

- Da $x(t)$ reell, gilt

$$f(\omega) = f^*(-\omega)$$

und die Information der positiven Frequenzen reicht (Faktor 2 zur Varianzerhaltung)

- Erhaltung der Freiheitsgrade
- Durch Abzählen folgt: Alle $f(\omega)$, $\omega \neq \omega_k$ sind durch Linearkombinationen der $f(\omega_k)$ darzustellen.

5.1.2.2 Zeitkontinuierliche vs. zeitdiskrete Prozesse

Vom zeitkontinuierlichem Prozeß zur gesampten Zeitreihe:

Das Aliasing-Problem:

- Befindet sich im zeitkontinuierlichen Prozess Power oberhalb der Nyquist Frequenz, so wird diese im Spektrum des gesampten Prozesses an der Nyquistfrequenz "gespiegelt" unterhalb der Nyquistfrequenz auftauchen.

ZEICHNUNG ALIASING im Zeitraum, Unterschied sin und cos bei NyquistFrequenz

FOLIE Recipes

- Kennt man von drehenden Rädern
- Darum: Analog Tiefpaßfiltern **vor** Sampling.

5.1.2.3 Wachstumsverhalten harmonischer peaks

Oder: Unterschied Fourierreihen und FT(Mischender Prozeß)

Bei "richtiger" Normierung:

$$\sum_{k=1}^{N/2} 2\pi S(\omega_k) = \sum_{i=1}^N x^2(i)$$

- Periodische Prozesse: $S(\omega) \neq 0$ nur für endlich viele ω
- Mischende Prozesse: $S(\omega) \neq 0 \forall \omega$

Daraus folgt:

- Periodische Prozesse: $S(\omega) \propto N$
- Mischende Prozesse: $S(\omega) = \text{const}$

Statistikeinschub

- Schätzer $\hat{\Theta}_N$ heißt erwartungstreu, unverzerrt oder unbiased, wenn gilt

$$\langle \hat{\Theta}_N \rangle = \Theta$$

- Mitunter gilt dies nur asymptotisch, dann asymptotisch erwartungstreu, asymptotisch unverzerrt, asymptotisch unbiased.
- Ein Schätzer $\hat{\Theta}$ heißt konsistent, wenn gilt

$$\lim_{N \rightarrow \infty} (\hat{\Theta}_N - \Theta) \xrightarrow{prob} 0$$

i.e. Bias und Varianz gehen mit N gegen 0.

5.1.2.4 Spektralschätzer

Es gibt at least 5 verschiedene Typen von Spektralschätzern.

Der hier Empfohlene beruht auf der Glättung des Periodogramms:

5.1.2.4.1 Das Periodogramm

- Betrachte zuerst ARMA-Prozesse. Deren Spektrum:

$$S_{ARMA}(\omega_k) = \frac{|\Theta(e^{i\omega_k})|^2}{|\phi(e^{i\omega_k})|^2} S_{WN}(\omega_k)$$

I.e. betrachte $S_{WN}(\omega_k)$

- Die Fouriertrafo von WN war:

$$f(\omega_k) = \sum e^{-i\omega_k t} \epsilon(t) = \frac{1}{2\sqrt{\pi}} \eta(\omega_k), \quad \eta(\omega_k) \sim N_C(0, \sigma^2)$$

- Spektrum war

$$S(\omega_k) = \langle |f(\omega_k)|^2 \rangle$$

- $|f(\omega_k)|^2$ hat extra Namen: Periodogramm

$$Per(\omega_k) = |f(\omega_k)|^2$$

- Hoffnung: Periodogramm ist ein guter Schätzer für das Spektrum
- Da

$$Per(\omega_k) = |f(\omega_k)|^2 = (\operatorname{Re}(f(\omega_k)))^2 + (\operatorname{Im}(f(\omega_k)))^2$$

gilt für $x(t) = \epsilon(t)$:

$$Per_{WN}(\omega_k) \sim \chi_2^2$$

Bemerkung: Das Periodogramm ist also frequenzweise Chi-Quadrat-verteilt mit zwei Freiheitsgeraden. Dies ist übrigens eine Exponentialverteilung mit Parameter 2.

- Für nichtweiße (i.a. nichtlineare) Prozesse hilft der Zentrale Grenzwertsatz, und es gilt allgemein :

$$Per(\omega_k) \sim \frac{1}{2} S(\omega_k) \chi_2^2, \quad \omega_k \neq 0, \pi$$

unabhängig von N . (Für $\omega_k = 0, \pi$: $Per(\omega_k) \sim S(\omega_k) \chi_1^2$, da nur $\cos(\omega_k t)$ beiträgt).

- Da

$$\langle \chi_2^2 \rangle = 2, \quad \text{Var}(\chi_2^2) = 4 \quad \text{SD}(\chi_2^2) = 2$$

ist die Standardabweichung des Periodogramms unabhängig von N

Damit:

Das Periodogramm ist kein konsistenter Schätzer für das Spektrum !

- Statt kleiner werdender Varianz des Schätzers erhält man bessere Auflösung im Frequenzraum, i.e. mehr natürliche Fourier-Frequenzen.

Leakage und Tapern:

Noch schlimmer:

- Die FT sieht die Zeitreihe $x(t), t = 1, \dots, T$ als Ausschnitt einer unendlich langen Reihe $y(t), t \in \mathbb{Z}$:

$$x(t) = w_u(t)y(t), \quad w_u(t) = \begin{cases} 1 & \text{if } 1 \leq t \leq T \\ 0 & \text{sonst} \end{cases}$$

- Da Multiplikation im Zeitraum eine Faltung im Frequenzraum darstellt, ist Spektralschätzung "verschmiert", vor allem wird Power von Peaks in Täler transportiert :

$$f_x(\omega_k) = \sum_i f_y(\omega_{k+i}) f_w(\omega_i)$$

- **Periodogramm ist (sogar) ein verzerrter Schätzer.**
- Diesen Effekt nennt man wortmalerisch Leakage, er ist für $w_u(t)$ maximal schlimm.
- "Lösung": Wähle $w(t)$, das sanfter verläuft, z.B. Bartlett window (Dreiecksfenster)

$$w_B(t) = \begin{cases} 1 - \left| \frac{t - \frac{1}{2}T}{\frac{1}{2}T} \right| & \text{if } 1 \leq t \leq T \\ 0 & \text{sonst} \end{cases}$$

- Dieses nennt man Tapern.

BILDER Recipes p.547f

- Das resultierende Periodogramm ist mit

$$g = \frac{T}{\sum_{t=1}^T w^2(t)}$$

zu normieren.

Einschub für's Gemüt:

- Warum sehen wir Schwingungen beim AR[1] Prozeß ?
- Weil Spektrum im niederfrequenten größte Power hat ...
- ... und die χ_2^2 -Verteilung dort irgendwo einen größten Wert annimmt.

ZEICHNUNG dazu

5.1.2.4.2 Schätzung des Spektrums durch Glättung des Periodogramms

Zentral:

Weil das (wahre) Spektrum glatt, kann man Spektren schätzen, in dem man das Periodogramm glättet mit Kern k_j :

$$\hat{S}(\omega_i) = \sum_{j=-h}^h k_j Per(\omega_{i+j})$$

Dieses ergibt mit $N \rightarrow \infty$, $h \rightarrow \infty$, und $h/N \rightarrow 0$ einen konsistenten Schätzer.

BILDER Folienfilm

- Die Anzahl der Freiheitsgrade ist :

$$\nu = \frac{2}{\sum_{i=-h}^h k_i^2}$$

- Beachte: Durch die Multiplikation mit dem Taper-Fenster werden benachbarte Frequency bins korreliert.
- Das ändert die effektive Anzahl der Freiheitsgrade weiter:

$$\nu = \frac{q_2^2}{q_4} \frac{2}{\sum_{i=-h}^h k_i^2}, \quad q_2 = \sum_i w_i^2, \quad q_4 = \sum_i w_i^4$$

- Für Statistik bei getaperten Daten, siehe [317, 59]

5.1.2.5 Alternative Spektralschätzer:

5.1.2.5.1 (i) ACF Fourier transformieren

Greife zurück auf Mathematiker-Definition:

$$\hat{S}(\omega) = \sum_{\tau} e^{-i\omega\tau} \widehat{ACF}(\tau)$$

Schätzer der $ACF(\tau)$

$$\frac{1}{N} \text{ oder } \frac{1}{N - \tau} ?$$

- Naheliegender ist es (Annahme $\langle x \rangle = 0$)

$$\widehat{ACF}(\tau) = \frac{1}{N - \tau} \sum_{t=1}^{N-\tau} x(t)x(t + \tau)$$

zu definieren.

- Dieser Schätzer ist unverzerrt, aber hat für große τ eine große Varianz.

- Nun gilt $\forall \phi$ -mischenden Prozesse

$$\lim_{\tau \rightarrow \infty} \widehat{ACF}(\tau) = 0$$

Der Schätzer:

$$\widehat{ACF}(\tau) = \frac{1}{N} \sum_{t=1}^{N-\tau} x(t)x(t+\tau)$$

hat eine kleinere Varianz, ist verzerrt, aber in die "richtige" Richtung, da ACF für $\tau \rightarrow$ groß gegen 0 geht.

- Vor allem ist er positiv definit, was der erste nicht ist. D.h. Fouriertransformation vom ersten kann negative Werte ergeben, was wenig sinnvoll ist.
- M.K.z.: $\widehat{ACF}(\tau)$ ist konsistenter Schätzer.
- Aber die geschätzte ACF hat (schon für lineare Prozesse) eine komplizierte Korrelationsstruktur ihrer Schätzfehler:

$$\begin{aligned} & \langle (\widehat{ACF}(\tau_1) - ACF(\tau_1))(\widehat{ACF}(\tau_2) - ACF(\tau_2)) \rangle = \\ & \frac{1}{N} \sum_{k=1} \{ ACF(k+\tau_1) + ACF(k-\tau_1) - 2ACF(\tau_1)ACF(k) \} \\ & \quad \{ ACF(k+\tau_2) + ACF(k-\tau_2) - 2ACF(\tau_2)ACF(k) \} \end{aligned}$$

ZEICHNUNG $ACF(\tau)$ und $\widehat{ACF}(\tau)$ von AR[2]

95 % Konfidenzintervall für WN

$$\pm 1.96 \frac{1}{\sqrt{N}} \sigma^2$$

Übung:

ACF-Schätzung

Berechne die Autokorrelationsfunktion für einen AR[2] Prozeß mit $T = 20$ und $\tau = 100$ bis zum lag 500. Was fällt auf ?

Spektralschätzung:

- Fourier-Trafo der empirischen ACF ergibt das (χ_2^2 zappelnde) Periodogramm:

$$\hat{S}(\omega_k) = \sum_{\tau=0}^M e^{-i\omega_k \tau} \widehat{ACF}(\tau) \equiv Per(\omega_k) \quad ,$$

obwohl die geschätzte ACF eine (relativ) glatte Funktion ist.

- Erklärung: Der Punktweise Schätzfehler ist $O(1/N)$, aufsummiert gibt das $NO(1/N) = O(1)$ Schwankungen im Fourierraum.
- Schätzen durch Abschneiden:

- Wählt man ein Fenster $W(\tau)$, das für $\tau \rightarrow \infty$ schnell genug gegen 0 geht, so ergibt:

$$\hat{S}(\omega) = \sum_{\tau=0}^M e^{-i\omega \tau} W(\tau) \widehat{ACF}(\tau) \equiv \sum_{j=-h}^h k_j Per(\omega_{k+j})$$

einen konsistenten Schätzer

- Hat $W(\tau)$ endlichen Träger $[0, M]$, $M \ll N$, hat, so muß die $\widehat{ACF}(\tau)$ nur bis lag M ausgerechnet werden. Gesamtrechnaufwand $O(MN)$, statt $O(N^2)$ wie bei Periodogrammansatz mit naiver FT.

Methode der Wahl vor Erfindung der Fast-Fourier-Trafo ($O(N \log N)$).

- Glättung des Periodogramms entspricht Multiplikation der ACF.
- $W(\tau)$ und k_j hängen über FT zusammen.
Vorteil der k_j -Formulierung: Man sieht direkt, was man tut.

5.1.2.5.2 (ii) Periodogramme mitteln

- Unterteile Zeitreihe in L Stücke
- Mittlere deren Periodogramme

Sei $M = N/L$

$$Per_l(\omega_k) = \left| \sum_{t=1}^M e^{-i\omega_k t} x((l-1)M + t) \right|^2$$

$$\hat{S}(\omega_k) = \frac{1}{L} \sum_{l=1}^L Per_l(\omega_k)$$

- Äquivalent dazu, das Periodogramm der Gesamtreihe mit einem Rechteckfenster der Breite $2h + 1 = L$ zu glätten.
- Diskussion der Frequenzauflösung.

Ausflug: Spektralanalyse chaotischer Prozesse

5.1.2.5.3 (iii) Multi tapering [282]

Das Tapern mit einer Funktion wie

$$w(i) = 1 - \left| \frac{i - \frac{1}{2}N}{\frac{1}{2}N} \right|$$

vernichtet Information der Randpunkte, die an sich nicht schlechter sind als die Punkte in der Mitte der Zeitreihe.

Die obige Strategie:

- Tapern
- FT
- Periodogramm
- Schätzen durch glätten

kann deswegen ersetzt werden durch:

- Tapere die Zeitreihen mit orthogonalen Funktionen $v_j(i)$, ergibt $x_j(i)$.
- $FT_j(\omega_k)$ der einzelnen Ergebnisse (sind unabhängig voneinander)
- Periodogramme der einzelnen FTs
- Schätzen durch Mitteln der $Per_j(\omega_k)$

Optimale Taper hierfür

- U_j : Discrete prolate spheroidal wave functions

FOLIE Thomson Taper

- Aber entsprechende Sinüsse gehen auch.

Erhaltung des grundsätzlichen Problems:

- Varianzreduktion vs. Leakage

Schöne Darstellung:

<http://www.ms.washington.edu/~s520/chpt-07/chpt-07.ps>

5.1.2.5.4 (iv) AR-Spektralschätzung

- Fitte AR[p] Prozeß an die Daten, siehe Kap. 6.2.
- Rechne zugehöriges Spektrum aus.
- Vorteil: Kann beliebig nah benachbarte peaks auflösen, was Periodogramm-Glätter nicht kann.
- Hype der 70er Jahre [47].
- Problem:
Was passiert wenn AR Modell nicht paßt ? Nichtlinearität, Beobachtungsrauschen, ...
- "Lösung" :
Überparameterisierung. Schafft die Probleme, die man eigentlich lösen wollte.
- Für Ordnung $p =$ Anzahl der Daten N ergibt das AR[p]-Spektrum das Periodogramm
- mit $N \rightarrow \infty, p \rightarrow \infty, p/N \rightarrow 0$ gibt das konsistenten Spektralschätzer

Synpose des Erhaltungssatzes des Problem:

- Periodogramm glätten: $N \rightarrow \infty, h \rightarrow \infty, h/N \rightarrow 0$

- ACF fenstern: $N \rightarrow \infty, M \rightarrow \infty, M/N \rightarrow 0$
- Periodogramme mitteln: $N \rightarrow \infty, L \rightarrow \infty, L/N \rightarrow 0$
- AR-Prozesse anpassen: $N \rightarrow \infty, p \rightarrow \infty, p/N \rightarrow 0$

Kein Entkommen, No Free Lunch !

5.1.2.5.5 (v) Whittle Schätzer [325]

Statistikeinschub:

Maximum Likelihood Schätzer:

- Hat man ein parametrisiertes Modell:

$$y_i = y(x_i, a) + \epsilon_i$$

induziert dies eine Wahrscheinlichkeit $p(y_i, x_i, a)$.

- Diese Wahrscheinlichkeit in Abhängigkeit der Parameter heißt Likelihood. Im hiesigen Falle:

$$L(y_1, \dots, y_N, x_1, \dots, x_N | a) \propto \prod_{i=1}^N p(y_i, x_i | a)$$

da y_i nur von x_i und ϵ_i abhängt. Die Likelihood ist formal keine Wahrscheinlichkeit, e.g. keine Normierung.

Die absoluten Werte haben keine Bedeutung. Relative Werte können zur Modellselektion, siehe Kap. 8 dienen.

- Im Falle Gaußverteilter Fehler:

$$L(y_1, \dots, y_N | a) \propto \prod_{i=1}^N \left\{ \exp \left[-\frac{1}{2} \left(\frac{y_i - y(x_i, a)}{\sigma_i} \right)^2 \right] \Delta y \right\} \quad (\dagger)$$

Maximierung von (\dagger) entspricht Minimierung des negativen Logarithmus von (\dagger) :

$$\mathcal{L} = \left[\sum_{i=1}^N \frac{[y_i - y(x_i, a)]^2}{2\sigma_i^2} \right] - N \log \Delta y$$

ergibt χ^2 Fitting.

- Hat man eine parametrisierte Form des Spektrums in Abhängigkeit von Parametern des Prozesses, wie z.B. bei FARIMA Prozessen, ergibt sich

$$L(\text{Per}(\omega_0), \dots, \text{Per}(\omega_{N/2}); a) = \prod_{k=0}^{N/2} \frac{1}{S(\omega_k, a)} e^{-\text{Per}(\omega_k)/S(\omega_k, a)}$$

Bemerkung: Eine χ^2 -Verteilung mit zwei Freiheitsgraden entspricht einer Exponentialverteilung mit Parameter $\tau = 2$: $\chi_2^2 \sim \frac{1}{2}e^{-x/2}$. Für die Verteilung des Periodogramms an der Frequenz ω gilt: $\text{Per}(\omega) \sim \frac{1}{2}\chi_2^2 S(\omega)$

Mit $\mathcal{L} = -\log(L)$

$$\mathcal{L}(a) = \sum_{k=0}^{N/2} \log(S(\omega_k, a)) + \text{Per}(\omega_k)/S(\omega_k, a) \quad (5.9)$$

ergibt sich der Whittle-Schätzer.

Vorteil: Der Einfluß von Beobachtungsrauschen läßt sich durch eine additive Konstante in $S(\omega, a)$ leicht berücksichtigen.

Vorteil der empfohlenen Methode: Man kann datengetrieben die Glättungsbreite wählen, siehe [39, 289].

Besprechung von [289]

FOLIE FIG. 2, 3, 5

FOLIE QAT Fig. 1 Spektren immer logarithmisch plotten

5.1.2.6 Vertrauensintervalle fürs Spektrum

Erinnere: Anzahl der Freiheitsgrade des Spektralschätzers (modulo Tapereffekt):

$$\nu = \frac{2}{\sum_{i=-h}^h w_i^2}$$

Da χ^2 Verteilungen additiv sind, ist α Konfidenzintervall für $S(\omega_k)$:

$$\left(\frac{\nu \hat{S}(\omega_k)}{\chi_\nu^2(\alpha/2)}, \frac{\nu \hat{S}(\omega_k)}{\chi_\nu^2(1 - \alpha/2)} \right)$$

Für große ν geht das gegen Gaußsche (i.e. symmetrische) Intervalle.

QAT FIG. 5 again (Quantitative Analyse von Tremorzeitreihen)

5.1.2.7 Lomb Spektrum

[226]

Alles bisherige setze voraus, daß die Daten äquidistant gesampelt sind. Im Falle nichtäquidistanter Daten:

- verliert die Nyquist-Frequenz ihren eigentlichen Sinn
- sollte man nicht interpolieren
- den Algorithmus von Lomb verwenden:

$$Per(\omega) \propto \left(\frac{[\sum_i x(t_i) \cos(\omega(t_i - \tau))]^2}{\sum_i \cos^2(\omega(t_i - \tau))} + \frac{[\sum_i x(t_i) \sin(\omega(t_i - \tau))]^2}{\sum_i \sin^2(\omega(t_i - \tau))} \right)$$

mit

$$\tan(2\omega\tau) = \frac{\sum_i \sin(2\omega t_i)}{\sum_i \cos(2\omega t_i)}$$

- Ist so, als würde man $a \cos(\omega t) + b \sin(\omega t)$ an die Daten fitten und schauen, wie hoch der Anteil für ein bestimmtes ω ist.
- Es gibt Fast Lomb Algorithmus $O(N \log N)$.

5.1.2.8 Test auf white noise

Statistikeinschub:

Test auf Verträglichkeit einer empirischen mit einer theoretischen Verteilung
Kumulative Verteilung $cum(x)$

$$cum(x) = \int_{-\infty}^x p(x') dx'$$

Empirische kumulative Verteilung $cum_{emp}(x)$ basierend auf N Daten x_i :
Sortiere die x_i aufsteigend

$$cum_{emp}(x) = \frac{1}{N} (\text{Anzahl der } x_i \text{ mit } , x_i < x)$$

Kolmogorov-Smirnov Test basiert auf Testgröße

$$D = \max_x |cum(x) - cum_{emp}(x)|$$

Verwandte Tests:

- Cramer von Mises: Testgröße CvM mit

$$CvM = \int w(x)(cum(x) - cum_{emp}(x))^2$$

- deWet-Wenter: Testgröße dWW mit

$$dWW = \max_y |x_1 - x_2|$$

wobei $x_1 = cum^{-1}(y)$ und $x_2 = cum_{emp}^{-1}(y)$.

Vorteil KS-Test: Verteilung unter der Nullhypothese hängt nicht von der
wahren Verteilung $p(x)$ ab.

Ende Statistikeinschub

Anwendung aufs Periodogramm:

- Bilde das normierte kumulative Periodogramm

$$CP(\omega_r) = \frac{\sum_{k=1}^r Per(\omega_k)}{\sum_{k=1}^{N/2} Per(\omega_k)} \quad \text{with} \quad r = 1, 2, \dots, N/2$$

- Vergleiche dies mit der Diagonale des weißen Rauschens.
- Beachte:
 - Das normierte kumulative Periodogramm (oder auch Spektrum) erfüllt die formalen Eigenschaften einer Verteilungsfunktion.
 - Aber nicht die Verteilungseigenschaften der Realisierung: χ_2^2 vs. Poisson-Verteilung.
 - Es geht trotzdem gut. CLT

FOLIE QAT FIG. 4

5.1.2.9 Test auf harmonische Komponenten

Gesucht: Harmonische Komponente in weißem Rauschen [260].

Wird viel in der Astrophysik gemacht

- Perioden von Doppelsternsystemen
- Rotationsfrequenz Schwarzer Löcher

3 FOLIEN eps+amp*sin, Perio

- Weißes Rauschen der Varianz σ^2 :

$$Per(\omega_k)/\sigma^2 \sim \chi_2^2 \sim \frac{1}{2}e^{-x/2}$$

bzw. für die kumulative Verteilung

$$p[Per(\omega_k)/\sigma^2 \leq z] = \int_0^z dx \frac{1}{2}e^{-x/2} = 1 - e^{-z/2}$$

- Das gibt α Quantil für einzelne Frequenz.
- Aber Dilemma IV: Multiples testen, da man jeden Frequenzbin anschaut.

- Bonferroni Korrektur

$$\gamma = \left\{ \max_{1 \leq k \leq N/2} (Per(\omega_k)) \right\} / \sigma^2$$

$$\begin{aligned} p[\gamma > z] &= 1 - p[\gamma \leq z] \\ &= 1 - p[(Per(\omega_k)/\sigma^2 \leq z, \forall k)] \\ &= 1 - (1 - e^{-z/2})^{N/2} \end{aligned}$$

- Nun ist σ^2 i.a. nicht bekannt.
- Unter H_0 kann man es aus den Daten schätzen. Der Schätzer der dies berücksichtigt, Fischers g -Statistik [76] sieht im wesentlichen aus wie der ebige.

Harmonische Komponente im allgemeinen. [309]

- Verwende einen (gegenüber Peaks) robusten Spektralschätzer mit der impliziten Darstellung

$$0 = \sum_{i=-h}^h w_i \Psi \left(\frac{Per(\omega_{k+i})}{\hat{S}(\omega_k)} - 1 \right) \quad \forall k$$

mit z.B.

$$\Psi(x) = \begin{cases} x & \text{if } x < c \\ 0 & \text{if } x \geq c \end{cases}$$

Motivation:

$$\begin{aligned} \hat{S}(\omega_k) &= \sum_{i=-h}^h w_i Per(\omega_{k+i}) \\ 1 &= \sum_{i=-h}^h w_i \frac{Per(\omega_{k+i})}{\hat{S}(\omega_k)} \\ 0 &= \sum_{i=-h}^h w_i \left(\frac{Per(\omega_{k+i})}{\hat{S}(\omega_k)} - 1 \right), \quad \text{wg. } \sum_{i=-h}^h w_i = 1 \end{aligned}$$

- Teile $Per(\omega_k)$ durch $\hat{S}(\omega_k)$
- Verwende Ergebnis von oben

5.1.2.10 Konfidenzintervall für die Peakfrequenz

Für (nichtlineare) Oszillationen ist die Peakfrequenz eine charakteristische Größe. Gesucht: Konfidenzintervall für sie.

5.1.2.10.1 Statistikeinschub: Bootstrap Konfidenzintervalle

Bootstrap=An den eigenen Stiefeln aus dem Sumpf ziehen

Literatur:

- B. Efron and R.J. Tibshirani: An Introduction to the Bootstrap [73]
- E. Mammen: When does bootstrap work? : asymptotic results and simulations [182]
- A.C. Davison and D.V. Hinkley: Bootstrap Methods and their Application [65].

Monte Carlo ((Re-)sampling) Methode zur Ermittlung von Konfidenzintervallen und Tests, wenn keine asymptotischen Methoden zur Verfügung stehen.

Beispiel: Standardfehler des Mittelwertes der Gaußverteilung (ist σ/\sqrt{N}).

- Wenn man M Sample der Länge N hätte.
- Bestimme

$$\mu_j = \frac{1}{N} \sum_{i=1}^N x_{ij}, \quad j = 1, \dots, M$$

- Die μ_j sind Realisierungen der ZV : $\hat{\mu} = \frac{1}{N} \sum_{i=1}^N X_i$.
- Betrachte ihre Verteilung. Deren $\alpha/2$ und $1 - \alpha/2$ Quantile sind gerade die Konfidenzintervalle.

- Durchführung:
 - Sortiere μ_j
 - $\mu_{\text{int}_{\text{floor}}(M\alpha/2)}$ gibt $\alpha/2$ Quantil
 - $\mu_{\text{int}_{\text{ceil}}(M(1-\alpha/2))}$ gibt $1 - \alpha/2$ Quantil
- Problem: Man hat keine M Sample :-(
- Idee: Jedes x_i liefert einen Beitrag $1/N$ zur empirischen kumulativen Dichte. Die Wahrscheinlichkeit, jedes x_i zu erhalten, ist gleichgroß.
- Damit:
 - Produziere M Bootstrap-Sample durch Ziehen mit Zurücklegen.
- Dann: Verfahre wie oben.

Beachte, es gibt Fälle, in denen diese Idee nicht funktioniert, siehe [182], z.B.

- Eigenwerte einer Kovarianzmatrix
- Grenzen einer kompakten Verteilung

Zwei Bootstrap-Spielarten

- Nichtparametrischer Bootstrap: Á la oben: Ziehen mit Zurücklegen
- Parametrischer Bootstrap:
 - Schätze Parameter einer Verteilung aus Daten
 - Sample aus der Verteilung

Wichtig:

Untersuche (simulativ), ob die Coverage stimmt:

- Wird der wahre Wert in $(1 - \alpha)100$ Prozent der Fälle abgedeckt ?
- Wenn seltener: Procedere Quatsch.
- Wenn häufiger: Procedere konservativ.

Ende Statistikeinschub

Konfidenzintervall für die Peakfrequenz [290]

Erinnere:

$$Per(\omega_k) \sim \frac{1}{2}S(\omega_k)\chi_2^2, \quad (\ddagger)$$

Procedere:

- Schätze das Spektrum
- Simuliere M Periodogramme nach Gl. (\ddagger).
- Für jedes simulierte Periodogramm schätze das Spektrum und ermittle die Peakfrequenz ω_i^* .
- Ermittle Konfidenzintervalle aus den Quantilen der empirischen Verteilung der ω_i^* .

Fragen:

- Wie sieht bias und Varianz des Schätzers für Peakfrequenz aus?
- Wie hängt Coverage von der Breite des Spektralschätzers ab ?

3 FOLIEN PEAKCON

Das Verfahren ist konservativ. Erklärung über Peak-Unterschätzung in der initialen Schätzung.

Anmerkung: Näher an der originalen bootstrap Idee ist es [79, 63]:

- das Periodogramm durch das geschätzte Spektrum zu teilen $\epsilon_k = Per(\omega_k)/\hat{S}(\omega_k)$
- aus dem Ergebnis zu bootstrafen ϵ_k^*
- Bootstrap-Periodogramme aus $Per^*(\omega_k) = \hat{S}(\omega_k)\epsilon_k^*$ zu bilden.

5.1.2.11 Test auf Gleichheit zweier Peakfrequenzen

5.1.2.11.1 Statistikeinschub: Bootstrap Tests

- Frage bei Konfidenzintervall Ermittlung: Stimmt Coverage ?
- Entspricht beim Testen Verteilung unter H_0
- Zusätzlich beim Testen: Power of the Test.

Setting :

$$H_0 : \theta = \theta_0$$

$$H_1 : \theta \neq \theta_0$$

- Sei $\hat{\theta}$ die Funktion des Samples $x_i, i = 1, \dots, N$.
 $\hat{\theta}^*$ eine gebootstrapte Version.
- Naive Idee: Untersuche $\hat{\theta} - \theta_0$, i.e. Bootstrappe : $\hat{\theta} - \theta_0$ und schaue, ob 0 im Intervall.
- Problem:
Wenn H_0 falsch, dann ist $|\hat{\theta} - \theta_0|$ in der selben Größe wie $|\hat{\theta}^* - \theta_0| \Rightarrow$ der Test hat keine Power.
- Besser:
Bootstrappe : $\hat{\theta}^* - \hat{\theta}$ und schaue, ob θ_0 im Intervall.
- Noch besser:
Wenn Information über die Skala σ von $\hat{\theta}$ vorliegt, bootstrappe:

$$\frac{\hat{\theta}^* - \hat{\theta}}{\hat{\sigma}^*}, \text{ nicht } \hat{\theta}^* - \hat{\theta} \text{ oder } \frac{\hat{\theta}^* - \hat{\theta}}{\hat{\sigma}}$$

Dieses nennt man Pivoting, führt schneller in die Aymptotik.

Beide Tips aus [99].

Ende Statistikeinschub

5.1.2.11.2 Test auf eine Differenz von Peakfrequenzen [291]

Betrachte zwei Prozesse

- Peakfrequenzen ω_1 und ω_2
- Betrachte $\theta = \omega_1 - \omega_2$
- $H_0 : \theta = \theta_0 = 0$

Procedere:

- Schätze die Spektren und ihre Peakfrequenzen: $\hat{\omega}_1, \hat{\omega}_2$.
- Schätze die Differenz der Peakfrequenzen durch : $\hat{\theta} = \widehat{\Delta\omega} = \hat{\omega}_1 - \hat{\omega}_2$.
- Simuliere Paare von Periodogrammen nach Gl. (‡)
- Schätze die Spektren und die Differenz $\widehat{\Delta\omega}^*$.
- (i) Ermittle die Verteilung $(\widehat{\Delta\omega}^* - \widehat{\Delta\omega})$ und verwerfe H_0 , wenn $\widehat{\Delta\omega}$ außerhalb der Quantile liegt.
- (ii) Eine natürliche Skale von $\hat{\omega}_1$ und $\hat{\omega}_2$
 - Varianz Peakfrequenz $\propto 1/(\text{Krümmung am Peak})$
 - Quadratische Näherung um peak
 - Krümmung $\propto 1/(\text{Peakbreite } b_i)^2$

ZEICHNUNG Halbwertsbreiten, quadratische Näherung

Pivotisiere mit b_1^*, b_2^* und bilde die Verteilung von

$$\frac{\widehat{\Delta\omega}^* - \widehat{\Delta\omega}}{\sqrt{b_1^{2*} + b_2^{2*}}}$$

Verwerfe H_0 , wenn $\widehat{\Delta\omega}/\sqrt{b_1^2 + b_2^2}$ außerhalb der Quantile liegt.

Power of the Test - Untersuchung

- Peakfrequenz für AR[2] Prozesse:

$$\omega_{peak} = 2\pi \arccos(\cos(2\pi/T) \cosh(1/\tau))$$

- Prozeß 1: AR[2] mit $T=50$, $\tau = 100$
- Prozeß 1: AR[2] mit $T=50.15$, $\tau = 500$

FIG 1 und 2 aus Diffcon

Kurvenvertauschung wg. konservativem Niveau bei Variante (i)

Beide Verfahren konservativ, weil Spektren am peak unter- und zu breit geschätzt werden.

Bemerkung zu beiden Verfahren: Es wird angenommen, daß es keine Korrelationen

$$\begin{aligned} \langle Per(\omega_i) Per(\omega_j) \rangle &\propto \delta_{ij} \\ \langle Per(\omega_i) Per(\omega_j) Per(\omega_{i+j}) \rangle &\propto \delta_{ij} \end{aligned}$$

gibt. Sonst darf man nicht einzeln bootstrappen.

- Vernachlässigung "kurzreichweitiger" (immer positiver) Korrelationen durchs Tapern führt zu Überschätzung der Variabilität, macht konservativ.
- Vernachlässigung "langreichweitiger" Drei-Punkt-Korrelationen im Falle nicht linearer Oszillatoren kein Problem, da diese zwischen den Harmonischen auftreten.

5.1.2.12 Tests auf schwache Stationarität:

Für lineare Prozesse [228]

- Logarithmierte geschätzte Spektren $\hat{Y}(i, j) = \log(\hat{S}(t_i, \omega_j))$ zu i Zeiten in j Frequenzbändern dann Streuung frequenzunabhängig.

$$\log(\text{Per}(\omega_k)) = \log\left(\frac{1}{2}S(\omega_k) \chi_2^2\right) = \log\left(\frac{1}{2}S(\omega_k)\right) + \log(\chi_2^2)$$

$$\hat{Y}(i, j) = \log S(i, j) + \epsilon(i, j), \quad \epsilon(i, j) \sim N(0, \sigma^2)$$

-

$$Y(i, j) = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ij}$$

- Berechne: Zwei faktorielle ANOVA

Teste γ_{ij} (Interaction) , α_i (Stationarität) (β_j (white noise)) auf Null

Interpretation γ_{ij} :

Wenn $\gamma_{ij} = 0$, dann

$$S(t_i, \omega_j) = c^2(t)S(\omega) \iff x(t) = c(t)x_0(t), \quad x_0(t) \text{ stationär}$$

”Uniformly modulated process”

Item	DoF	Sum of Squares
Between times	I-1	$S_T = J \sum_{i=1}^I (Y_{i.} - Y_{..})^2$
Between frequencies	J-1	$S_F = I \sum_{j=1}^J (Y_{.j} - Y_{..})^2$
Interaction	(I-1)(J-1)	$S_I = \sum_{i=1}^I \sum_{j=1}^J (Y_{ij} - Y_{i.} - Y_{.j} + Y_{..})^2$
Total	IJ-1	$S_0 = \sum_{i=1}^I \sum_{j=1}^J (Y_{ij} - Y_{..})^2$

Durch Logarithmieren Residualvarianz unabhängig von ω und bekannt:
Teststatistik χ^2 - statt F -verteilt.

- 1. Schritt: $\gamma_{ij} = 0$?

$$S_I/\sigma^2 \sim \chi_{(I-1)(J-1)}^2$$

2. Schritt:

Stationär:

$$S_T/\sigma^2 \sim \chi_{(I-1)}^2$$

- Problem:

Geht für nichtlineare Prozesse wegen der Kopplungen höherer Ordnung (Cramér Darstellung) schief.

Für nichtlineare Prozesse [327]:

Zerstöre die Korrelationen höherer Ordnung von Hand.

Es bleiben die Dilemmata I-V:

Welche Verletzung von Stationarität ist für welche Frage/Methode wie schlimm ?

Übung :

Naiver Stationaritätstest mit AR[1] Prozeß [137]

5.1.3 Schätzung für periodisch korrelierte Prozesse

Bifrequenzperiodogramm:

$$Per(\omega_p, \omega_q) = f(\omega_p)f^*(\omega_q) \in \mathbb{C}$$

Schätze Bifrequenzspektrum durch:

$$Bi\widehat{Fre}S(\omega_p, \omega_q) = \sum_{m=-h}^h Per(\omega_{p+m}, \omega_{q+m})$$

Glättung entlang der Nebendiagonalen, da in dieser Richtung stetig.
Spektralkohärenz $\gamma(\omega_p, \omega_q)$ durch

$$\gamma(\omega_p, \omega_q) = \sum_{p=-h}^h \frac{Bi\widehat{Fre}S(\omega_p, \omega_q)}{\sqrt{\widehat{S}(\omega_p)\widehat{S}(\omega_q)}}$$

Kohärente Mittelung entlang Nebendiagonalen:

$$st_{co}(\omega_d) = |\gamma(\omega_p, \omega_{p+d})|^2, \quad h = N, \text{ modulo Randeffekte}$$

Inkohärente Mittelung :

$$st_{in}(\omega_d) = \frac{1}{L} \sum_{l=1}^L |\gamma(\omega_{pl}, \omega_{pl+d})|^2$$

Test auf periodische Korrelationen: Analog zum Test auf harmonische Komponenten in WN [92]

Anmerkung (für die Liebhaber der EMG Analyse):

$$st_{co}(\omega) \propto |S_z(\omega)|^2, \quad z(t) = x^2(t)$$

5.1.4 Nachweis von Fractional Browian Motion oder FGN

Probleme:

- Man muß $1/f^\alpha$ Skalierung für kleine f nachweisen
- Mache log-log-plot
- Da sieht vieles linear aus
- Frühe Kritik an Mandelbrot-paper : AR ist auch nicht abzulehnen [190, 121]

ZEICHNUNG AR[1] Spektrum

FOLIE NATURE ASTRO PAPER

FOLIE EL NINO paper

- Konzept überschätzt, es ist ein inverses Problem, dafür gilt:
"If your only tool is a hammer, every problem will be a nail"

Für statistische Analyse solcher Daten: [21, 22, 279, 185]

5.1.5 Zum Merken

"Spektralanalyse = lineare Analyse"

Aber:

Was man alles im Spektrum sieht:

- Oszillationen

- Oszillator oder Relaxator(en)
- Höhere Harmonische bei nichtlinearen Schwingern und Grad der Nichtlinearität
- Varianz des Beobachtungsrauschen im Hochfrequenten
- Stochastisch vs. chaotisch im Hochfrequenten [270, 269]
- $1/f^\alpha$ Rauschen
- Instationaritäten (Beispiel: Run down von Ionenkanal)

Übung :

Spektralanalyse, oder was macht der peak beim Rössler ?

5.2 Kreuzspektralanalyse

5.2.1 Analytisches

Analog zum (Auto-)spektrum: Das Kreuzspektrum für zwei Prozesse $x(t)$ und $y(t)$

$$CS(\omega) = \langle f_x(\omega) f_y^*(\omega) \rangle = \sum e^{-i\omega\tau} CCF(\tau)$$

mit Kreuzkorrelationsfunktion:

$$CCF(\tau) = \langle y(t)x(t - \tau) \rangle$$

$CS(\omega)$ ist komplex, bilde

- Kohärenzspektrum

$$Coh(\omega) = \frac{|CS(\omega)|}{\sqrt{S_x(\omega)S_y(\omega)}}, \quad Coh(\omega) \in [0, 1]$$

und

- Phasenspektrum

$$CS(\omega) = |CS(\omega)| e^{i\Phi(\omega)} \quad . \quad (5.10)$$

Bedeutung Kohärenzspektrum:

- Wenn $y(t)$ eine lineare Funktion von $x(t)$ ist, $y(t) = L(x(t))$, folgt:
 $Coh(\omega) = 1$

$$\begin{aligned} y(t) &= \sum_i a_i x(t-i) = \phi(B)x(t) \\ f_y(\omega) &= \phi(e^{-i\omega})f_x(\omega) \\ Coh^2(\omega) &= \frac{|\langle f_y(\omega)f_x^*(\omega) \rangle|^2}{\langle f_y(\omega)f_y^*(\omega) \rangle \langle f_x(\omega)f_x^*(\omega) \rangle} \\ &= \frac{|\langle \phi(e^{-i\omega})f_x(\omega)f_x^*(\omega) \rangle|^2}{\langle \phi(e^{-i\omega})f_x(\omega)\phi(e^{i\omega})f_x^*(\omega) \rangle \langle f_x(\omega)f_x^*(\omega) \rangle} \\ &= \frac{|\phi(e^{-i\omega})|^2 \langle f_x(\omega)f_x^*(\omega) \rangle^2}{|\phi(e^{-i\omega})|^2 \langle f_x(\omega)f_x^*(\omega) \rangle \langle f_x(\omega)f_x^*(\omega) \rangle} \\ &= 1 \end{aligned}$$

- Maß für linearen Zusammenhang, lineare Vorhersagbarkeit.
- Kohärenz zwischen zwei Sinüssen macht keinen Sinn.
- Für mischende Prozesse Kausalität, siehe aber Graphische Modelle, S. 113.

Fünf Gründe für $Coh(\omega) \neq 1$:

- Prozesse sind unabhängig
- Nichtlineare Beziehung

$$y(t) = x^2(t) \quad \implies Coh(\omega) = 0.$$

$$y(t) = x^3(t) \quad \implies Coh(\omega) > 0.$$

I.a. Taylorentwicklung mit führenden linearen Term.

- Andere Einflüsse auf $y(t)$ als nur von $x(t)$
- Misalignment [105], siehe

Übung: Misalignment

Schätze das Kohärenz- und Phasenspektrum zwischen einem AR[2] Prozess und seinem Rauschen

- Beobachtungsrauschen

Wenn $y(t) = L(x(t))$, aber $y(t)$ mit Beobachtungsrauschen belegt ist

$$Coh(\omega) = \sqrt{\frac{S_y(\omega)}{S_y(\omega) + \sigma_{ob}^2}} .$$

Wenn $x(t)$ und $y(t)$ BR enthalten:

$$Coh(\omega) = \sqrt{1 - \frac{\sigma_x^2 S_y + S_x \sigma_y^2 + \sigma_x^2 \sigma_y^2}{(S_x + \sigma_x^2)(S_y + \sigma_y^2)}} ,$$

FOLIEN OBS-NOISE

Bemerkung:

Linearität der Prozesse ist keine Voraussetzung für sinnvolle Interpretation. Nichtlinearität bewirkt "nur" Korrelationen höherer Ordnung der Schätzer zwischen verschiedenen Frequenzen (Cramér Darstellung).

Bedeutung Phasenspektrum:

- Wenn $y(t) = L(x(t))$, dann $\Phi(\omega) = f(L(\cdot))$.

Merke: $\Phi(\omega)$ hängt **nicht** von $x(t)$ ab. $x(t)$ kann auch z.B. chaotisch, z.B. Rössler-System sein.

Beispiele:

- Reines Delay : $y(t) = x(t - \Delta)$ dann folgt

$$y(t) = B^\Delta(x(t))$$

damit

$$f_y(\omega) = e^{-i\Delta\omega} f_x(\omega)$$

und

$$\Phi(\omega) = \Delta\omega$$

ein linearer Phasenverlauf mit Steigung = Delay.

- Wenn $y(t)$ die Ableitung von $x(t)$ ist, i.e. $y(t) = \dot{x}(t)$, folgt

$$f_y(\omega) = \sum e^{-i\omega t} \frac{d}{dt} x(t) = i\omega f_x(\omega)$$

Mit

$$e^{i\pi/2} = i$$

folgt

$$\Phi(\omega) = \pi/2 \quad .$$

- Für AR[1] Prozeß und sein treibendes Rauschen folgt

$$\begin{aligned} x(t) &= ax(t-1) + \epsilon(t) \\ (1 - aB)x(t) &= \epsilon(t) \\ (1 - ae^{-i\omega})f_x(\omega) &= f_\epsilon(\omega) \\ f_x(\omega) &= \frac{1}{1 - ae^{-i\omega}} f_\epsilon(\omega) \\ CS(\omega) &= \langle f_x(\omega) f_y^*(\omega) \rangle \\ &= \left\langle \frac{1}{1 - ae^{-i\omega}} f_\epsilon(\omega), f_\epsilon^*(\omega) \right\rangle \end{aligned}$$

$$\Phi(\omega) = \arg \left(\frac{1}{1 - ae^{-i\omega}} \right) = \operatorname{atan} \left(\frac{a \sin(\omega)}{1 - a \cos(\omega)} \right)$$

ZEICHNUNG AR[1] Phasenspektrum

- Für AR[2] Prozeß und sein treibendes Rauschen folgt

$$\Phi(\omega) = \arctan\left(\frac{a_1 \sin \omega + a_2 \sin 2\omega}{1 - a_1 \cos \omega - a_2 \cos 2\omega}\right) .$$

ZEICHNUNG AR[2] Phasenspektrum

- Nullstelle Gain

Das Gain eines linearen Systems ist definiert als :

$$G(\omega) = \frac{|CS_{xy}(\omega)|}{S_x(\omega)}$$

hat $G(\omega)$ eine Nullstelle, macht $\phi(\omega)$ einen Sprung um π [223].

Erklärung mit Thomas' Folien.

Späteres Ziel:

Durch Fit von (parameterisierten) Phasenspektren an empirische etwas über die Prozesse lernen.

5.2.1.1 Diskret vs. Kontinuierlich Part II

Phasenspektrum: Obige Rechnungen gelten für zeitdiskrete Prozesse.

Zeitkontinuierliche Rechnungen:

- AR[1]

$$\dot{x} = -\alpha x + \epsilon$$

Ansatz Frequenzweise $\epsilon(t) = e^{i\omega t}$, $x(t) = A(\omega)e^{i\omega t - \phi(\omega)}$

$$\Phi(\omega) = \text{atan}(\omega/\alpha)$$

Für vernünftiges Sampling:

FOLIE Phase konti vs. discrete AR1

$$\Phi_{AR1,cont}(\omega) \cong \Phi_{AR1,discr}(\omega) + \omega/2 .$$

- AR[2] Die klassische Resonanzformel

ZEICHNUNG Phase konti vs. discrete AR2

$$\Phi_{AR2,cont}(\omega) \cong \Phi_{AR2,discr}(\omega) + \omega \quad .$$

Folge:

Möchte man durch Fit parametrisierter Phasenspektren an empirische etwas über Zusammenhang der Prozesse lernen, so ergibt sich im

- AR[1] Fall artifizielles Delay mit Zeitschritt $\Delta = \text{Samplingintervall}/2$.
- AR[2] Falle art. Delay $\Delta = \text{Samplingintervall}$.

Eventuell:

Korrelationskoeffizient und Fischer-Trafo

5.2.1.2 Filtern

Literatur:

- A.V. Oppenheim and R.W. Schaffer: Digital signal processing [211]
- L.R. Rabiner and B. Gold: Theory and Application of Digital Signal Processing [230]

Lineare Filter sind eine andere Sichtweise auf ARMA[p,q] Prozesse:

$$\Theta(B)x(t) = \phi(B)\epsilon(t) \quad ,$$

nur daß $\epsilon(t)$ jetzt durch ein zu filterndes Signal $v(t)$ ersetzt wird

$$\Theta(B)x(t) = \phi(B)v(t) \quad .$$

Die Eigenschaften des Filters hängen davon nicht ab, sondern nur von der Transferfunktion:

$$H = \frac{\phi(e^{-i\omega})}{\Theta(e^{-i\omega})}$$

Für diese wünscht man sich Eigenschaften, gegeben maximale Werte p, q : Filter-Design, siehe z.B. [211, 230]

In der Regel möchte man:

- Scharfe Kanten im Frequenzraum
- Linearen Phasenverlauf

Das ist nicht gleichzeitig zu erfüllen.

Charakterisierende Eigenschaften von Filtern:

- Magnitude (Gain=log(Magnitude) ("dB"))

$$mag(\omega) = |H(\omega)|^2$$

- Phase

$$\Phi(\omega) = \arg(H(\omega))$$

Gruppengeschwindigkeit :

$$v_G = \frac{d}{d\omega} \Phi(\omega)$$

Verschiedene Tiefpaß-Filter Typen:

- Butterworth-Filter: Maximal flaches Gain an $\omega = 0$
- Bessel-Filter: Maximal linearer Phasenverlauf bei $\omega = 0$ (wenig Verzerrung)
- Chebyshev-Filter: Maximal gutes Stufenverhalten im Gain
- Elliptische-Filter: Chebyshev-Filter mit begrenztem "Ribbeln"

JEWELLS FOLIE

Reiner AR-Filter : Infinite Impulse Response (IIR)

Reiner MA-Filter : Finite Impulse Response (FIR)

2 wichtige Fälle:

- On-line (analog, zeitkontinuierlich):

Filterkonstruktion, wichtige Stichworte:

- z -Transformation
- Bilineare transformation
- s -Transformation, s -plane

$$s = \frac{1 - z^{-1}}{T}$$

Speziell Anti-aliasing Tiefpass-filtern:

- Alles über der Nyquist Frequenz muß so gut wie möglich weg,
 - aber die Signale sollen unverzerrt bleiben \implies lineare Phase \implies Bessel-Filter möglichst hoher Ordnung.
- Offline bandpass-filtern
 - FT
 - $f(\omega) = 0$ unerwünschte Frequenzen
 - FT $^{-1}$

Nicht, wenn man spiky oder scharfkantige Signale hat, bewirkt dann Überschwingen.

Zusammenhang Kreuzspektralanalyse – Filtern klar ?

5.2.2 Schätzerei

Geschätzt wird $Coh(\omega)$ und $\Phi(\omega)$ über Kreuzperiodogramm und Autoperiodogramme

$$\widehat{Coh}(\omega_k) = \frac{|\sum_{i=-h}^h w_i f_x(\omega_{k+i}) f_y^*(\omega_{k+i})|}{\sqrt{\sum_{i=-h}^h w_i |f_x(\omega_{k+i})|^2 \sum_{i=-h}^h w_i |f_y(\omega_{k+i})|^2}}$$

Wenn nicht geglättet, folgt $Coh(\omega) = 1$.

Fundamental: Glätten oder nicht Glätten offenbart die getroffene Annahme über

- "Ich analysiere einen mischenden Prozeß"

oder

- "Ich analysiere einen periodischen Prozeß"

Konfidenzintervall für Test auf $Coh(\omega) = 0$:

$$s = \sqrt{1 - \alpha^{\frac{2}{\nu-2}}} \quad ,$$

mit effektiver Anzahl von Freiheitsgraden, siehe S. 79:

$$\nu = \frac{2}{\sum_{i=-h}^h w_i^2} \quad .$$

Schätzung Phasenspektrum:

$$\widehat{CS}(\omega) = \sum_{i=-h}^h w_i f_x(\omega_{k+i}) f_y^*(\omega_{k+i})$$

Phase:

$$\widehat{\Phi}(\omega) = \text{atan} \left(\frac{\text{Im}(\widehat{CS}(\omega))}{\text{Re}(\widehat{CS}(\omega))} \right)$$

Konfidenzintervall Phasenspektrum

$$\text{var}(\widehat{\Phi}(\omega)) = \frac{1}{\nu} \left(\frac{1}{\widehat{Coh}^2(\omega)} - 1 \right) \quad , \quad (5.11)$$

Gilt nur für $Coh(\omega) > 0$:

- Fehler ist Funktion der Kohärenz.
- Das ist sinnvoll: Je geringer der lineare Zusammenhang ist, desto schlechter kann man seine Phasenbeziehung wissen.
- Da $\Phi(\omega)$ 2π periodisch, ist es nur auswertbar, wenn der Fehler kleiner als, say $\pi/5$ ist.

Ausflug CCF

- Info von Kohärenz und Phase ist 1:1 zu Info in CCF und ACF

- Ist dort i.a. schwer sichtbar, siehe [292, 293]
- Liegt u.a. an der Korrelationsstruktur der Schätzfehler (linearer Fall) (Bartell-Formula [15]):

$$\begin{aligned} \langle (CCF(h) - \widehat{CCF}(h))(CCF(k) - \widehat{CCF}(k)) \rangle = \\ \frac{1}{N} \sum_{j=-N}^N ACF_1(j)ACF_2(j+k-h) + CCF(j+k)CCF(-h+j) - \\ CCF(h)\{ACF_1(j)CCF(j+k) - ACF_2(j)CCF(-k+j)\} - \\ CCF(-j)\{ACF_1(j)CCF(-j-k) - ACF_2(j)CCF(k-j)\} + \\ CCF(h)CCF(k)\{\frac{1}{2}ACF^2(j) - CCF^2(j) + \frac{1}{2}ACF_2(j)^2\} \end{aligned}$$

d.h. die wahre CCF, ACFs bestimmen die Korrelationsstruktur in boshafter Weise.

- Spezialfall: Unkorreliert, $CCF(\tau) = 0$ durch Wischen:

$$\langle (CCF(h) - \widehat{CCF}(h))(CCF(k) - \widehat{CCF}(k)) \rangle = \frac{1}{N} \sum_{j=-N}^N ACF_1(j)ACF_2(j+k-h)$$

- speziell:

$$\langle (CCF(h) - \widehat{CCF}(h))^2 \rangle = \frac{1}{N} \sum_{j=-N}^N ACF_1(j)ACF_2(j)$$

Fehlerbalken natürlicherweise größer als bei WN.

Übung:

Koherenz zwischen zwei Rösslern

Besprechung von Phys. Tremor und CMMs, Lancet-paper [292, 293, 168, 115]

Für Rückkopplung, siehe [3]

Für zahlreiche Anwendungen, siehe [294]

5.2.2.1 Delay - Schätzung mit Hilberttransformation

[103, 206, 29, 57]

- Wichtigste Anwendung der Analyse des Phasenspektrums: Delay-Schätzung.

Beispiel: Erdölsuche ZEICHNUNG

- Problem: I.d.R.: Außer reinem Delay zusätzlich Filtereigenschaften in den Prozessen.
- Beispiel: Nervenleitung ist ein Tiefpaßfilter
- Daraus folgt: Anpassung einer (lokalen) Gradienten kann sinnlos sein.

Formalisiert:

$$\sum_{i=0}^p b_i B^i y(t) = \sum_{i=0}^q c_i B^i x(t - \tau)$$

Nach Invertierung folgt:

$$y(t) = \sum_{j=\tau}^{\infty} a_j x(t - j)$$

- τ interessierendes Delay
- a_j Transferfunktion, macht das Problem

In den Fourierraum (mit $\Delta = \tau$):

Aus

$$y(t) = \int_{\Delta}^{\infty} a(t') x(t - t') dt'$$

folgt mit

$$f_a(\omega) = \int_0^{\infty} a(t + \Delta) e^{-i\omega t} dt$$

$$CS(\omega) = \langle f_y f_x^* \rangle = \langle e^{i\Delta\omega} f_a f_x f_x^* \rangle = e^{i\Delta\omega} f_a(\omega) S_x(\omega)$$

damit

$$\Phi(\omega) = \Delta\omega + \arg(f_a(\omega))$$

Einschub: Hilberttransformation [103, 206] :

- Ist die Funktion (in Polarkoordinaten)

$$\tilde{X}(z) = \log|X(z)| + i \arg(X(z))$$

in der oberen Halbebene well-behaved, sind Real- und Imaginärteil durch die Hilberttrafo verknüpft :

$$\arg(X(e^{i\omega})) = \frac{1}{2\pi} CP \int_0^\pi d\theta \log|X(\theta)| \left(\cot\left(\frac{\omega - \theta}{2}\right) + \cot\left(\frac{\omega + \theta}{2}\right) \right)$$

- Physikern aus der Streutheorie bekannt, wo Amplituden gemessen werden, und unter Annahme Linearität und Kausalität Phasen berechnet werden (Kramers-Kronig-Relationen).
- Setzt man :

$$|X(\theta)| = \frac{|CS(\theta)|}{S_x(\theta)}$$

liefert Hilberttrafo eine Phasenfunktion:

$$\Phi_{MP}(\omega) = \frac{1}{2\pi} CP \int_0^\pi d\theta \log \frac{|CS(\theta)|}{S_x(\theta)} \left(\cot\left(\frac{\omega - \theta}{2}\right) + \cot\left(\frac{\omega + \theta}{2}\right) \right)$$

- Diese hat die Minimum-Phase Eigenschaft: Phasenfunktion mit der schnellsten Übertragungsgeschwindigkeit, die mit

$$|X(\theta)| = \frac{|CS(\theta)|}{S_x(\theta)}$$

verträglich ist. Siehe Unbehauen: Systemtheorie, Kap. 2.7, 4.7 oder 7.2
Siehe vor allem: [18], p. 798f

Ende Einschub

Man kann zeigen: Jedes Phasenspektrum eines kausalen Prozesses gehorcht:

$$\Phi(\omega) = \Delta\omega + \Phi_{MP}(\omega) + AllPol(\omega)$$

$AllPol(\omega)$ i.d.R. unphysikalisch und gleich Null, dann:

$$\Phi_{MP}(\omega) = \arg(f_a(\omega))$$

Delay-schätzung:

- Idee : Bestimme Delay Δ aus :

$$\min_{\Delta} \sum \left(\hat{\Phi}(\omega_k) - (\Delta\omega_k + \Phi_{MP}(\omega_k)) \right)^2$$

- Problem: Phase 2π - periodisch, speziell $\hat{\Phi}(\omega_k)$ hat i.d.R. Phasensprünge.
- Lösung :

$$\max_{\Delta} \sum \cos \left(\hat{\Phi}(\omega_k) - (\Delta\omega_k + \Phi_{MP}(\omega_k)) \right)$$

- Problem:
Gewichtung der Beiträge ?
- Lösung, siehe Gl. (5.11):

$$\max_{\Delta} \sum \frac{\widehat{Coh}^2(\omega_k)}{1 - \widehat{Coh}^2(\omega_k)} \cos \left(\hat{\Phi}(\omega_k) - (\Delta\omega_k + \Phi_{MP}(\omega_k)) \right)$$

ist fast MLE.

FOLIE Lindemann-Simus, Noise Untersuchung erwähnen.

5.2.3 Graphische Modelle

Literatur für den nichtzeitreihen Fall :

- D.R. Cox and N. Wermuth: Multivariate Dependencies [53]
- S.L. Lauritzen: Graphical Modelling [172]
- J. Whittaker: Graphical Models in Applied Multivariate Statistics [324]
- Grundsätzliches Problem:
Methoden für bivariaten Fall angewandt im multivariaten Fall können Scheinkorrelationen ergeben:
 - Sitze der Republikaner und Sonnenflecken
 - Die Babies und die Störche

ZEICHNUNG Babies, Störche und soziokulturelle Entwicklung

- Idee:
Betrachte Korrelation (Kohärenz) zwischen x und y unter Berücksichtigung von $z =$ alle andere Information.
- Im Regressionsfall:
FOLIE MURDERER, Aufklärung der Überraschung
Säuglinge, Aufklärung des Vorurteils

Formalisiert:

- Betrachte multivariate Gaußverteilung

$$\vec{x} = N(\vec{\mu}, \Sigma)$$

- Teile \vec{x} in $\vec{x} = (\vec{x}_1, \vec{x}_2)$
Entsprechend

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

- Unbedingte Verteilungen:

$$\vec{x}_i = N(\vec{\mu}_i, \Sigma_{ii})$$

- Bedingte Verteilung von \vec{x}_2 gegeben \vec{x}_1 hat Erwartungswert:

$$E(\vec{x}_2|\vec{x}_1) = \vec{\mu}_2 + \Sigma_{21}\Sigma_{11}^{-1}(\vec{x}_1 - \vec{\mu}_1)$$

und Kovarianz:

$$Cov(\vec{x}_2|\vec{x}_1) = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$$

- Partielle Kovarianz zwischen X und Y (Aufspaltung von \vec{x}_2), gegeben Z (\vec{x}_1):

$$Cov((\vec{X}, \vec{Y})|\vec{Z}) = Cov(\vec{X}, \vec{Y}) - Cov(\vec{X}, \vec{Z})Var^{-1}(Z)Cov(\vec{Z}, \vec{Y})$$

Bezeichnung: "rauspartialisieren"

- Bedingte lineare Unabhängigkeit zwischen X und Y , gegeben Z :

$$Cov((\vec{X}, \vec{Y})|\vec{Z}) = 0 \quad \text{Bezeichnung: } X \Upsilon Y | Z$$

Allgemein, in terms of Dichten:

$$p(x|y, z) = p(x|z)$$

"Knowing Z , reading Y is irrelevant for reading X "

- Aber Phänomen: "Married parents of a joint child"

ZEICHNUNG

Beispiel: 2 Würfel, Ergebnis addieren

"Modellmisspezifikation"

"Knowing Z , reading Y , you know everything about X "

- Aufgabe:
Konstruiere den "Moral Graph"

Für Zeitreihen [95, 61, 60, 294]:

- Besonders wichtig der Fall:

ZEICHNUNG: $A \rightarrow B \rightarrow C$

- \vec{x} von oben entspricht $\vec{f}(\omega)$, (Kreuz-)spektren den (Ko-)Varianzen
- Partielles Kreuzspektrum:

$$PCS_{X_1 X_2 | Y}(\omega) = CS_{X_1 X_2}(\omega) - CS_{X_1 Y}(\omega) S_Y(\omega)^{-1} CS_{Y X_2}(\omega)$$

Daraus partielles Kohärenzspektrum $Coh_{part}(\omega)$

$$Coh_{part}(\omega) = \frac{|PCS_{X_1 X_2 | Y}(\omega)|}{S_{X_1 | Y} S_{X_2 | Y}}$$

und partielles Phasenspektrum $\Phi_{part}(\omega)$.

- Konfidenzintervall für $Coh_{part}(\omega) = 0$:

$$s = \sqrt{1 - \alpha^{\frac{2}{\nu - 2L - 2}}} \quad .$$

ν effektive DoF des Spektralschätzers, $L = |Y|$

Analyseansatz:

- Vergleiche (normale) Kreuzspektralanalyseergebnisse mit denen der partiellen
- Verschwindet Koherenz beim Partialisieren, war es eine indirekte Verbindung
- Entsteht sie: Married parents of a joint child

FOLIE SIEGFRIEDS HAND Phase und SNR erwähnen.

Vorsicht:

- Sind Komponenten stark korreliert, unterschätzen Graphische Modelle die Kopplungen zu den weniger stark korrelierten Komponenten. Erklären.
- Die Hoffnung, einfach hochdimensionale Zeitreihen hineinstecken zu können und die "Verschaltung" herauszubekommen, hat sich nicht erfüllt.

Weitere Methoden:

- Granger Causality
- Directed partial coherence [251, 13, 254]
- Directed transfer function [149, 150]

AUSBAUEN

Basieren auf Fitten von multivariaten AR Prozessen mit Darstellung der Ergebnisse im Frequenzraum.

Zeit-abhängige Version: [326, 255]

5.3 Bispektrum

Literatur:

- C.L. Nikias and M.R. Raghuveer: Bispectrum Estimation: A digital signal processing framework [210]
- T. Subba Rao, M.M. Gabr: An Introduction to bispectral Analysis and bilinear Time Series Models [275] eher mathematisch

Cramér Darstellung revisited:

$$x(t) = \sum_k e^{i\omega_k t} f(\omega_k)$$

Für stationäre Prozesse gilt:

$$\begin{aligned}
\langle f(\omega_k) \rangle &= 0 \\
\langle f(\omega_1)f^*(\omega_2) \rangle &= \begin{cases} 0 & \text{für } \omega_1 \neq \omega_2 \\ S(\omega_1) & \text{für } \omega_1 = \omega_2 \end{cases} \\
\langle f(\omega_1)f(\omega_2)f^*(\omega_3) \rangle &= \begin{cases} 0 & \text{für } \omega_1 + \omega_2 \neq \omega_3 \\ B(\omega_1, \omega_2) & \text{für } \omega_1 + \omega_2 = \omega_3 \end{cases}
\end{aligned}$$

Trispektrum

$$B(\omega_1, \omega_2) = \langle f(\omega_1)f(\omega_2)f^*(\omega_1 + \omega_2) \rangle$$

Schätzung durch Glättung in 2 D:

$$\widehat{B}(\omega_k, \omega_l) = \sum_{i=-h}^h \sum_{j=-h}^h f(\omega_{k+i})f(\omega_{l+j})f^*(\omega_{k+i} + \omega_{l+j})$$

oder durch Stücke mitteln oder über Tripelkorrelationsfunktion

Phasenkopplung (ω_1, ω_2) an $\omega_1 + \omega_2$ entsteht aus quadratischer Nichtlinearität
Betrachte:

$$\begin{aligned}
\ddot{x} &= -\gamma(x, y)\dot{x} - \omega_x^2 x + Force_x \\
\ddot{y} &= -\gamma(x, y)\dot{y} - \omega_y^2 y + Force_y \\
\gamma &= ax + by + cxy + \dots \\
z &= x + y \text{ Beobachtung} \\
\sin \omega_x \sin \omega_y &= \frac{1}{2} \cos(\omega_x + \omega_y) - \cos(\omega_x - \omega_y)
\end{aligned}$$

Anwendung: [159, 158]

Test auf Linearität [120, 232]

Für lineare Prozesse ist das Bispektrum im

- Gaußschen Falle $B(\omega_1, \omega_2) = 0$
- im nichtGaußschen Falle $B(\omega_1, \omega_2) = \kappa$, $\kappa =$ Schiefe des Prozesses.

Bemerkung: Sei $\{X_t\}$ ein stationärer Prozess zur Ordnung drei. Dann existiert das dritte Moment und hängt nur von t_1, t_2 ab:

$$E[X_t X_{t+t_1} X_{t+t_2}] = E[X_0 X_{t_1} X_{t_2}] = m(t_1, t_2)$$

Das zentrale Moment dritter Ordnung ist definiert als

$$C(t_1, t_2) = E[(X_t - \mu)(X_{t+t_1} - \mu)(X_{t+t_2} - \mu)]. \quad (5.12)$$

5.3.1 Bispektrum

Die bispektrale Dichtefunktion oder kurz das Bispektrum $f(\omega_1, \omega_2)$ ist definiert durch:

$$f(\omega_1, \omega_2) = \frac{1}{(2\pi)^2} \sum_{t_1=-\infty}^{\infty} \sum_{t_2=-\infty}^{\infty} C(t_1, t_2) e^{-it_1\omega_1 - it_2\omega_2}, \quad -\pi \leq \omega_1, \omega_2 \leq \pi$$

und damit

$$C(t_1, t_2) = \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} e^{it_1\omega_1 + it_2\omega_2} f(\omega_1, \omega_2) d\omega_1 d\omega_2 \quad (5.13)$$

Das Bispektrum $f(\omega_1, \omega_2)$ ist eine komplexwertige Funktion und existiert für alle ω_1, ω_2 falls

$$\sum_{t_1=-\infty}^{\infty} \sum_{t_2=-\infty}^{\infty} |C(t_1, t_2)| < \infty$$

Aus (5.13) folgt für $t_1 = t_2 = 0$:

$$C(0, 0) = E[(X_t - \mu)^3] = \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} f(\omega_1, \omega_2) d\omega_1 d\omega_2$$

Ähnlich wie das Spektrum das zweite zentrale Moment, die Varianz, in Anteile von Frequenzbändern zerlegt, ist das Bispektrum die Zerlegung des dritten zentralen Momentes nach Beiträgen im Frequenzbereich $(\omega_1, \omega_2) \times (\omega_1 + d\omega_1, \omega_2 + d\omega_2)$. Für einen linearen gaußschen Prozess $\{X_t\}$ ist $C(t_1, t_2) \equiv 0$ und damit auch $f(\omega_1, \omega_2) \equiv 0$. Demnach kann das Bispektrum dazu benutzt werden, Abweichungen eines Prozesses vom linearen gaußschen Falle festzustellen.

5.3.2 Kreuzbisppektrum

to be added

5.3.3 Wavelets

Gabor-Transformation

$$\exp\{(t - t_0)^2/\sigma\} \cos(\omega_k t)$$

Spektrogramm

Becker DA checken

[64, 68]

FOLIEN aus Gerschenfeld

Rauschunterdrückung mittels Wavelet-Shrinkage [67]

Als Anwendung: [113, 114]

Kapitel 6

Parameterschätzung in dynamischen Systemen

6.1 Parameterschätzung

Literatur:

- D.R. Cox and D.V. Hinkley: Theoretical Statistics [51]
- E.L. Lehmann: Theory of Point Estimation [174]

6.1.1 Maximum Likelihood Schätzer

Sei X eine parametrisierte Zufallsvariable mit Dichte $p(x, a)$.
Gegeben N Realisierungen, dann heißt

$$L(x_1, \dots, x_N | a) = \prod_{i=1}^N p(x_i, a)$$

die Likelihood: "Prob(a), daß gegebenen Daten entstehen"

- $L(x_1, \dots, x_N | a)$ ist in Abhängigkeit von a zu lesen
- Daten sind gegeben
- Likelihood ist keine Wahrscheinlichkeit, da $\int p(x, a) da$ nicht normiert.
Im Gegensatz zu $\int p(x, a) dx = 1$

Maximum Likelihood Schätzer:

- Wähle Parameter a so, dass Likelihood maximal wird
- Intuitiv sinnvoll
- Formal:

$$\frac{\partial L(x_1, \dots, x_N | a)}{\partial a} = 0$$

- Logarithmiere:

$$\mathcal{L}(a) = \log L(a) = \sum_{i=1}^N \log p(x_i, a)$$

Ersetzt schwierig handhabbares Produkt durch einfachere Summe

- M.k.z.: MLE-Schätzer unter milden Bedingungen asymptotisch unverzerrt. Widerspruchsbeweis (Cox/Hinkley p. 288f, hübsch)

Cramér-Rao Schranke

Erinnerung: Effizienz bei statistischen Test

- Im folgenden alle Indices unterdrückt
- Betrachte Score:

$$V = \frac{\partial}{\partial a} \mathcal{L}(x, a) = \frac{\partial}{\partial a} \log p(x, a) = \frac{1}{p(x, a)} \frac{\partial}{\partial a} p(x, a) \quad (\dagger)$$

- Der Erwartungswert ist

$$\begin{aligned} \langle V \rangle &= \int dx p(x, a) \frac{\partial}{\partial a} \log p(x, a) \\ &= \int dx p(x, a) \frac{1}{p(x, a)} \frac{\partial}{\partial a} p(x, a) \\ &= \int dx \frac{\partial}{\partial a} p(x, a) \\ &= \frac{\partial}{\partial a} \int dx p(x, a) \\ &= 0 \end{aligned}$$

- Für Varianz

$$\text{Var}(V) := \left\langle \left(\frac{\partial}{\partial a} \mathcal{L}(x, a) \right)^2 \right\rangle$$

folgt, mit Differentiation von

$$\langle V \rangle = 0 = \int dx p(x, a) \frac{\partial}{\partial a} \log p(x, a)$$

$$0 = \int dx \frac{\partial}{\partial a} p(x, a) \frac{\partial}{\partial a} \log p(x, a) + \int dx p(x, a) \frac{\partial^2}{\partial a^2} \log p(x, a)$$

Mit (†) folgt für 1. Summanden:

$$\int dx p(x, a) \left(\frac{\partial}{\partial a} \log p(x, a) \right)^2 = \text{Var}(V)$$

und damit:

$$\text{Var}(V) = E \left(-\frac{\partial^2}{\partial a^2} \mathcal{L}(x, a) \right)$$

- $E \left(-\frac{\partial^2}{\partial a^2} \mathcal{L}(x, a) \right)$ heißt Fischer Informations Matrix.

Betrachte unverzerrten Schätzer $\hat{\theta}(x)$ für Parameter a , i.e. $\langle \hat{\theta}(x) \rangle = a$.

- Erinnerung: Varianz des Schätzers ergibt Vertrauensintervalle.
- Betrachte:

$$\begin{aligned} \langle V \hat{\theta}(x) \rangle &= \int dx p(x, a) \frac{1}{p(x, a)} \frac{\partial}{\partial a} p(x, a) \hat{\theta}(x) \\ &= \int dx \frac{\partial}{\partial a} p(x, a) \hat{\theta}(x) \\ &= \frac{\partial}{\partial a} \int dx p(x, a) \hat{\theta}(x) \\ &= \frac{\partial}{\partial a} \langle \hat{\theta}(x) \rangle \\ &= \frac{\partial}{\partial a} a \\ &= 1 \end{aligned}$$

Betrachte Cauchy-Schwarz Ungleichung:

$$\begin{aligned} \langle (V - \langle V \rangle)(\hat{\theta} - \langle \hat{\theta} \rangle) \rangle^2 &\leq \langle (V - \langle V \rangle)^2 \rangle \langle (\hat{\theta} - \langle \hat{\theta} \rangle)^2 \rangle \\ \langle V\hat{\theta} - V\langle \hat{\theta} \rangle - \langle V \rangle\hat{\theta} + \langle V \rangle\langle \hat{\theta} \rangle \rangle^2 &\leq \text{Var}(V)\text{Var}(\hat{\theta}) \\ \langle V\hat{\theta} \rangle^2 &\leq \text{Var}(V)\text{Var}(\hat{\theta}) \\ \text{Var}(\hat{\theta}) &\geq \frac{1}{\text{Var}(V)} = \frac{1}{\langle \left(-\frac{\partial^2}{\partial a^2} \mathcal{L}(x, a)\right) \rangle} \end{aligned}$$

- Krümmung der Log-Likelihood bestimmt Varianz der Schätzer.
- M.k.z.: Maximum likelihood Schätzer nimmt untere Grenze an, also

$$\text{Var}(\hat{\theta}_{\text{MLE}}) = \frac{1}{\langle \left(-\frac{\partial^2}{\partial a^2} \mathcal{L}(x, a)\right) \rangle}$$

- Effizienz:

$$\text{Eff}(\hat{\Theta}) = \frac{\text{Var}(\Theta_{\text{MLE}})}{\text{Var}(\Theta)} \leq 1$$

MLE-Schätzer ist das beste Pferd im Stall.

Lineare Fehlerfortpflanzung:

$$\hat{\Theta}_1 = g(\hat{\Theta}_2), \quad \text{Var}(\hat{\Theta}_1) = (g'(\hat{\Theta}_2))^2 \text{Var}(\hat{\Theta}_2)$$

ZEICHNUNG dazu

[51] bias-correction p.310

Konkrete Beispiele

- Mittelwert der Gaußverteilung,

$$p(x_i, \mu) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x_i - \mu)^2}{2\sigma^2}}$$

Log-Likelihood

$$\mathcal{L}(\mu) = \log \frac{N}{\sqrt{2\pi}\sigma} - \sum_{i=1}^N \frac{(x_i - \mu)^2}{2\sigma^2}$$

$$\frac{\partial \mathcal{L}(\mu)}{\partial \mu} = \sum_i \frac{(x_i - \mu)}{\sigma^2}$$

$$\sum_i (x_i - \mu) = 0$$

Damit:

$$\hat{\mu}_{MLE} = \frac{1}{N} \sum_i x_i$$

$$\frac{\partial^2 \mathcal{L}(\mu)}{\partial \mu^2} = - \sum_i \frac{1}{\sigma^2} = - \frac{N}{\sigma^2}$$

σ^2 bestimmt Krümmung: Je grösser, desto kleiner die Krümmung und damit die Varianz größer

$$\text{Var}(\hat{\mu}) = - \frac{1}{\frac{\partial^2 \mathcal{L}(\mu)}{\partial \mu^2}} = \frac{\sigma^2}{N}$$

N -Abhängigkeit macht übliche \sqrt{N} -Abhängigkeit

- Lineare Regression:

$$y_i = ax_i + \epsilon_i, \quad \epsilon_i \sim N(0, \sigma^2)$$

Log-Likelihood

$$\mathcal{L}(a) \propto \sum_{i=1}^N (y_i - ax_i)^2$$

$$\frac{\partial \mathcal{L}(a)}{\partial a} = \sum (y_i - ax_i)x_i$$

$$\sum_i (y_i - ax_i)x_i = 0$$

$$\sum (y_i x_i - ax_i^2) = 0$$

$$\hat{a}_{MLE} = \frac{\sum_i y_i x_i}{\sum_i x_i^2}$$

$\sigma_i \neq \sigma_j$ - Fall diskutieren.

- Rückwärts gelesen: Schätzt man auf Grund kleinster Quadrate, hat man Gauß-Verteilung angenommen
- Lokationsparameter bei Cauchy-Verteilung

$$\hat{\mu}_{MLE} = \text{Median}$$

- -

$$\hat{\mu} = \frac{1}{N} \sum x_i$$

hat bei Cauchy-Verteilung große Varianz (∞)

- Median-Schätzer hat bei Gauß-Verteilung grössere Varianz.
- Alles analog zur Test-Theorie

6.1.2 Methods of Moments

- Likelihood manchmal schwer zu berechnen, Beispiel siehe Kapitel 6.2.4
- Dann MoM eine Alternative
- Ansatz:
Berechne Momente ...
 - ... aus den Daten: Mo_{emp}
 - ... und aus Modell parametrisierte theoretische Momente $Mo_{theo}(\theta)$.

- Definiere Schätzer über:

$$Mo_{emp} = Mo_{theo}(\hat{\theta}_{MM})$$

- I.d.R. gilt

$$Var(\theta_{MM}) > Var(\theta_{MLE})$$

Ist das Problem linear in den Parametern und betrachtet man 2. Momente, gilt:

$$\hat{\theta}_{MM} = \hat{\theta}_{MLE}, \quad Var(\theta_{MM}) = Var(\theta_{MLE})$$

6.1.3 Ausflug Bayesianismus

Bisher (Frequentisten): Es gibt "wahre Parameter"
Bayesianismus:

- Parameter sind auch Zufallsvariablen
- Alle Wahrscheinlichkeiten sind bedingte Wahrscheinlichkeit

$$p(\theta|Data) \propto p(Data|\theta)p(\theta)$$

Die Likelihood $p(Data|\theta)$ wird durch den prior $p(\theta)$ verziert.
Liefert Maximum a posteriori (MAP)-Schätzer und seine Verteilung.

ZEICHNUNG prior, likelihood, a posteriori

- Prior Einfluß : $O(1)$ Likelihood Einfluß : $O(N)$
- Prior kann sinnvoll Vorwissen einbauen
Akkumulation von Information bei Folge von Experimenten, experimentelle Versuchplanung [171]
- Liefert (im frequentistischen Sinne) verzerrte Schätzer

Speziell wichtig bei schlecht-gestellt inversen Problemen

$$\begin{aligned} y &= Ax + \epsilon \\ \hat{x} &= A^{-1}y \end{aligned}$$

Ist A singular oder schlecht konditioniert, i.e. fast singular, grosse

$$\text{Konditionszahl} = \frac{\text{größter Eigenwert}}{\text{kleinster Eigenwert}}$$

wird x zwar unverzerrt geschätzt, aber Schätzer hat riesen Varianz

$$MSE = \langle (\hat{\theta} - \theta)^2 \rangle = bias^2 + Var(\hat{\theta})$$

ZEICHUNG: Trade-off: bias vs. Varianz

Sinnvolle priors:

- kleine $|x|$: $p(x) \propto e^{-|x|}$
- $f(x)$ glatt: $p(f(x)) \propto e^{\partial^2 / \partial x^2 f(x)}$
 - MLE unverzerrt, aber riesige Varianz
 - Grosse Varianz Reduktion für kleinen bias.
 - Siehe James Stein Schätzer [140, 72]
- Monte Carlo Markov Chain [88, 153, 212]
 - Stochastischer Prozeß auf den Parametern
 - Stationäre Dichte ist die gewünschte, sample davon
 - wichtige Technik auch für Likelihood Berechnung

6.2 ARMA Modelle

Literatur:

- R.H. Shumway and D.S. Stoffer: Time Series Analysis and Its Application [267]
- P.J. Brockwell, R.A. Davis *Time Series: Theory and Methods* [36].

6.2.1 AR Modelle

Betrachte AR[p] Prozeß:

$$x_t = a_1 x_{t-1} + \dots + a_p x_{t-p} + \epsilon_t \quad \epsilon_t \sim N(0, \sigma^2)$$

- Multiplikation mit $x_{t-j}, j = 0, \dots, p$ ergibt (mit $\gamma(\tau) = ACF(\tau)$):

$$\begin{aligned} \gamma(0) &= a_1 \gamma(1) + \dots + a_p \gamma(p) + \sigma^2 \\ \gamma(\tau) &= a_1 \gamma(\tau - 1) + \dots + a_p \gamma(\tau - p), \quad \tau = 1, 2, \dots, p \end{aligned}$$

oder

$$\vec{\gamma} = \Gamma \vec{a}, \quad \sigma^2 = \gamma(0) - \vec{\gamma}' \vec{a}$$

mit

$$\Gamma = \{\gamma(k-j)\}_{j,k=1}^p, \quad \vec{a}_p = (a_1, \dots, a_p)'$$

- Dieses sind die Yule-Walker Gleichungen. Γ hat Toeplitz-Struktur.
- Methods of Moments: Ersetze $\vec{\gamma}$ durch $\hat{\vec{\gamma}}$:

$$\hat{\vec{a}} = \hat{\Gamma}^{-1} \hat{\vec{\gamma}}, \quad \hat{\sigma}^2 = \hat{\gamma}(0) - \hat{\vec{\gamma}}' \hat{\Gamma}^{-1} \hat{\vec{\gamma}}$$

- Es gilt:

$$\sqrt{N}(\hat{\vec{a}} - \vec{a}) \rightarrow N(0, \sigma^2 \Gamma^{-1}) \quad \text{oder} \quad \hat{\vec{a}} \rightarrow N\left(\vec{a}, \frac{\sigma^2 \Gamma^{-1}}{N}\right)$$

Zur Berechnung muß Γ invertiert werden.

- Ausweg: **Durbin-Levinson Algorithmus** (alle $\hat{\cdot}$ weggelassen)

$$\begin{aligned}\phi_{nn} &= \frac{\gamma(n) - \sum_{k=1}^{n-1} \phi_{n-1,k} \gamma(n-k)}{1 - \sum_{k=1}^{n-1} \phi_{n-1,k} \gamma(k)}, & \phi_{00} &= 0 \\ \phi_{nk} &= \phi_{n-1,k} - \phi_{nn} \phi_{n-1,n-k}, & k &= 1, \dots, n-1\end{aligned}$$

Für $n = p$, ist $a_i = \phi_{pi}$.

- Nebeneffekt: Liefert PACF(τ) in Form der ϕ_{nn} .

Erster Schritt zur Modellselektion, da:

$$\sqrt{N}(\phi_{hh}) = \sqrt{N} a_h \rightarrow N(0, \sigma^2), \text{ für } h > p$$

FOLIE BD 237

- Für AR ist YW-MM-Schätzer MLE, da i.w. least squares Schätzer in einem linearen Gaußschen Modell.

6.2.2 MA Modelle

Yule-Walker Ansatz:

- Transformiere

$$\begin{aligned}x(t) &= \epsilon(t) + \theta \epsilon(t-1) \\ x(t) &= (1 + \theta B)\epsilon(t) \\ \left(\frac{1}{1 + \theta B}\right) x(t) &= \epsilon(t) \\ \sum_{j=0}^{\infty} (\theta B)^j x(t) &= \epsilon(t) \\ x(t) &= \sum_{j=1}^{\infty} -(\theta B)^j x(t) + \epsilon(t) \\ x(t) &= \sum_{j=1}^{\infty} -\theta^j x(t-j) + \epsilon(t)\end{aligned}$$

- Methods of Moments nicht effizient, weil Modell nichtlinear in den Parametern.
- Betrachte dafür MA[1]-Prozeß:

YW-Gleichungen:

$$\begin{aligned}\gamma(0) &= \sigma^2(1 + \theta^2) \\ \gamma(1) &= \sigma^2\theta\end{aligned}$$

Es folgt:

$$\widehat{ACF}(1) = \frac{\hat{\theta}}{1 + \hat{\theta}^2}$$

und

$$\hat{\theta} = \frac{1 - \sqrt{1 - 4\widehat{ACF}(1)^2}}{2\widehat{ACF}(1)}$$

Da

$$\sqrt{N} \widehat{ACF}(1) \sim N(ACF(1), (1 - 2ACF(1)^2 + 4ACF(1)^4))$$

folgt mit Fehlerfortpflanzung:

$$\sigma^2(\theta) = \left(\frac{\partial \hat{\theta}}{\partial \widehat{ACF}(1)} \right)^2 (1 - 2ACF(1)^2 + 4ACF(1)^4)$$

und somit

$$\hat{\theta}_{YW} \sim N\left(\theta, \frac{1 + \theta^2 + 4\theta^4 + \theta^6 + \theta^8}{N(1 - \theta^2)^2}\right)$$

6.2.3 Innovation-Algorithmus

- Der Yule-Walker-Schätzer konstruiert einen Schätzer $\hat{x}(i)$ für $x(i)$ basierend auf der Basis $x(i-1), \dots, x(i-p)$.
- Eine andere und orthogonale Basis ist gegeben durch die "Innovationen" $x(i) - \hat{x}(i)$, i.e. das $\epsilon(i)$.

Es gilt:

$$\begin{aligned} \langle (x(i) - \hat{x}(i)) x(j) \rangle &= 0, \quad j = 1, \dots, i-1 \\ \langle (x(i) - \hat{x}(i)) (x(j) - \hat{x}(j)) \rangle &= 0, \quad j = 1, \dots, i-1 \end{aligned}$$

- Darstellung von $\hat{x}(i)$ in der Basis der Innovationen:

$$\hat{x}(i) = \sum_{j=1}^q \theta_{q,j} (x(i-j) - \hat{x}(i-j)) \quad (\ddagger)$$

$\theta_{q,j}$ ergeben sich aus der Rekursion

$$\begin{aligned} v_0 &= \gamma(0) \\ \hat{\theta}_{q,k} &= v_k^{-1} \left(\gamma(m-k) - \sum_{j=0}^{k-1} \hat{\theta}_{m,j} \hat{\theta}_{k,j} v_j \right) \quad k = 0, \dots, m-1 \\ v_m &= \gamma(0) - \sum_{j=0}^{m-1} \hat{\theta}_{m,j}^2 v_j \end{aligned}$$

- Beweis: Multipliziere (\ddagger) mit $(x(j) - \hat{x}(j))$ und nutze Orthogonalität.
- Asymptotik:

$$\hat{\theta}_{Inno} \sim N \left(\theta, \frac{1}{N} \right)$$

- Effizienz:

$$\theta = 0.5$$

$$Eff(Inno/YW) = 1/2.7$$

6.2.4 Maximum Likelihood Schätzer

Es gibt keine schöne geschlossene Formel.

- Betrachte MA[1]

$$x(t) = \epsilon(t) + \theta\epsilon(t-1)$$

In Analogie zu AR[1]

$$x(t) = ax(t-1) + \epsilon(t)$$

bei MA[1]

$$\hat{x}(t) = \theta\epsilon(t-1)$$

- Nichtlineare Optimierung von

$$S = \sum_i (x(i) - \hat{x}(i))^2 = \sum_i \hat{\epsilon}(i)^2$$

Da $\hat{x}(i)$ von $\hat{\epsilon}(i-1)$ abhängt, Selbstkonsistenzproblem

- Startwerte für die Parameter z.B. aus Innovationsalgorithmus.

Beispiel MA[1] (alle $\hat{\cdot}$ weggelassen):

-

$$S(\theta) = \sum_i \epsilon_i(\theta)^2$$

1. Ordnung Entwicklung:

$$\epsilon_i(\theta) \approx \epsilon_i(\theta_0) - (\theta - \theta_0)z_i(\theta_0)$$

mit

$$z_i(\theta_0) = \frac{\partial \epsilon_i(\theta_0)}{\partial \theta}$$

- Lineare Näherung eingesetzt:

$$Q(\theta) = \sum (\epsilon_i(\theta_0) - (\theta - \theta_0)z_i(\theta_0))^2$$

- Aus linearer Regression und Numerik:

$$\theta_{j+1} = \theta_j + \frac{\sum_i z_i(\theta_j)\epsilon_i(\theta_j)}{\sum_i z_i^2(\theta_j)} \quad \text{iterieren} \quad (\ddagger)$$

Woher $z_i(\theta_j)$?

- Schreibe MA-Prozeß

$$x_i = \epsilon_i + \theta\epsilon_{i-1}$$

formal um in:

$$\epsilon_i(\theta) = x_i - \theta\epsilon_{i-1}(\theta)$$

Partielle Ableitung nach θ , x_i als Daten gegeben, Produktregel:

$$z_i(\theta) = \epsilon_{i-1} - \theta z_{i-1}(\theta) \quad (\text{Startwerte } \epsilon_0 = z_0 = 0)$$

Damit (\ddagger) iterieren bis zur Konvergenz.

- Bei MA[1] schon kompliziert, i.a. nicht praktikabel
- Asymptotik MLE Schätzer:

$$\hat{\theta}_{MLE} \sim N\left(\theta, \frac{1 - \theta^2}{N}\right)$$

- Effizienz für $\theta = 0.5$:

$$Eff(MLE/YW) = 1/3.6$$

$$Eff(MLE/Inno) = 1/1.33$$

6.2.5 ARMA[p,q] Modelle

Es gibt keine straightforward Lösung. Geht nur über nichtlineare numerische Optimierung, z.B.:

- Startschätzungen für AR-Parameter aus Yule-Walker – Gleichungen
- Residuen aus AR[p]-Fit als Startschätzungen für $\epsilon(t)$
- Dann: Minimiere

$$\sum_t (\hat{\epsilon}(t))^2 = \sum_t (x(t) - (\sum_i a_i x(t-i) + \sum_i b_i \hat{\epsilon}(t-i)))^2$$

oder:

Fitte theoretisches Spektrum des ARMA[p,q] Prozesses an das Periodogramm, erinnere Whittle-Schätzer, Kap. 5.1.2, Gl. (5.9).

6.2.6 Diskret vs. Kontinuierlich III

Eine zeitkontinuierliche lineare stochastische DGL p -ter Ordnung ergibt einen ARMA[p,p-1] Prozess [220]. Beweis siehe unten.

Nebenbemerkung:

Als dynamische Modelle sind ARMA[p,q] eine Nullmenge, aber Spektraltyp Box/Jenkins Epoche 70er Jahre [33].

6.3 Nichtlineare Differenzgleichungen

- Es gibt keine natürliche Ordnung im Nichtlinearen.
- Eine naheliegende Erweiterung von linearen AR[p]-Prozessen ins Nichtlineare:

$$x(t) = \sum_{i=1}^p a_i x(t-i) + \sum_{i=1}^p \sum_{j=1}^p b_{ij} x(t-i)x(t-j) + \sum_{i=1}^p \sum_{j=1}^p \sum_{l=1}^p c_{ijl} x(t-i)x(t-j)x(t-l) + \dots + \epsilon(t)$$

- Anzahl N_p der Polynome in Abhängigkeit der Regressionsordnung p und des Polynomgrades K :

$$N_p = \binom{p+K}{K}$$

Parameterschätzung analog zur nichtlinearen Regression.

- Beispiel: $p=1$

Bezeichnung: $y_i = x(t)$, $x_i = x(t-1)$

$$E = \sum_i (y_i - \sum_k a_k x_i^k)^2$$

- Ableitung nach a_l ergibt:

$$\sum_l a_l \sum_i x_i^k x_i^l = \sum_i y_i x_i^k$$

Mit

$$g_{kl} = \sum_i x_i^k x_i^l = \sum_i x_i^{k+l}, \quad b_k = \sum_i y_i x_i^k$$

folgen Normalgleichungen

$$\sum_l g_{kl} a_l = b_k$$

- Probleme:
 - g_{kl} i.d.R. schlecht konditioniert
 - a_l hängen von K ab.
- Ausweg: Orthogonale Basis p_i bezüglich der Dichte $w(x)$ wählen.
Ergodizität ...

$$\langle p_i p_j \rangle = \int_x dx w(x) p_i(x) p_j(x) = \sum_t p_i(x(t)) p_j(x(t)) = \delta_{ij} c_{ij}$$

Dann g_{kl} diagonal.

Löst beide Probleme.

- Problem: $w(x)$ nicht bekannt.
- Lösung: Gram-Schmidt'sche Orthogonalisierung
oder rekursiv nach [78]

$$\begin{aligned}
 p_0(x_i) &= 1 \\
 p_1(x_i) &= x_i p_0(x_i) - \alpha_{10} p_0(x_i) \\
 p_j(x_i) &= x_i p_{j-1}(x_i) - \alpha_{j,j-1} p_{j-1}(x_i) - \alpha_{j,j-2} p_{j-2}(x_i) \\
 \alpha_{kl} &= \frac{\sum_i x_i p_{k-1}(x_i) p_l(x_i)}{\sum_i p_l(x_i)^2}
 \end{aligned}$$

Gibt empirisch orthogonale Polynome

- Siehe auch Stichwort:
Volterra modelling [186, 165, 256]
Für Anwendungen siehe: [89, 38]

6.3.1 Diskret vs. Kontinuierlich IV

Ist zu Grunde liegende Dynamik zeitkontinuierlich, ist die Interpretation der Parameter sehr schwierig, hängt z.B. der Sampling Rate ab.

- Für Vorhersage o.k.
- Für Verständnis schwierig

6.3.2 Diskret vs. Kontinuierlich V

Ist zu Grunde liegende Dynamik zeitkontinuierlich und das Rauschen Gauß, gilt dies nicht mehr für zeitdiskrete Dynamik, siehe Kap. 6.8.

Andere Basisfunktionen (remember: keine natürliche Ordnung):

- "Neuronale Netze", siehe andere Vorlesung

$$y_i = \sum_k w_k f(a_k x_i)$$

mit sigmoidalem $f(x_i)$, z.B.

$$f(x) = \text{atan}(x)$$

allgemein:

$$x(t) = \sum_k w_k f\left(\sum_l a_{kl} x(t-l)\right)$$

- Radiale Basis Funktionen

$$y_i = \sum_k w_k e^{-\frac{(x_i - \mu_k)^2}{\sigma^2}}$$

- ...

- "Vorteil" gegen Polynome:

Für Polynome gilt $|Pol(x)| \rightarrow \infty$ für $x \rightarrow \pm\infty$. NN und RBF bleiben beschränkt. Kann einem aber auch was vorgaukeln.

- Nachteile:

Nichtlinear in den Parametern: Muß numerisch, iterativ optimiert werden (Euphemismus: "Lernen")

Interpretierbarkeit der Parameter ganz schwierig.

6.4 EM Algorithmus

Literatur: [66, 278]

Expectation-Maximization Algorithmus:

- Iterative Berechnung von ML-Schätzern, wenn nicht alle Variablen beobachtet werden können.
- Bezeichnung:
 y_i : Beobachtete Variablen, x_i : Unbeobachtete Variablen, θ : Parameter

- Unser Beispiel

$$\begin{aligned}x(t) &= Ax(t-1) + \epsilon(t) \\y(t) &= Cx(t) + \eta(t)\end{aligned}$$

- Erinnerung: Bedingte Wahrscheinlichkeit, Pik As

$$p(x, y) = p(x|y)p(y)$$

- Voraussetzung: $p(x_{1\dots N}|y_{1\dots N}, \theta)$ muß berechenbar sein.

Jede Iteration $\theta^{(i)} \rightarrow \theta^{(i+1)}$ besteht aus zwei Schritten:

1. E-Step

Bilde den Erwartungswert von $\ln p(y_{1\dots N}, x_{1\dots N} | \theta)$ bezüglich der Dichte $p(x_{1\dots N} | y_{1\dots N}, \theta^{(i)})$:

$$A(\theta, \theta^{(i)}) = \int_{x_{1\dots N}} dx_{1\dots N} \ln p(y_{1\dots N}, x_{1\dots N} | \theta) p(x_{1\dots N} | y_{1\dots N}, \theta^{(i)})$$

2. M-Step

Maximiere $A(\theta, \theta^{(i)})$ bezüglich θ und setze

$$\theta^{(i+1)} = \arg \max_{\theta} A(\theta, \theta^{(i)}) .$$

Satz:

Dieses bewirkt einer Maximierung der ("incomplete data") Log-Likelihood :

$$L(\theta | y_{1\dots N}) = \ln p(y_{1\dots N} | \theta)$$

bezüglich θ .

Beweis: [328]

Zeige :

$$\mathcal{L}(\theta^{(i+1)} | y_{1\dots N}) \geq \mathcal{L}(\theta^{(i)} | y_{1\dots N})$$

Wegen

$$p(y_{1\dots N} | \theta) = \frac{p(x_{1\dots N}, y_{1\dots N} | \theta)}{p(x_{1\dots N} | y_{1\dots N}, \theta)}$$

gilt

$$\mathcal{L}(\theta | y_{1\dots N}) \equiv \ln p(y_{1\dots N} | \theta) = \ln p(x_{1\dots N}, y_{1\dots N} | \theta) - \ln p(x_{1\dots N} | y_{1\dots N}, \theta).$$

Multipliziere mit $p(x_{1\dots N} | y_{1\dots N}, \theta^{(i)})$, integriere über alle x_i (fällt für $\mathcal{L}(\theta | y_{1\dots N})$ raus). Ergibt:

$$\mathcal{L}(\theta | y_{1\dots N}) = A(\theta, \theta^{(i)}) - B(\theta, \theta^{(i)})$$

mit $A(\theta, \theta^{(i)})$ von oben und

$$B(\theta, \theta^{(i)}) = \int_{x_{1\dots N}} dx_{1\dots N} \ln p(x_{1\dots N} | y_{1\dots N}, \theta) p(x_{1\dots N} | y_{1\dots N}, \theta^{(i)}).$$

Nun gilt immer

$$B(\theta^{(i)}, \theta^{(i)}) \geq B(\theta, \theta^{(i)}).$$

Beweis:

Mit

$$g(x, y) = p(x | y, \theta), \quad f(x, y) = p(x | y, \theta^{(i)})$$

ist

$$B(\theta, \theta^{(i)}) - B(\theta^{(i)}, \theta^{(i)}) = \int dx f(x, y) \ln \frac{g(x, y)}{f(x, y)}.$$

Nun gilt aber per Taylorentwicklung mit $0 \leq \lambda \leq 1$

$$\ln \frac{g}{f} = \frac{g}{f} - 1 - \frac{(g/f - 1)^2}{2(1 + \lambda(g/f - 1)^2)},$$

und somit folgt :

$$\begin{aligned} B(\theta, \theta^{(i)}) - B(\theta^{(i)}, \theta^{(i)}) &= \int dx f \left[\frac{g}{f} - 1 - \frac{(g/f - 1)^2}{2(1 + \lambda(g/f - 1)^2)} \right] \\ &= - \int dx f \frac{(g/f - 1)^2}{2(1 + \lambda(g/f - 1)^2)} \leq 0. \end{aligned}$$

Damit folgt:

$$B(\theta, \theta^{(i)}) \leq B(\theta^{(i)}, \theta^{(i)}) .$$

Gleichheit bei $g = f$, d.h. Konvergenz.

Wir bilden nun:

$$\mathcal{L}(\theta | y_{1\dots N}) - \mathcal{L}(\theta^{(i)} | y_{1\dots N}) = (A(\theta, \theta^{(i)}) - A(\theta^{(i)}, \theta^{(i)})) + (B(\theta^{(i)}, \theta^{(i)}) - B(\theta, \theta^{(i)}))$$

Für

$$\theta^{(i+1)} = \arg \max_{\theta} A(\theta, \theta^{(i)})$$

gilt nun

$$A(\theta^{(i+1)}, \theta^{(i)}) \geq A(\theta^{(i)}, \theta^{(i)}) \quad B(\theta^{(i)}, \theta^{(i)}) \geq B(\theta^{(i+1)}, \theta^{(i)})$$

und somit

$$\mathcal{L}(\theta^{(i+1)} | y_{1\dots N}) \geq \mathcal{L}(\theta^{(i)} | y_{1\dots N}) .$$

$$\mathcal{L}(\theta^{(i+1)} | y_{1\dots N}) = \mathcal{L}(\theta^{(i)} | y_{1\dots N}) ,$$

wenn nix mehr zu optimieren ist.

Bemerkungen: :

- Die bedingte Dichte:

$$p(x_{1\dots N} | y_{1\dots N}, \theta)$$

zerfällt für

- unabhängige x_i in :

$$\prod_i p(x_i | y_{1\dots N}, \theta)$$

- 1. Ordnungs Prozesse x (modulo Anfangswert) in :

$$\prod_i p(x_i, x_{i+1} | y_{1\dots N}, \theta)$$

- Diese Berechnung hängt vom Modell ab:

- * Zustandsraummodell (Kap. 6.5): Kalman-Filter/Smoothing-Filter
- * Hidden Markov Modell (Kap. 6.7): Forward/Backward Algorithmus, Baum-Welsh Algorithmus

Häufig kann

$$\theta^{(i+1)} = \arg \max_{\theta} A(\theta, \theta^{(i)}) .$$

analytisch berechnet werden.

- Konkret sieht das Verfahren so aus:
 1. Wähle (falsche) Anfangswerte für $\theta^{(0)}$
 2. Schätze die unbeobachteten $x_j^{(i)}$ und ihre Verteilung (falsch)
 3. Nehme die Schätzungen ernst und schätze $\theta^{(i+1)}$ (falsch)
 4. gehe zu 2. bis Konvergenz (richtig)
- EM Algorithmus hat lineare Konvergenz
 - Vorteil gegen (quadratische) 'brute force optimierung': Constraints z.B. wegen
 - * Stationarität im ZRM
 - * Übergangswahrscheinlichkeiten a_{ij} im HMM: $0 \leq a_{ij} \leq 1$
 - * Positivität: Ratenkonstanten q_{ij} in HMMs $q_{ij} \geq 0$
 - * Zur Beschleunigung der Konvergenz, siehe: [141, 193, 142]
 werden automatisch eingehalten.
 - Häufig: Erst EM, dann z.B. Quasi-Newton

Zusammenfassung:

Der EM Algorithmus kommt in der Zeitreihenanalyse in den Fällen zur Anwendung, wenn wie beim HMM und ZRM im Prozeß zwei Arten von Stochastik vorliegen: Eine nicht beobachtbare stochastische Dynamik, die in der Beobachtung noch einmal verrauscht ist. Der EM-Algorithmus ist in diesem Falle nicht die einzig mögliche Lösung, aber die numerisch stabilste.

6.5 Zustandsraummodell

Literatur:

- R.H. Shumway and D.S. Stoffer: Time Series Analysis and Its Application [267]
- A. Gelb: Applied Optimal Estimation [87]
- A.C. Harvey: Forecasting structural time series models and the Kalman filter [109]

$$\begin{aligned}\vec{x}(t) &= A\vec{x}(t-1) + \vec{\epsilon}(t), & \vec{\epsilon}(t) &\sim N(0, Q) \\ \vec{y}(t) &= C\vec{x}(t) + \vec{\eta}(t), & \vec{\eta}(t) &\sim N(0, R)\end{aligned}$$

I.d.R.: $\vec{y}(t) = y(t)$, $R \in \mathbb{R}$

Spektrum:

$$S(\omega) = C(1 - Ae^{-i\omega})^{-1}Q \left((1 - Ae^{i\omega})^{-1} \right)^T C^T + R.$$

6.5.1 Gründe fürs ZRM

- AR[p] hat nicht genug Freiheitsgrade.

Betrachte AR[2] Prozeß mit Parametern, so daß zwei unabhängige Relaxatoren gegeben sind.

DoFs: 2 Relaxationszeiten + 2 Varianzen

Aber AR[2] hat nur 3 Dofs: a_1 , a_2 und σ^2 des treibenden Rauschens.

$$\begin{aligned}\vec{x}(t) &= \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix} \vec{x}(t-1) + \vec{\epsilon}(t), & \vec{\epsilon}(t) &\sim N(0, Q) & Q &= \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}, \\ \vec{y}(t) &= \begin{pmatrix} 1 \\ 1 \end{pmatrix} \vec{x}(t)\end{aligned}$$

- Natürliche Formulierung für ARMA[p,q] Prozesse

Aus

$$X(t) = \sum_{l=1}^p a_l X(t-l) + \epsilon(t) + \sum_{l=0}^q b_l \epsilon(t-l)$$

wird mit Zustandsvektor:

$$X(t) = (X_1(t), X_2(t), \dots, X_m(t))$$

mit $X_1(t) = X(t)$, $m = \max(p, q+1)$, $\mathbf{C} = (1, 0, \dots, 0)$ ein ZRM

$$X(t) = \mathbf{A}X(t-1) + \mathbf{B}\epsilon(t)$$

mit

$$\mathbf{A} = \begin{pmatrix} a_1 & 1 & 0 & \dots & 0 \\ a_2 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{m-1} & 0 & 0 & \dots & 1 \\ a_m & 0 & 0 & \dots & 0 \end{pmatrix},$$

$$\mathbf{B} = \begin{pmatrix} 1 \\ b_1 \\ \vdots \\ b_{m-2} \\ b_{m-1} \end{pmatrix},$$

wobei $a_i = 0$ für $i > p$ und $b_j = 0$ für $j > q$ gesetzt werde.

\mathbf{Q} ergibt sich als dyadisches Produkt von \mathbf{B} und ist singulär mit Rank 1.

- Das "Error-in-Variables" Problem [84, 44]

Betrachte AR[1] Prozeß

$$x(t) = ax(t-1) + \epsilon(t)$$

Parameter a kann unverzerrt geschätzt werden mit (Multipliziere mit $x(t-1)$ und bilde den Erwartungswert):

$$\hat{a} = \frac{\sum x(t-1)x(t)}{\sum x(t-1)x(t-1)} \quad . \quad (6.1)$$

Wenn die Beobachtung nur unter Rauschen möglich ist, gilt:

$$y(t) = x(t) + \eta(t), \quad \eta(t) \sim \mathcal{N}(0, R) \quad ,$$

Schätzt man a durch Gl. (6.1) basierend auf $y(t)$, ist der Erwartungswert von \hat{a}

$$\langle \hat{a} \rangle = \frac{\langle y(t-1)y(t) \rangle}{\langle y(t-1)y(t-1) \rangle} = a \frac{1}{1 + \frac{R}{\langle x(t)^2 \rangle}}$$

mit $\langle x(t)^2 \rangle = 1/(1-a^2)$.

Der Parameter a wird unterschätzt. Der bias hängt vom Signal-zu-Rausch Verhältnis ab. Im Regressionsfall heißt dies "Error-in-variables" - Problem [84]. Unterschätzung des funktionellen Zusammenhangs gilt auch für nichtlineare Differenzen-Prozesse, unabhängig davon, ob Polynome, NN oder RBF als Basis gewählt werden.

6.5.2 Eichfixierung

ZRM hat zwei Eichfreiheiten:

- Nichttriviale kontinuierliche Eichtransformation

$$\begin{aligned} A' &= LAL^{-1} \\ X' &= LX \\ \epsilon' &= L\epsilon \\ C' &= CL^{-1} \\ Q' &= LQL^{-1} \end{aligned}$$

Macht Grat in der Likelihood und damit Parameterschätzung unmöglich.

Fixierung durch $Q = 1$ oder $A = \text{Blockdiagonalgestalt}$.

– Triviale Freiheit: $X \rightarrow -X$, $C \rightarrow -C$ iss egal

6.5.3 Diskret vs. Kontinuierlich Part VI

Betrachte lineare stochastische DGL 2. Ordnung (o.B.d.A $\epsilon \sim \mathcal{N}(0, 1)$):

$$\ddot{x} = a\dot{x} + bx + \epsilon$$

Erster Ordnung System:

$$\begin{aligned}\dot{x}_1 &= x_2 \\ \dot{x}_2 &= ax_2 + bx_1 + \epsilon \\ y &= x_1\end{aligned}$$

Oder

$$\begin{aligned}\dot{\vec{x}}(t) &= A\vec{x}(t) + \vec{\epsilon}(t) \\ y(t) &= C\vec{x}(t)\end{aligned}$$

mit

$$A = \begin{pmatrix} 0 & 1 \\ a & b \end{pmatrix}, \quad \vec{\epsilon}(t) \sim \mathcal{N}(0, Q) \quad Q = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \quad C = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

Q singular mit Rank 1

Integration mit Euler-Verfahren

Erinnere: I.d.R: $\delta t \ll \Delta t$: Häufig iterieren zwischem dem Samplen

Eine Integration:

$$\begin{aligned}\vec{x}(t) &= \vec{x}(t - \delta t) + \delta t A \vec{x}(t - \delta t) + \sqrt{\delta t} \vec{\epsilon}(t) \\ y(t) &= C \vec{x}(t)\end{aligned}$$

Zwei Integrationen, führt auf :

$$\vec{x}(t) = \tilde{A}\vec{x}(t - 2\delta t) + \sqrt{2\delta t}\vec{\epsilon}(t - 2\delta t)$$

mit

$$\tilde{A} = (1 + \delta t A)^2, \quad \vec{\epsilon}(t) \sim \mathcal{N}(0, \tilde{Q})$$

\tilde{Q} ist voll besetzt und nicht singuläre Kovarianzmatrix.

Allgemein:

Kontinuierlicher AR[p] geht nach "vollem" ZRM.

Oft gute Näherung: ARMA[p,p-1] - Prozeß [220].

6.5.4 Kalman und Smoothing Filter

Das Wahre ist verborgen. Man muß es schätzen lernen.

Voraussetzung bei EM Algorithmus: $p(x_{1...N}|y_{1...N}, \theta)$ muß berechenbar sein.

Gegeben Beobachtungen $Y(t)$ und ein parametrisiertes ZRM.

- Das Kalman-Filter schätzt rekursiv die unbeobachteten $X(t)$ und ihre Verteilung: Da Gaußsch und linear: 1. und 2. Momente.
 - Von $p(x_{1...N}|y_{1...N}, \theta)$
 - wegen Prozess 1. Ordnung nach $\prod_t p(x_t x_{t-i}|y_{1...N}, \theta)$
 - wegen Gaußsch und linear: 1. und 2. Momente.
- Eingeführt : Kalman 1960 [148]
- Erste wichtige Anwendung: Apollo Mission [136]:
 y_i Radarmessungen, ZRM: Flugdynamik, x_i Position.

6.5.4.1 Die Idee des Kalman-Filters

$$\begin{aligned} X(t) &= \mathbf{A}X(t-1) + \varepsilon(t), \quad \varepsilon \sim N(0, Q) \\ Y(t) &= \mathbf{C}X(t) + \eta(t) \quad \eta \sim N(0, R) \end{aligned}$$

Eine optimale Schätzung $\hat{X}(t)$ für $X(t)$ basierend auf $Y(t')$, $t' = 1, \dots, t$ bedeutet, daß

- der Fehler $X(t) - \hat{X}(t)$ minimal ist
- $\langle (X(t) - \hat{X}(t)) Y(t') \rangle = 0$ gilt.

Der Schätzfehler muß also wieder senkrecht auf der von den $Y(t')$, $t = 1, \dots, t$ aufgespannten Ebene stehen, bzw. die Schätzfehler sollen unkorreliert sein mit der Beobachtung der Vergangenheit.

Bezeichnung: $X(t | \tilde{t})$: Schätzer $\hat{X}(t)$ unter Berücksichtigung der Beobachtungen $Y(t')$, $t' = 1, \dots, \tilde{t}$

Da $X(t | t-1)$ alle Information aus den Beobachtungen $Y(1)$ bis $Y(t-1)$ enthält, muß sich $X(t | t)$ aus $X(t | t-1)$ und $Y(t)$ ergeben.

Schätzvorgang in zwei Schritten:

1. Prädiktion von $X(t)$ vor der Beobachtung von $Y(t)$: $X(t | t-1)$

$$X(t | t-1) = \mathbf{A}X(t-1 | t-1). \quad (6.1)$$

Die Prädiktion $Y(t | t-1)$ basierend auf $X(t | t-1)$ ist dann

$$Y(t | t-1) = \mathbf{C}X(t | t-1). \quad (6.2)$$

2. Korrektur von $X(t | t-1)$ nach der Beobachtung $Y(t)$.

Korrektur wird proportional zum Prognosefehler der Beobachtung $Y(t) - Y(t | t-1)$ angesetzt.

Damit ergibt sich für $X(t | t)$:

$$X(t | t) = X(t | t-1) + \mathbf{K}(t)(Y(t) - Y(t | t-1)) \quad (6.3)$$

mit $\mathbf{K}(t)$: Kalman-Gain, noch zu ermittelnde Matrix.

Um die versteckte Größe X komplett zu charakterisieren, braucht man noch die Kovarianzmatrizen.

$$\begin{aligned}\mathbf{P}(t | t) &= \langle (X(t) - X(t | t)) (X(t) - X(t | t))^T \rangle \\ \mathbf{P}(t | t-1) &= \langle (X(t) - X(t | t-1))(X(t) - X(t | t-1))^T \rangle\end{aligned}$$

6.5.4.2 Herleitung des Algorithmus

Man erhält aus (6.3) für den Schätzfehler

$$X(t) - X(t | t) = X(t) - X(t | t-1) - \mathbf{K}(t)(Y(t) - Y(t | t-1)) \quad (6.4)$$

Mit Gln. (6.1) und (6.2) folgt

$$Y(t) - Y(t | t-1) = \mathbf{C}(X(t) - X(t | t-1)) + \eta(t) \quad (6.5)$$

und somit

$$X(t) - X(t | t) = (\mathbf{1} - \mathbf{K}(t) \mathbf{C})(X(t) - X(t | t-1)) + \mathbf{K}(t)\eta(t).$$

Für die Schätzfehler-Matrix $\mathbf{P}(t | t)$ ergibt sich somit

$$\mathbf{P}(t | t) = (\mathbf{1} - \mathbf{K}(t) \mathbf{C})\mathbf{P}(t | t-1)(\mathbf{1} - \mathbf{K}(t) \mathbf{C})^T + \mathbf{K}(t) \mathbf{R} \mathbf{K}(t)^T.$$

Wähle $\mathbf{K}(t)$ so, daß

$$\text{Spur}(\mathbf{P}(t | t)) = \langle (X(t) - X(t | t))^2 \rangle$$

minimal wird.

Führt auf:

$$(\mathbf{1} - \mathbf{K}(t) \mathbf{C})\mathbf{P}(t | t-1)\mathbf{C}^T = \mathbf{K}(t) \mathbf{R}$$

und somit folgt

$$\mathbf{K}(t) = \mathbf{P}(t | t-1)\mathbf{C}^T(\mathbf{C}\mathbf{P}(t | t-1)\mathbf{C}^T + \mathbf{R})^{-1}. \quad (6.6)$$

Die Rekursionsgleichungen für $\mathbf{P}(t | t)$ und $\mathbf{P}(t | t-1)$ erhält man wie folgt:
Aus

$$X(t) - X(t | t-1) = \mathbf{A}(X(t-1) - X(t-1 | t-1)) + \epsilon(t)$$

folgt

$$\begin{aligned}\mathbf{P}(t | t-1) &\equiv \langle (X(t) - X(t | t-1))(X(t) - X(t | t-1))^T \rangle \\ &= \mathbf{A}\mathbf{P}(t-1 | t-1)\mathbf{A}^T + \mathbf{Q}\end{aligned}$$

Aus (6.4) folgt weiterhin

$$\begin{aligned}P(t | t) &= P(t | t-1) \\ &\quad + [\mathbf{K}(t) \langle (Y(t) - Y(t | t-1))(Y(t) - Y(t | t-1))^T \rangle \mathbf{K}^T(t)] \\ &\quad - \langle (X(t) - X(t | t-1))(Y(t) - Y(t | t-1))\mathbf{K}^T(t) \\ &\quad - \mathbf{K}(t) \langle (Y(t) - Y(t | t-1))(X(t) - X(t | t-1)) \rangle \rangle \quad (6.7)\end{aligned}$$

Nun ist mit Gl.(6.5)

$$\langle (Y(t) - Y(t | t-1))(Y(t) - Y(t | t-1))^T \rangle = \mathbf{C}(t)\mathbf{P}(t | t-1)\mathbf{C}(t) + \mathbf{R}(t)$$

und

$$\langle (X(t) - X(t | t-1))(Y(t) - Y(t | t-1))^T \rangle = \mathbf{C}\mathbf{P}(t | t-1).$$

Der zweite und dritte Term auf der rechten Seite von (6.7) lauten nun

$$\mathbf{K}(t)(\mathbf{C}\mathbf{P}(t | t-1)\mathbf{C}^T + \mathbf{R})\mathbf{K}^T(t) - \mathbf{P}(t | t-1)\mathbf{C}^T\mathbf{K}^T(t)$$

und wegen des Ausdrucks (6.6) für $\mathbf{K}(t)$ heben sie sich gegenseitig auf. Es verbleibt

$$\mathbf{P}(t | t) = \mathbf{P}(t | t-1) - \mathbf{K}(t)\mathbf{C}\mathbf{P}(t | t-1) = (\mathbf{1} - \mathbf{K}(t)\mathbf{C})\mathbf{P}(t | t-1)$$

That's it.

6.5.4.3 Zusammenfassung und Bemerkungen

Zusammengefasst - Das Kalman-Filter:

$$\mathbf{P}(t | t-1) = \mathbf{A}\mathbf{P}(t-1 | t-1)\mathbf{A}^T + \mathbf{Q} \quad (6.8)$$

$$\mathbf{K}(t) = \mathbf{P}(t | t-1)\mathbf{C}^T(\mathbf{C}\mathbf{P}(t | t-1)\mathbf{C}^T + \mathbf{R})^{-1} \quad (6.9)$$

$$\mathbf{P}(t | t) = (\mathbf{1} - \mathbf{K}(t)\mathbf{C})\mathbf{P}(t | t-1) \quad (6.10)$$

$$X(t | t-1) = \mathbf{A}X(t-1 | t-1)$$

$$Y(t | t-1) = \mathbf{C}X(t | t-1)$$

$$X(t | t) = X(t | t-1) + \mathbf{K}(t)(Y(t) - Y(t | t-1))$$

Bemerkungen:

- Kalman-Filter liefert die wahrscheinlichste Trajektorie $X(t)$, gegeben die Daten und das Modell (Maximum a posteriori-Schätzer). Dieses ist i.a. keine typische Trajektorie, sondern eine geglättete.
- Die Größen $\mathbf{P}(t | t - 1)$, $\mathbf{K}(t)$ und $\mathbf{P}(t | t)$ können off-line berechnet werden.
- Gleichungen (6.8-6.10) stellen Fixpunktgleichungen für sie dar.
- Startwerte: $\mathbf{P}(0 | 0) = \text{Var}(Y)$, $X(0 | 0) = 0$
- Für $\mathbf{R} \rightarrow 0$ wird die Beobachtung genau. Dann folgt aus (6.6), falls $\mathbf{C}^{-1}(t)$ existiert, daß $\mathbf{K}(t) = \mathbf{C}^{-1}(t)$.

Aus Gl.(6.3) wird dann

$$X(t | t) = X(t | t - 1) + \mathbf{C}^{-1}(t)[Y(t) - \mathbf{C}X(t | t - 1)] = \mathbf{C}^{-1}(t)Y(t),$$

wie erwartet.

Für $\mathbf{R} \neq 0$, hat Kalman-geschätztes $X(t | t)$ immer kleinere Varianz als die naive Schätzung:

$$X(t | t) = \mathbf{C}^{-1}(t)Y(t),$$

- Kalman-Größen liefern auch Varianz der Vorhersagefehler von Y :

$$\Delta(t | t - 1) = \langle (Y(t) - Y(t|t - 1))^2 \rangle = \mathbf{C} \mathbf{P}(t | t - 1) \mathbf{C}^T + \mathbf{R}$$

mit anschaulicher Interpretation.

- Kalman-Gain sagt, wie sicher sich das Modell ob seines Zustands ist:

$$\mathbf{Q} = 0 (\mathbf{R} \neq 0) \implies \mathbf{P}(t | t - 1) \rightarrow 0 \implies \mathbf{K}(t) \rightarrow 0$$

in Anbetracht des Beobachtungsrauschens: $C = 1$:

$$K = \frac{P}{P + R} = \frac{P}{\Delta} = \frac{\text{interne Unsicherheit}}{\text{externe Unsicherheit}}$$

- Kalman-Filter ist online fähig.
- Kalman-Filter ist optimales (online) (Tiefpass-) Filter.
 $X(t|t)$ hinkt $X(t)$ hinterher.
- \mathbf{A} , \mathbf{Q} , \mathbf{C} und \mathbf{R} können auch zeitabhängig sein.

Übung: Rettet Riker

Anschauliche Erklärung analog zu Kap. 5.2.3 Graphische Modelle : Understanding Kalman filter paper [191]

6.5.4.4 Das Glättungsfilter

Kalman Filter liefert $X(t | t)$.

”Rückwärts-Kalman-Filter”, ”Smoothing-Filter” liefert $X(t | N)$.

Mit den (End-)Startwerten

$$X(N | N), \quad \mathbf{P}(N | N)$$

und

$$\mathbf{B}(t) := \mathbf{P}(t | t) \mathbf{A}^T \mathbf{P}^{-1}(t + 1 | t)$$

ergeben sich die Rückwärtsrekursionen

$$X(t | N) = X(t | t) + \mathbf{B}(t)(X(t + 1 | N) - X(t + 1 | t)), \quad t = N - 1, \dots, 1$$

und

$$\mathbf{P}(t | N) = \mathbf{P}(t | t) + \mathbf{B}(t)(\mathbf{P}(t + 1 | N) - \mathbf{P}(t + 1 | t))\mathbf{B}^T(t).$$

$Y(t)$ gehen nicht direkt ein, sondern über die ”von oben” kommenden $X(t | N)$.

Anwendung: [136]

Während $X(t|t)$ hinter $X(t)$ herhinkt, zeigt $X(t|N)$ keinen Delay.

6.5.5 Parameterschätzung im Zustandsraummodell

Literatur: [192] [266]

Wähle ZRM mit Eichfixierung :

$$\begin{aligned} X(t) &= \mathbf{A} \mathbf{X}(t-1) + \epsilon(t), & \epsilon(t) &\sim WN(\mathbf{0}, \mathbf{1}) \\ Y(t) &= \mathbf{C} \mathbf{X}(t) + \eta(t), & \eta(t) &\sim WN(\mathbf{0}, \mathbf{R}). \end{aligned}$$

mit

$$X(t) \in R^m, \quad Y(t) \in R^n$$

Parameter:

$$\Theta = \{R_{ij}, C_{i\alpha}, A_{\alpha\beta}\}.$$

Damit lautet die (online-)Likelihoodfunktion

$$\begin{aligned} P_Y(Y_N; \Theta) &= (2\pi)^{-Nn/2} \prod_{t=1}^N (\det \mathbf{\Delta}(t | t-1))^{-1/2} \\ &\quad \cdot \exp\left\{-\frac{1}{2}(Y(t) - Y(t | t-1))^T (\mathbf{\Delta}(t | t-1))^{-1} (Y(t) - Y(t | t-1))\right\} \end{aligned}$$

Diese lässt sich numerisch iterativ optimieren:

- Wähle Startwerte für Θ
- Nutze Kalman-Filter zur Berechnung von $Y(t | t-1)$ und $\mathbf{\Delta}(t | t-1)$

Optimiere mit :

- Powell (ohne Ableitungen)
- Quasi-Newton (mit Ableitungen (numerisch oder analytisch))
- ...

Problem: Stationaritätsbedingungen an \mathbf{A} sind schwer zu gewährleisten.

Besser :

6.5.5.1 EM - Algorithmus

Vorbemerkung:

Die Abhängigkeit von $\mathbf{P}(0 | 0)$ und $X(0 | 0)$ wird unterdrückt.

Strenggenommen: Conditional Likelihood

Betrachte:

$$P_{Y,X}(Y_N, X_N; \Theta) = P_{Y|X}(Y_N | X_N; \Theta) P_X(X_N | \Theta)$$

mit

$$\begin{aligned} P_X(X_N | \Theta) &= \\ & (2\pi)^{-Nm/2} \prod_{t=1}^N \cdot \exp\left\{-\frac{1}{2}(\mathbf{X}(t) - \mathbf{A} \mathbf{X}(t-1))^T (\mathbf{X}(t) - \mathbf{A} \mathbf{X}(t-1))\right\} \\ P_{Y|X}(\mathbf{Y}_N | \mathbf{X}_N; \Theta) &= \\ & (2\pi)^{-Nn/2} \prod_{t=1}^N (\det \mathbf{R})^{-1/2} \exp\left\{-\frac{1}{2}(\mathbf{Y}(t) - \mathbf{C} \mathbf{X}(t))^T \mathbf{R}^{-1}(\mathbf{Y}(t) - \mathbf{C} \mathbf{X}(t))\right\}. \end{aligned}$$

(erinnere $Q=1$)

Ergibt (modulo egalere Konstanten):

$$\begin{aligned} \ln P_{Y,X}(Y_N, X_N; \Theta) &= -\frac{1}{2}N \ln(\det \mathbf{R}) \\ & - \frac{1}{2} \sum_{t=1}^N (\mathbf{Y}(t) - \mathbf{C} \mathbf{X}(t))^T \mathbf{R}^{-1}(\mathbf{Y}(t) - \mathbf{C} \mathbf{X}(t)) \\ & - \frac{1}{2} \sum_{t=1}^N (\mathbf{X}(t) - \mathbf{A} \mathbf{X}(t-1))^T (\mathbf{X}(t) - \mathbf{A} \mathbf{X}(t-1)) \end{aligned}$$

E-Step des EM-Algorithmus':

- Bilde Erwartungswert von $\ln P_{Y,X}(Y_{1..N}, X_{1..N}; \Theta)$ bezüglich $P(X_{1..N} | Y_{1..N}; \Theta^{(i)})$
- Benutze Ergoden-Theorem :

$$\int dx p(x|y, \Theta) \ln p(x, y|\Theta) = \sum_t \ln p(x, y|\Theta)$$

- Ergibt : $A(\Theta, \Theta^{(i)})$.
- Gegeben Y_N und $\Theta^{(i)}$, liefert Kalman- und Glättungsfilter $X(t | N)$ und $P(t|N)$

Im einzelnen:

$$\langle X(t) \rangle_{Y_N; \Theta^{(i)}} = X(t | N) \quad (6.11)$$

und

$$\begin{aligned} \langle X(t)X(t)^T \rangle_{Y_N; \Theta^{(i)}} &= \langle (X(t) - X(t | N))(X(t) - X(t | N))^T \rangle_{Y_N; \Theta^{(i)}} \\ &\quad + X(t | N)X(t | N)^T \\ &= \mathbf{P}(t | N) + X(t | N)X(t | N)^T. \end{aligned} \quad (6.12)$$

Der Rechenrick im Detail

- Betrachte

$$\begin{aligned} \mathbf{P}(t | N) &= \langle (X(t) - X(t | N))(X(t) - X(t | N))^T \rangle_{Y_N; \Theta^{(i)}} \\ &= \langle X(t) X(t)^T \rangle_{Y_N; \Theta^{(i)}} \\ &\quad - \langle X(t) X(t | N)^T \rangle_{Y_N; \Theta^{(i)}} \\ &\quad - \langle X(t | N) X(t)^T \rangle_{Y_N; \Theta^{(i)}} \\ &\quad + \langle X(t | N)X(t | N)^T \rangle_{Y_N; \Theta^{(i)}} \end{aligned}$$

- $X(t | N)$ läßt sich jeweils aus Erwartungswert rausziehen

$$\begin{aligned} &= \langle X(t) X(t)^T \rangle_{Y_N; \Theta^{(i)}} \\ &\quad - \langle X(t) \rangle_{Y_N; \Theta^{(i)}} X(t | N)^T \\ &\quad - X(t | N) \langle X(t)^T \rangle_{Y_N; \Theta^{(i)}} \\ &\quad + X(t | N)X(t | N)^T \end{aligned}$$

Mit Gl. (6.11) folgt Gl. (6.12)

Analog ergibt sich

$$\begin{aligned} \langle X(t)X(t-1)^T \rangle_{Y_N; \Theta^{(i)}} &= \langle (X(t) - X(t | N))(X(t-1) - X(t-1 | N))^T \rangle_{Y_N; \Theta^{(i)}} \\ &\quad + X(t | N)X(t-1 | N)^T \\ &= \mathbf{B}(t | N)\mathbf{P}(t-1 | N) + X(t | N)X(t-1 | N)^T. \end{aligned}$$

Beachte:

- $\mathbf{B}(t | N)$, $\mathbf{P}(t | N)$ und $X(t | N)$ sind mit Parametersatz $\Theta^{(i)}$ berechnet wurden.
- \mathbf{R} , \mathbf{A} , \mathbf{C} gehören zum Parametersatz Θ .

Mit den Abkürzungen

$$\begin{aligned}\mathbf{Q}^{(1)} &= \sum_{t=1}^N Y(t)Y(t)^T \\ \mathbf{Q}^{(2)} &= \sum_{t=1}^N X(t | N)Y(t)^T \\ \mathbf{Q}^{(3)} &= \sum_{t=1}^N \langle X(t)X(t)^T \rangle_{Y_N; \Theta^{(i)}} \\ \mathbf{Q}^{(4)} &= \sum_{t=2}^N \langle X(t-1)X(t)^T \rangle_{Y_N; \Theta^{(i)}} \\ \mathbf{Q}^{(5)} &\approx \mathbf{Q}^{(3)} = \sum_{t=2}^N \langle X(t-1)X(t-1)^T \rangle_{Y_N; \Theta^{(i)}}\end{aligned}$$

erhält man somit für die zu maximierende Funktion

$$\begin{aligned}A(\Theta, \Theta^{(i)}) &= -\frac{1}{2}N \ln(\det \mathbf{R}) \\ &\quad - \frac{1}{2}\text{Tr}(\mathbf{R}^{-1}(\mathbf{Q}^{(1)} - \mathbf{C}\mathbf{Q}^{(2)} - \mathbf{Q}^{(2)T}\mathbf{C}^T + \mathbf{C}\mathbf{Q}^{(3)}\mathbf{C}^T)) \\ &\quad - \frac{1}{2}\text{Tr}(\mathbf{Q}^{(3)} - \mathbf{A}\mathbf{Q}^{(4)} - \mathbf{Q}^{(4)T}\mathbf{A}^T + \mathbf{A}\mathbf{Q}^{(5)}\mathbf{A}^T)\end{aligned}$$

M-Step des EM-Algorithmus:

Abhängigkeit $A(\Theta, \Theta^{(i)})$ von Parametern \mathbf{A} , \mathbf{C} , \mathbf{R} so einfach, daß man Maximum analytisch bestimmen kann.

Tricks, z.B.:

$$\frac{\partial \ln(\det \mathbf{R})}{\partial R_{ij}} = \text{Tr}(\mathbf{R}^{-1} \frac{\partial \mathbf{R}}{\partial R_{ij}})$$

Das sehr anschauliche Ergebnis:

$$\mathbf{A} = (\mathbf{Q}^{(4)})^T (\mathbf{Q}^{(5)})^{-1} = \sum_{t=1}^N \langle X(t-1)X(t)^T \rangle_{Y_N; \Theta^{(i)}} / \sum_{t=1}^N \langle X(t-1)X(t-1)^T \rangle_{Y_N; \Theta^{(i)}}$$

$$\mathbf{C} = (\mathbf{Q}^{(2)})^T (\mathbf{Q}^{(3)})^{-1} = \sum_{t=1}^N X(t | N)Y(t)^T / \sum_{t=1}^N \langle X(t)X(t)^T \rangle_{Y_N; \Theta^{(i)}}$$

$$\begin{aligned} \mathbf{R} &= \frac{1}{N} (\mathbf{Q}^1 - \mathbf{C}\mathbf{Q}^2 - \mathbf{Q}^{2^T} \mathbf{C}^T + \mathbf{C}\mathbf{Q}^3 \mathbf{C}^T) \\ &= \frac{1}{N} \sum_{t=1}^N (Y(t) - \mathbf{C}\mathbf{X}(t | N))(Y(t) - \mathbf{C}\mathbf{X}(t | N))^T + \mathbf{C}\mathbf{P}(t | N)\mathbf{C}^T \end{aligned}$$

- Mit diesem Parameterupdate auf zum E-Step des nächsten Iterationsschritt im EM-Algorithmus
- Endeder EM-Iteration, wenn relative Änderung der Parameter klein, i.e. z.B. 10^{-6}
- Bemerkung: Nicht über relative Änderung der Likelihood abbrechen, kann "Durststrecken" geben.

Modellwahl Kriterien

- Residuen weiß (KS-Test)
- Residuen Gausssch (qq-plot)
- Vergleiche Spektrum des Modells mit Periodogramm
- Knick in der erklärten Varianz

Anwendung besprechen : "Modeling noisy time series" [284]

FOLIEN PHYSTREM

6.5.5.2 Nichtlineare ZRMs & Extended Kalman Filter

[87, 109]

Für nichtlineare ZRMs:

$$\begin{aligned}\vec{x}(t) &= \vec{f}(\vec{x}(t-1)) + \vec{\epsilon}(t) \\ y(t) &= g(\vec{x}(t)) + \eta(t)\end{aligned}$$

läßt sich das (extended) Kalman-Filter nutzen, wenn man setzt:

$$\begin{aligned}\mathbf{A} &= \left. \frac{\partial \vec{f}}{\partial \vec{x}} \right|_{\vec{x}=\vec{x}_{t-1}|t-1} \\ \mathbf{C} &= \left. \frac{\partial g}{\partial \vec{x}} \right|_{\vec{x}=\vec{x}_t|t-1}\end{aligned}$$

- Dies führt i.d.R zu verzerrten Schätzern.
- Ausmaß der Verzerrung hängt vom Modell ab.
- Grund:
Auf Grund der Nichtlinearität ist (lineare) Beschreibung durch Mittelwert und (Co-)varianzen nicht mehr hinreichend.

ZEICHNUNG Auswirkung der Krümmung

- Ferner: Man muß pro E-Step Kalman- und Smoothing-Filter mehrfach anwenden.

6.5.5.3 Unscented filtering

[146, 145, 319, 306] (“Unparfümiert”, da keine Linearisierung)

Verbesserung des Extended Kalman Filters:

Statt Beschreibung durch

- lineare Näherung der Abbildung
- Anwendung auf Mittelwert und (Co-)varianzen

hier

- exakte nichtlineare Abbildung von z.B. 5 und 95% Quantilen (called "σ Punkte")
- Konstruktion einer Gaußverteilung (speziell Mittelwert) aus Bildern der σ Punkte

6.5.5.4 Joined Approach

Nehme Parameter in Dynamik auf mit

$$\Theta(t) = \Theta(t - 1)$$

Update über Kalman-Filter.

... ausführen ...

ZEICHNUNG zu Unscented Filtering

Scheint besser zu sein.

6.5.5.5 Markov Chain Monte Carlo (MCMC) Methoden

Literatur:

- Allgemein: [88]
- Zeitreihen: [212]
- ZRM: Hürzeler-Diss [135]
- Bayes'scher Ansatz: Es gibt keine wahren Parameter. Parameter sind auch Zufallsvariablen
- Gegeben Daten, Modellstruktur und prior-Verteilung für Parameter.
- Formuliere stochastischen Prozeß im Parameterraum, dessen stationäre Verteilung die (richtige) Dichte des Schätzers ist.
- Sample aus der stationären Verteilung.

6.5.5.6 Zustandsraummodell als Trendschätzer

Literatur: [160, 161]

$$\begin{aligned}x(t) &= x(t-1) + \epsilon(t) \\ y(t) &= x(t) + \eta(t)\end{aligned}$$

- Dynamische Gleichung sieht aus wie Brownian Motion
- Ist es aber de facto nicht.
- Ist Constraint auf die erste Ableitung:

$$x(t) - x(t-1) = \epsilon(t)$$

Bedeutung

- Modell versucht, die Varianz der Daten in einen Trend und Beobachtungsrauschen zu zerlegen.
- Die spezielle Wahl dieser Lösung dieses schlecht-gestellt inversen Problems hat etwas Magisches.

Auch höhere Ableitungen sind möglich. Beispiel 2. Ableitung

$$\begin{aligned}\vec{x}(t) &= A\vec{x}(t-1) + \vec{\epsilon}(t) \\ y(t) &= (1, 0)\vec{x}(t) + \eta(t)\end{aligned}$$

mit

$$A = \begin{pmatrix} 2 & -1 \\ 1 & 0 \end{pmatrix}, \quad \vec{\epsilon}(t) \sim \mathcal{N}(0, Q) \quad Q = \begin{pmatrix} \sigma^2 & 0 \\ 0 & 0 \end{pmatrix}$$

Kommt aus

$$\ddot{x} \approx x(t+1) - 2x(t) + x(t-1) \quad x(t+1) = 2x(t) - x(t-1) + \epsilon(t+1)$$

Ferner geht

$$\begin{aligned}a(t) &= a(t-1) + \epsilon(t) \\ y(t) &= a(t)y(t-1) + \eta(t)\end{aligned}$$

was einen zeitvariablen AR[1] Prozeß darstellt.

Erlaubt zwei-abhängige AR-Spektren

Zur Theorie langsam instationärer Systeme, siehe [62]

Ist der Prozeß stationär, liefert er einen Schätzer für a

6.6 Markov Prozesse

6.6.1 Motivation, Definition, Eigenschaften

Markov Prozeß: s diskrete Zustände ("Alphabet")

Beispiele:

- Gen-Sequenzen (A,C,G,T) [70]
- Sprache (Phoneme: "sch", "te")
- Ionenkanäle (Konfirmationen) [118, 250]

FOLIE Ionenkanal und patch clamp

Markov-Bedingung (1. Ordnung):

Alle verfügbare Information über $x(t+1)$ ist durch $x(t)$ gegeben:

$$p(x(t+1)|x(t), x(t-1), \dots) = p(x(t+1)|x(t))$$

Auch vektorielle AR[1]-Prozesse und ODEs fallen darunter¹

Zentrale Größe für Markov Prozesse: Übergangswahrscheinlichkeiten

$$a_{ij} = p(x(t+1) = j | x(t) = i), \quad i, j = 1, \dots, s,$$

mit Constraint:

$$\sum_j a_{ij} = 1, \quad \forall i, \quad \text{oder } A \mathbf{1} = \mathbf{1}$$

Formale Analogie zu AR[1] [102]

¹Mitunter werden auch Prozesse, die die Markov Bedingung erfüllen als Markov-Prozesse bezeichnet, nicht so hier.

Erinnere: AR Model

$$p(x(t+1)|x(t)) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x(t+1)-ax(t))^2}{2\sigma^2}}$$

Betrachte s -dimensionale ZV $\vec{\xi}(t)$:

$$\xi(t)_i = 1, \text{ if } x(t) = i, \quad \xi(t)_j = 0 \text{ für } j \neq i$$

(Bemerkung: Liegevektoren)

Wenn $x(t) = i$, dann

$$Prob(\xi(t+1)_j = 1 | x(t) = i) = a_{ij}$$

damit:

$$\langle (\vec{\xi}(t+1) | \vec{\xi}(t) \rangle = \vec{\xi}(t) A$$

Auf Grund der Markov-Bedingung gilt:

$$\vec{\xi}(t+1) = \vec{\xi}(t) A + \vec{\nu}(t) \tag{6.13}$$

resp.

$$\vec{\xi}^T(t+1) = A^T \vec{\xi}^T(t) + \vec{\nu}^T(t)$$

mit merkwürdigem, diskret-wertigen, aber weißen Rauschen

$$\vec{\nu}(t) = \vec{\xi}(t) - \langle (\vec{\xi}(t+1) | \vec{\xi}(t) \rangle$$

mit komponentenweisem Mittelwert 0.

Übung:

Zeige für ein 3-Zustandsmodell, daß $\langle \nu_i(t) \rangle = 0$ gilt.

Kovarianzmatrix von $\vec{\nu}(t)$ ist singular.

Beweis:

Da

$$\vec{\xi}(t) \mathbf{1} = 1, \quad A \mathbf{1} = \mathbf{1}$$

folgt aus Gl.(6.13) :

$$\vec{\nu}(t) \mathbf{1} = 0$$

6.6.1.1 Stationäre Dichte

Der Erwartungswert von $\vec{\xi}(t)$ definiert die stationäre Dichte

$$\pi = \langle \vec{\xi}(t) \rangle$$

Aus Gl. (6.13) folgt somit:

$$\langle \vec{\xi}(t+1) \rangle = \langle \vec{\xi}(t) \rangle A \implies \pi = \pi A$$

- A hat einen Eigenwert 1.
- Trotzdem ist der analoge AR[1] Prozeß stationär, da die singuläre Richtung der Kovarianzmatrix grade zu dem zugehörigen Eigenvektor gehört, siehe [102].
- Alle anderen Eigenwerte sind < 1 und reell.

Betrachte ZRM (ohne Beobachtungsrauschen):

$$\begin{aligned}\vec{\xi}^T(t) &= A^T \vec{\xi}^T(t-1) + \vec{v}^T(t) \\ y(t) &= (1, 2, \dots, s) \vec{\xi}^T(t)\end{aligned}$$

Damit:

Merke:

Spektrum (Markov-Prozess(1.Ordnung)) = Spektrum (AR[1]) mit lauter Relaxatoren.

Instationaritätsanekdote

Analogon zu ACF bei kontinuierlichem Zustandsraum:

Aufenthaltsdauer-Verteilung.

Markov Prozess macht exponential-verteilte Aufenthaltsdauern, da

$$Prob(x(t) = i \text{ für } t = 1, \dots, \tau) \propto a_{ii}^\tau$$

Nichtmarkovprozesse, z.B.: fraktale Modelle ($D > 1$)

$$\begin{aligned}\tilde{a}_{ii}(t_{in}) &= 1 - (1 - a_{ii}) t_{in}^{1-D} \\ \tilde{a}_{ij}(t_{in}) &= a_{ij} t_{in}^{1-D} \quad i \neq j \quad ,\end{aligned}$$

produziert algebraisch-verteilte Aufenthaltsdauern.

6.6.1.2 Aggregierte Markov Prozesse

Mindestens 2 Zustände sind an Ausgabe nicht unterscheidbar.

ZEICHNUNG Aggregierte Modelle

Unterscheidung dynamisch

m Zustände aggregiert : Aufenthaltsdauer des aggregierten Zustands m-exponentiell verteilt:

$$Prob(x_{aggr}(t) = i \text{ für } t = 1, \dots, \tau) = \sum_{j=1}^m c_j e^{\tau/\gamma_j}$$

6.6.1.3 Topologie eines Markovprozesses

Erlaubte und verbotene Übergänge

ZEICHNUNG Topologie

Erläuterung an Ionenkanälen

Betrachtet man einen Markov Prozeß 1. Ordnung in kontinuierlicher Zeit, so gilt für die Wahrscheinlichkeit $P_i(t)$ die Master-Gleichung:

Hier einfach:

$$\dot{P}_i = \sum_j q_{ji} P_j - \sum_j q_{ij} P_i$$

mit anschaulicher Bedeutung des Zu- und Abflusses von Wahrscheinlichkeit. Dieses läßt sich zusammenfassen zu:

$$\dot{P}_i = \sum_j P_j q_{ji}$$

mit Normierung:

$$q_{ii} = - \sum_{j \neq i} q_{ij}, \quad \forall i$$

Integration gibt den Zusammenhang zwischen Ratenkonstantenmatrix $Q = q_{ij}$ und Übergangswahrscheinlichkeitsmatrix $A = a_{ij}$

$$A = \exp(Q\Delta t) = 1 + Q\Delta t + \frac{1}{2}(Q\Delta t)^2 + \dots$$

- Effektiv zu berechnen per Padé Approximation [205]
- Paper: 19 dubious ways to compute the exponential of a matrix [201, 202]
- Bemerkung: Die Umkehrung ist nicht einfach, da Logarithmus von Matrix nicht straightforward.

6.6.1.4 Diskret vs. kontinuierlich VII

- Ist Prozeß zeitkontinuierlich und hat Topologie mit verbotenen Übergängen ($q_{ij} = 0$ für einige ij), so ist a_{ij} i.a. voll besetzt.
- Liegt wieder am Hochfrequenten.
- Begründen. Es gibt gesampelt immer alle möglichen Übergänge.
- Zeit-diskret kann man gar nicht von Topologie reden, da es keine verbotenen Übergänge gibt.
- Merke:
Zeitdiskretisierte kontinuierliche Dynamik kann keine Topologie entscheiden.

6.6.2 Parameterschätzung in (nicht aggregierten) Markov Modellen

Da a_{ij} Übergangswahrscheinlichkeiten darstellen, gelten die Nebenbedingungen:

$$0 \leq a_{ij} \leq 1 \quad (6.14)$$

$$\sum_j a_{ij} = 1 \quad . \quad (6.15)$$

Mit der Startverteilung π_i ist die Likelihood einer Realisierung gegeben durch:

$$\mathcal{L}(x_1, \dots, x_N | a_{ij}) = \pi_{x_1} \prod_t a_{x_t x_{t+1}}$$

Unter Vernachlässigung des Anfangszustandes ergibt Logarithmieren :

$$L(x_1, \dots, x_N | a_{ij}) = \sum_t \log a_{x_t x_{t+1}} = \sum_{i=1}^s \sum_{j=1}^s n_{ij} \log a_{ij}$$

mit

$$n_{ij} = \sum_t 1_{(x_t=i, x_{t+1}=j)} \quad (6.16)$$

Maximieren der Likelihood unter Berücksichtigung von $\sum_j a_{ij} = 1$ per Lagrange Multiplern λ_i :

$$\tilde{L}(x_1, \dots, x_N | a_{ij}) = \sum_{i=1}^s \sum_{j=1}^s n_{ij} \log a_{ij} + \sum_i \lambda_i (\sum_j a_{ij} - 1)$$

Ableitung nach a_{ij} ergibt:

$$n_{ij} \frac{1}{\hat{a}_{ij}} + \lambda_i = 0$$

$$\hat{a}_{ij} = \frac{n_{ij}}{-\lambda_i}$$

Die Nebenbedingung:

$$\sum_j \hat{a}_{ij} = 1 = \frac{\sum_j n_{ij}}{-\lambda_i}$$

führt auf die anschauliche Gleichung:

$$\hat{a}_{ij} = \frac{n_{ij}}{\sum_j n_{ij}} \quad (6.17)$$

Gl. (6.17) wird bei EM-Algorithmus für HMMs entsprechend zu Gl. (6.1) wieder auftauchen.

6.6.3 Identifizierbarkeit

6.6.3.1 Modellidentifizierbarkeit

Erinnerung aggregierte MM in Ionenkanal-Nomenklatur

ZEICHNUNG C-C-O, C-O-C und Kreis

Haben zwei Modelle die gleiche Likelihood, sind sie nicht zu unterscheiden. Die Modelle sind nichtidentifizierbar.

Übergangswahrscheinlichkeiten A

$$A = \exp(Q\Delta t) = \sum_{k=0}^{\infty} \frac{(Q\Delta t)^k}{k!}$$

mit Generatormatrix Q (zur praktischen Exponentiation der Q -Matrix, siehe [201, 205]).

$$Q = \left(\begin{array}{c|c} Q_{OO} & Q_{OC} \\ \hline Q_{CO} & Q_{CC} \end{array} \right)$$

und Zuständen:

$$(x_1(t) = O, \dots, x_{n_O}(t) = O; x_{n_O+1}(t) = C, \dots, x_{n_O+n_C}(t) = C)$$

Für nicht aggregierte MM gilt:

$$\mathcal{L}(x_1, \dots, x_N | a_{ij}) = \pi_{x_1} a_{x_1 x_2} \dots a_{x_{N-1} x_N}$$

Für aggregierte MM gilt:

$$\mathcal{L}(x_1, \dots, x_N | a_{ij}) = \sum_{x_1, \dots, x_N \in \{O, C\}} \pi_{x_1} a_{x_1 x_2} \dots a_{x_{N-1} x_N} \mathbf{1}, \quad \text{mit } \mathbf{1} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \quad (6.18)$$

Theorem [157]:

Seien Q und Q' zwei Generator-Matrizen von Markov Modellen. Existiert eine Ähnlichkeitstransformation

$$Q' = S^{-1}QS$$

mit

$$S = \left(\begin{array}{c|c} S_{OO} & 0 \\ \hline 0 & S_{CC} \end{array} \right)$$

und

$$\sum_j S_{ij} = 1 \quad \forall i$$

sind die Modelle nicht identifizierbar.

Comment: Dieses entspricht den nichttrivialen Eichfreiheiten des ZRMs, nur sind die Auswirkungen hier fundamentaler.

Beweis:

Aus

$$A = \exp(Q\Delta t) = \sum_{k=0}^{\infty} \frac{(Q\Delta t)^k}{k!}$$

folgt

$$A' = S^{-1}AS$$

Mit

$$\pi' = \pi S \quad \text{Bemerkung: ändert } \pi_O \text{ und } \pi_C \text{ nicht}$$

gilt für die Summanden in Gl. (6.18):

$$\begin{aligned} & \pi'_{x_1} a'_{x_1 x_2}, \dots, a'_{x_{N-1} x_N} \mathbf{1} \\ &= (\pi_{x_1} S) (S^{-1} a_{x_1 x_2} S), \dots, (S^{-1} a_{x_{N-1} x_N} S) \mathbf{1} \\ &= \pi_{x_1} a_{x_1 x_2}, \dots, a_{x_{N-1} x_N} \mathbf{1} \end{aligned}$$

Nichtunterscheidbare Modelle produzieren (natürlich) identische Verweilzeitverteilungen.

Beispiel:

Betrachte C-C-O Modell

ZEICHNUNG C-C-O mit a,b,c,d

mit Generator-matrix $Q^{(C-C-O)}$

$$Q^{(C-C-O)} = \left(\begin{array}{c|cc} -d & 0 & d \\ \hline 0 & -a & a \\ c & b & -(b+c) \end{array} \right)$$

Die Eigenwerte der Q_{CC} Submatrix sind:

$$\lambda_{1,2} = \frac{1}{2}(a+b+c) \pm \sqrt{(a+b+c)^2 - 4ac}$$

Wähle Ähnlichkeitsmatrix S :

$$S = \begin{pmatrix} 1 & 0 \\ 0 & S_{cc} \end{pmatrix}$$

mit

$$S_{cc} = \frac{1}{\lambda_1 - \lambda_2} \begin{pmatrix} -\lambda_2 & \lambda_1 \\ c - \lambda_2 & \lambda_1 - c \end{pmatrix}$$

ergibt sich:

$$\begin{aligned} S^{-1}Q^{(C-C-O)}S &= \\ &= \begin{pmatrix} -d & \frac{d(c-\lambda_2)}{\lambda_1-\lambda_2} & \frac{d(\lambda_1-c)}{\lambda_1-\lambda_2} \\ \lambda_1 & -\lambda_1 & 0 \\ \lambda_2 & 0 & -\lambda_2 \end{pmatrix} \end{aligned} \quad (6.19)$$

$$\begin{aligned} &= \begin{pmatrix} -d & f & g \\ e & -e & 0 \\ h & 0 & -h \end{pmatrix} \quad (6.20) \\ &= Q^{(C-O-C)} \end{aligned}$$

- Da $a, b, c, d > 0$, folgt $\lambda_1 > c > \lambda_2 > 0$
- Und damit $e, f, g, h > 0$
- Ergo: Das Ergebnis ist also eine vernünftige Generatormatrix.
- Der Schritt von Gl.(6.19) nach Gl. (6.20) kann umgedreht werden.

Damit folgt:

Für jedes C-C-O-Modell gibt es ein äquivalentes C-O-C Modell und vice versa.

- Gilt auch für C-C-O, C-O-C und Kreis-Modell.
- Da Kreis-Modell 6 Parameter hat, C-C-O, C-O-C aber 4 Parameter, können nicht alle Parameter des Kreismodells unabhängig aus Daten bestimmbar sein.

Die Modelle sind unterscheidbar, wenn Ratenkonstanten von Bedingungen abhängen und Versuche unter verschiedenen Bedingungen vorhanden.

Wenn Modelle nicht unterscheidbar sind, muß gelten:

$$\begin{aligned}\text{rank } Q'_{CO} &= \text{rank } Q_{CO} \\ \text{rank } Q'_{OC} &= \text{rank } Q_{OC}\end{aligned}$$

Ähnlichkeitstrafos erhalten den Rank, Beweis durch Hinschauen.

Demzufolge sind die Modelle

OCCC (rank $Q_{CO} = 1$) und OCCO (rank $Q_{CO} = 2$) unterscheidbar.

6.6.3.2 Parameteridentifizierbarkeit

Frage: Gegeben ein Modell, können alle Parameter bestimmt werden ?

Kontinuierliche Nichtidentifizierbarkeit bedeutet: Zu viele Parameter

Maximale Anzahl identifizierbarer Parameter in aggregierten Markov Modellen [80, 81]

Sei

- π_O (π_C) die stationäre Dichte der Offen (Geschlossen)zustände
- u_O (u_C) ein Vektor mit allen Komponenten = 1

Dann ist die Dichte $f_O(t)$ der Offenzeitverteilung:

$$f_O(t) = \pi_O e^{Q_{OO}t} Q_{OC} u_C$$

Diese läßt sich schreiben als:

$$f_O(t) = \sum_{i=1}^{N_O} \alpha_i e^{\lambda_i t}$$

Analog:

$$f_C(t) = \pi_C e^{Q_{CC}t} Q_{CO} u_O = \sum_{j=1}^{N_C} \beta_j e^{\omega_j t}$$

Ebenso erhält man für die zweidimensionale Verteilung $f_{OC}(t, s)$ für eine Offenzeit t gefolgt von einer Geschlossenzeit s

$$f_{OC}(t, s) = \pi_O e^{Q_{OO}t} Q_{OC} e^{Q_{CC}s} Q_{CO} u_O$$

und

$$f_{OC}(t, s) = \sum_{i=1}^{N_O} \sum_{j=1}^{N_C} \gamma_{ij} e^{\lambda_i t + \omega_j s}$$

mit den selben Parametern λ_i und ω_j wie bei den eindimensionalen Verteilungen.

Analog:

$$f_{CO}(t, s) = \sum_{i=1}^{N_O} \sum_{j=1}^{N_C} \delta_{ji} e^{\omega_j s + \lambda_i t}$$

M.K.z.:

Alle höher-dimensionalen Aufenthaltsverteilungen bestimmen sich aus den zwei-dimensionalen.

Alle Information über den Prozeß steckt in den 1 & 2D Verteilungen. Freie Anzahl deren Parameter legt die maximale Anzahl bestimmbarer Parameter in Q fest, in Umdrehung der obigen Gleichungen.

Maximale Anzahl identifizierbarer Parameter

- Es gibt $N_O + N_C$ Parameter λ_i, ω_j .

- $\int f_{OC}(t, s) ds dt = 1$, folgt: $N_O N_C - 1$ freie Parameter in γ_{ij}
- Die marginalen Dichten müssen übereinstimmen:

$$\int f_{OC}(t, s) ds = \int f_{CO}(s, t) ds$$

und

$$\int f_{OC}(t, s) dt = \int f_{CO}(s, t) dt$$

Dieses führt auf:

$$\sum_{j=1}^{N_C} \frac{\delta_{ji}}{\omega_j} = \sum_{j=1}^{N_C} \frac{\gamma_{ij}}{\omega_j}, \quad i = 1, \dots, N_O$$

und

$$\sum_{i=1}^{N_O} \frac{\delta_{ji}}{\lambda_i} = \sum_{i=1}^{N_O} \frac{\gamma_{ij}}{\lambda_i}, \quad j = 1, \dots, N_C$$

- Dieses gewährleistet Normierungsbedingung an $f_{CO}(s, t)$ und damit an δ_{ji} .
- Die Gleichungen geben für jede Spalte und Zeile von δ_{ji} ein Constraint, bleiben $(N_O - 1)(N_C - 1)$ Freiheitsgrade.
- Damit Gesamtanzahl der Freiheitsgrade, i.e. der (maximal) bestimm-
baren Parameter:

$$N_O + N_C + (N_O N_C - 1) + (N_O - 1)(N_C - 1) = 2N_O N_C$$

Beispiel C-C-O Kette vs. Kreis-Modell

Gateway-Zustand

Zerfällt ein Gating Schema bei Entfernung eines Zustands in zwei getrennte Graphen, von denen einer nur offene und der andere nur geschlossene Zustände enthält, so heißt dieser Zustand Gateway-Zustand.

Dann

- faktorisieren die Verteilungen $f_{OC}(t, s)$ und $f_{CO}(t, s)$.

- verringert sich die Anzahl identifizierbarer Parameter weiter, z.B.: das Loop Modell:

ZEICHNUNG: Nichtidentifizierbares Loop-Modell

Hinreichende Bedingung für die Nichtidentifizierbarkeit [313]
Erinnere Ähnlichkeitstransformation

$$Q' = S^{-1}QS$$

mit

$$S = \left(\begin{array}{c|c} S_{OO} & 0 \\ \hline 0 & S_{CC} \end{array} \right)$$

und

$$\sum_j S_{ij} = 1 \quad \forall i$$

Betrachte Ähnlichkeitstransformationen nahe der Identität

$$S(\vec{\epsilon} = 0) = 1$$

Parameternichtidentifizierbarkeit bedeutet,

- daß Q' gleiches gating Schema wie Q repräsentiert, gibt m Constraints
- aber S nicht-trivial ist.

Damit:

- S hat $(N_O(N_O - 1)) + (N_C(N_C - 1))$ Freiheitsgrade.
- m werden festgelegt.
- $\implies \text{Dim}(\vec{\epsilon}) = (N_O(N_O - 1)) + (N_C(N_C - 1)) - m$
- $\text{Dim}(\vec{\epsilon}) > 0 \implies$ Modell nicht identifizierbar.

Dieses Nichtidentifizierbarkeitskriterium kann Nichtidentifizierbarkeit auch zeigen, wenn $2N_O N_C$ Grenze nicht erreicht.

Loop Modelle mit gleichen O/C-Verweilzeiten

ZEICHUNG Loop-Modell

- Man kann konstruktiv (wie oben) zeigen:
Obiges Loop-Modell ist äquivalent zu dem Loop-Modell:

ZEICHNUNG: äquivalentes Loop-Modell (alle Zustände ausgetauscht zu oben)

- Betrachte obiges mit gleichen mittleren Offen- oder Geschlossenzeiten.
- Man kann zeigen:
Es gibt eine kontinuierliche zweiparametrische Transformation, die
 - die Likelihood invariant läßt
 - die das Gating-Schema invariant läßt
- Folge:
Sind Offen- oder Geschlossenzeiten identisch, sind die Parameter des Modell nicht identifizierbar.
- Fast gleiche Verweilzeiten:
 - Hesse Matrix schlecht konditioniert
FOLIE Condition number
 - Oder: Parameterschätzfehler groß.
FOLIE relative errors I
 - Sehr langsame Konvergenz der Schätzfehler.
FOLIE relative errors II

Ist das System im thermodynamischem Gleichgewicht, gilt für Loops:

$$\text{Produkt der Raten rechtsrum} = \text{Produkt der Raten linksrum}$$

Dieses legt eine Rate fest.

Detailed balance verletzt: Es gibt eine externe Energiequelle.

- Obige Eichtransformation hat zwei Freiheitsgrade
- Detailed balance legt einen fest
- Es gibt immer noch Nichtidentifizierbarkeit
- Speziell: Es gibt kontinuierlich Transformation zu Modellen, die detailed balance nicht erfüllen [313].

Merke:

Sind Offen- oder Geschlossenzeiten gleich, kann die Frage nach detailed balance nicht beantwortet werden.

Zusammenfassung:

Aggregierte Markov-Prozesse:

- Es gibt ununterscheidbare Topologien [157, 138, 315].
- Es gibt Obergrenze für Anzahl bestimmbarer Parameter [80, 81].
- Detailed balance ist bei identischen Aufenthaltszeiten nicht, bei ähnlichen schwer zu entscheiden [313].

6.6.4 Likelihood Ratio Tests

Beste Theorie-Literatur: [51]

Nomenklatur:

- Gegeben eine Modellklasse M mit Parametervektor $\theta \in R^r$.
- Wahrer Parameter: θ_0
- Geschätzter Parameter, basierend auf N Daten: $\hat{\theta}_N$

Erster LRT:

- $H_0 : M$ ist wahr

- $H_1 : M$ ist nicht wahr

Annahmen:

1. Die Parameter θ liegen nicht auf dem Rand des Parameterraums.
2. Die MLEs sind asymptotisch normal, i.e.:

$$\sqrt{N}(\hat{\theta}_N - \theta_0) = \mathcal{N}(0, \Sigma)$$

mit

$$\Sigma = -N \left(\frac{\partial^2 L(\hat{\theta}_N)}{\partial \theta_i \partial \theta_j} \right)^{-1}$$

3. Das Modell sei identifizierbar, i.e. θ ist eindeutig aus den Daten zu bestimmen, siehe Kap. 6.6.3.

dann gilt:

$$2 [L_N(\hat{\theta}_N) - L_N(\theta_0)] \sim \chi_r^2 \quad . \quad (6.21)$$

(Differenz der log-likelihoods ist *ratio* der likelihoods)

Beweis:

$$\begin{aligned} L_N(\theta_0) &= L_N(\hat{\theta}_N) + \frac{\partial}{\partial \theta_i} L_N(\hat{\theta}_N)(\theta_0 - \hat{\theta}_N) + \\ &\quad \frac{1}{2}(\theta_0 - \hat{\theta}_N) \frac{\partial^2}{\partial \theta_i \partial \theta_j} L_N(\hat{\theta}_N)(\theta_0 - \hat{\theta}_N) + O(|\theta_0 - \hat{\theta}_N|^3) \quad . \end{aligned}$$

- 2. Term RHS = 0 wegen MLE.
- Vernachlässigung von Termen höherer Ordnung
- Σ^{-1} dreht die Korrelationen der $\hat{\theta}_N$ raus.
- Löse auf nach $2(L(\hat{\theta}_N) - L(\theta_0))$

Da $L(\theta_0)$ nicht bekannt, eher von theoretischem Interesse.
Aber [288]:

- Schätze θ_0 aus allen Daten
- Viele $\hat{\theta}_N$ aus Datenstückchen
- Teste, ob Verteilung $2(L(\hat{\theta}_N) - L(\theta_0))$ stimmt

2. LRT: Gegeben zwei Modelle

- M_1 (ein Untermodell) mit r_1 Freiheitsgraden (θ_1) (von)
- M_2 (generelles Modell) mit r_2 Freiheitsgraden (θ_2)

Einfachster Fall:

- M_1 : 1. Komponente von $\theta_1 = 42$
- M_2 : 1. Komponente von $\theta_2 \subset R$
- $r_1 = r_2 - 1$

Annahmen:

1. Die Modellklassen sind geschachtelt ("nested"), i.e. eine ist Obermodellklasse des anderen
Gebe Beispiel für nicht-genested.
2. Die Modellklassen sind richtig spezifiziert, i.e. in ihnen ist das wahre Modell
3. Die MLEs sind asypmtotisch normal.
4. Die wahren Parameter liegen nicht auf dem Rand des Parameterraums
5. Alle Parameter sind identifizierbar unter der Nullhypothese.

Dann gilt asymptotisch:

$$2 [L_N(\hat{\theta}_2) - L_N(\hat{\theta}_1)] \sim \chi_{r_2-r_1}^2 \quad . \quad (6.22)$$

Beweis:

Teile Parametervektor θ in (ψ, λ)

Sei Untermodell gegeben durch Fixierung von ψ auf ψ_0

Durch Taylorentwicklung um Maximum Likelihood Schätzer, ergibt sich für den Likelihood Ratio

$$LR = 2 (L(\hat{\psi}, \hat{\lambda}) - L(\psi_0, \hat{\lambda}_0))$$

mit $\bar{\theta} = (\bar{\psi}, \bar{\lambda})$ den wahren Werten, folgt:

$$LR = 2L(\bar{\psi}, \bar{\lambda}) - \begin{pmatrix} \bar{\psi} - \hat{\psi} \\ \bar{\lambda} - \hat{\lambda} \end{pmatrix} I(\hat{\psi}, \hat{\lambda}) \begin{pmatrix} \bar{\psi} - \hat{\psi} \\ \bar{\lambda} - \hat{\lambda} \end{pmatrix} - (2L(\psi_0, \bar{\lambda}_0) - (\bar{\lambda}_0 - \hat{\lambda}_0) I(\hat{\lambda}_0) (\bar{\lambda}_0 - \hat{\lambda}_0))$$

Ist die Nullhypothese wahr gilt $\bar{\psi} = \psi_0$, und bewiesen.

Merke:

Asymptotik der LRTs aus asymptotischer Gaußianität der Schätzer.

Merke besonders:

LRT ...

- ... beurteilt höhere Erklärungsmöglichkeit komplexerer Modelle in Anbetracht der höheren Anzahl der Parameter (Occam's Razor).
- ... legt eine Skala der Beurteilung fest.

Bemerkung:

Für den Regressionsfall:

Asymptotisch: LRT = F-Test

Verwandte Tests [40]

Graphische Veranschaulichung des LRT

FOLIE [40]

Bestimmt Differenz der log-likelihoods durch Gucken an 2 Punkten.

6.6.4.1 Wald-Test

Idee:

- $(\hat{\theta}_2 - \theta_2)^2$ soll "klein" sein.
- Skala durch Krümmung der log Likelihood

Am obigen einfachsten Beispiel :

$$W = (\hat{\theta}_2 - \theta_2)^2 \frac{\partial^2}{\partial \theta_i \partial \theta_j} L_N(\hat{\theta}_2)$$

Im vektoriellen Fall:

$$W = (\hat{\theta}_2 - \theta_2)^T \frac{\partial^2}{\partial \theta_i \partial \theta_j} L_N(\hat{\theta}_2) (\hat{\theta}_2 - \theta_2)$$

Asymptotisch: Unter H_0 : $W \sim \chi_{r_2 - r_1}^2$

FOLIE [40]

Bestimmt Differenz der log-likelihoods durch Extrapolation basierend auf Obermodell (geschätzter Parameterwert und Krümmung der log-likelihood).

Entspricht Test auf Parameter mit Null verträglich. Eigentlicher LRT ist besser.

6.6.4.2 Lagrange Multiplier-Test

Idee:

- $\frac{\partial}{\partial \theta} L_N(\theta_1)$ soll "klein" sein.
- Skala durch Krümmung der log Likelihood

Sei

$$S(\theta) = \frac{\partial}{\partial \theta} L_N(\theta), \quad (S \text{ wie Score})$$

dann

$$LM = S(\theta_2)^T \frac{\partial^2}{\partial \theta_i \partial \theta_j} L_N(\theta_2) S(\theta_2)$$

FOLIE [40]

Asymptotisch: Unter H_0 : $LM \sim \chi_{r_2-r_1}^2$

Bestimmt Differenz der log-likelihood durch Extrapolation basierend auf Untermodell (Steigung und Krümmung der log-likelihood).

6.6.4.3 Nicht-Standard Testsituationen

"Nicht-Standard" meint: Annahmen für obige Resultate nicht erfüllt.

Häufigste Fälle

- Unter H_0 liegt Parameter auf dem Rand des Parameterraums [263, 312]
Folge: Schätzer kann nicht Gauß-verteilt sein.
- Parameter ist unter der Null nicht identifizierbar [106, 9]. Verteilung unklar.

Betrachte

ZEICHNUNG O-C-C vs. O-O-C $\begin{matrix} b \\ \xrightarrow{\quad} \\ \xleftarrow{\quad} \\ a \end{matrix}$ C nach # of States paper [314]

 H_0 : Wahre Model $\in \Gamma$ H_1 : Wahre Model $\notin \Gamma$

oder

 H_0 : $b = 0$ H_1 : $b \neq 0$

Folge:

1. b ist unter H_0 auf dem Rand des Parameterraums.
 \implies MLE kann nicht Gaußverteilt sein.
2. a ist nicht identifizierbar unter H_0 .

Ad [1.]:

- Ist der Rand simple und bekannt, gibt es analytische Resultate (sehr schön in [263])
- Beispiel: skalar, sonst á la oben: $b \geq 0$, Halbebene.

Statt Gaußverteilung :

- Potentiell negative Werte werden 0
- Potentiell positive Werte bleiben, wie gehabt
- Erklärende ZEICHNUNG

- Test Statistik:

$$2[L_N(\hat{\theta}_2) - L_N(\hat{\theta}_1)] \sim \frac{1}{2}\chi_0^2 + \frac{1}{2}\chi_1^2$$

χ_0^2 erklären, ZEICHNUNG Normal vs. Nichtstandard

- Merke:
Berücksichtigt man "Parameter auf dem Rand" nicht, wird der Standard-Test konservativ.

Ad [2.]: Schwierig

Parametrischer Bootstrap:

- Fitte Untermodell an die Daten
- Realisiere viele Zeitreihen
- Fitte jeweils Unter- und Obermodell
- Bestimme daraus empirische LR-Verteilung
- Basiere Test auf α Quantil der empirischen LR-Verteilung

6.6.4.4 Nicht geschachtelte Modelle

[312]

Sind die Modelle nicht geschachtelt (non-nested) könnte man allgemeines Obermodell als große Schachtel nehmen.

Verbietet sich für aggregierte Markov Modelle, da

- Anzahl der identifizierbaren Parameter begrenzt, siehe Kap. 6.6.3
- Dann ständig Nicht-Standard Test Situationen auftauchen würden

Ausweg für non-nested Models mit gleicher Anzahl offener und geschlossener Zustände siehe [316].

6.6.4.5 Ein Beispiel: Test auf detailed balance

ZEICHUNG Loop-Modell

Ist das System im thermodynamischem Gleichgewicht, gilt für Loops:

$$\text{Produkt der Raten rechtsrum} = \text{Produkt der Raten linksrum}$$

Dieses legt eine Rate fest.

Detailed balance verletzt: Es gibt eine externe Energiequelle.

LRT auf detailed balance

- H_0 : Detailed balance ist erfüllt
- H_1 : Detailed balance ist nicht erfüllt

$$2(L(\hat{\theta}_2) - L(\hat{\theta}_1)) \sim \chi_1^2$$

Aber:

Erinnere Dilemma III des Testens: Statistical significance vs. specific relevance

- Ratenkonstanten q_{ij} hängen mit Aktivierungsenergie E_{ij} über Arrhenius' Gesetz zusammen

$$q_{ij} \propto e^{-\left(\frac{E_{ij}}{kT}\right)}$$

ZEICHUNG zu Aktivierungsbergen

- Typische Skala gegeben durch
 - Membranpotential
 - Elementarladung.

$$E_0 \approx 10^{-20} J$$

- Detailed balance :

$$E_{12} + E_{23} + E_{31} = E_{13} + E_{32} + E_{21}$$

- respektive:

$$0 = \log \left(\frac{q_{12}q_{23}q_{31}}{q_{13}q_{32}q_{21}} \right) =: \log V$$

- Wenn detailed balance verletzt, bestimmt

$$kT \log V = E_0$$

die natürliche Skala.

- Beim Raumtemperatur:

$$\log V \approx 2.6$$

gibt kritischen (fachspezifischen Wert)

Kap. 6.6.3: Detailed balance Modelle mit gleichen Offen-, respektive Geschlossenzeiten sind nicht identifizierbar.

Folge [313]:

- Der Test hat in diesem Fall keine power.
- Bei ähnlichen Zeiten folgt verminderte Power.

6.7 Hidden Markov Modell

Literatur:

- R.J. Elliott, L. Aggoun, J.B. Moore: Hidden Markov Models : Estimation and Control [74]
- L.R. Rabiner: A tutorial on hidden Markov models and selected applications in speech recognition [229]
- J.D. Hamilton: Time Series Analysis [102]
- A.B. Poritz: Hidden Markov models: A guided tour [224]

Ist der Zustand $x(t)$ nicht direkt beobachtbar, sondern bestimmt nur die Wahrscheinlichkeit $p(y(t)|x(t))$ der Beobachtung $y(t)$, so spricht man von einem HMM.

Beispiel:

$$p(y(t)|x(t) = i) = \frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{(y-\mu_i)^2}{\sigma_i^2}}$$

HMM Zeichnung mit Aggregierung

- Formal ist Aggregierung im Markov Modell ein Spezialfall eines HMM.
- Inhaltlich sollten aber die Auswirkungen der Aggregierung von denen des Beobachtungsrauschen auseinandergelassen werden.

6.7.1 EM Algorithmus im HMM, Baum-Welsh-Algorithmus

Reminder: (Nicht aggregiertes) Markov Modell:

Likelihood

$$\mathcal{L}(x_1, \dots, x_N | a_{ij}) = \pi_{x_1} \prod_t a_{x_t x_{t+1}}$$

log Likelihood (i.w.)

$$L(x_1, \dots, x_N | a_{ij}) = \sum_t \log a_{x_t x_{t+1}} = \sum_{i=1}^s \sum_{j=1}^s n_{ij} \log a_{ij}$$

mit

$$n_{ij} = \sum_t 1_{(x_t=i, x_{t+1}=j)}$$

ergibt:

$$\hat{a}_{ij} = \frac{n_{ij}}{\sum_j n_{ij}}$$

Bei HMMs

- Problem:

Der Zustand $x(t)$ ist nicht mehr eindeutig aus der Beobachtung $y(t)$ zu erschließen.

- Damit:

Möchte man Likelihood berechnen, muß man jede Kombination der Zustände über die Zeit ("Pfad") gemäß ihrer Wahrscheinlichkeit berücksichtigen :

ZEICHNUNG des Netzes der Zustände

Bei S Zuständen und N Daten ergibt dies eine Komplexität von $O(S^N)$ Pfaden, was sehr schnell nicht mehr behandelbar wird.

- Lösung:

forward - procedure : Bestimmung der Likelihood mit Aufwand von $O(NS^2)$ [17] .

- Beruht auf Markov Eigenschaft des Hintergrundprozesses
- Erster Teil des E-Schrittes im EM - Algorithmus.

Definiere :

$$\alpha_t(i) = p(y_1, \dots, y_t, x_t = i | \theta)$$

$\alpha_t(i)$ kann rekursiv berechnet werden:

- Initialisierung :

$$\alpha_1(i) = \pi_i p_i(y_1)$$

$\pi_i p_i(y_1)$ erläutern.

Für das hier betrachtete Gaußsche Model ist

$$p_i(y_t) = \frac{1}{\sqrt{2\pi\sigma_i}} e^{-\frac{(y_t - \mu_i)^2}{2\sigma_i^2}}$$

- Rekursion :

$$\alpha_{t+1}(j) = \sum_i \alpha_t(i) a_{ij} p_j(y_{t+1}) \quad (6.23)$$

FOLIE Erklärendes Bild zu $\alpha_t(i)$ a la Rabiner, 1989, fig. 5

- Bemerkung: Am Ende ergibt sich die log-Likelihood durch :

$$L(\theta) = \log p(y_1, \dots, y_N | \theta) = \log \sum_i \alpha_N(i) \quad (6.24)$$

Dies kann zur Parameterschätzung per numerischer Optimierung genutzt werden, siehe Analogie zum Kalmanfilter.

Bemerkung:

Die $\alpha_t(i)$ werden sehr schnell sehr klein. Um numerischen Underflow zu vermeiden, werden sie jeweils auf $\sum_i \alpha_t(i) = 1$ normiert. Man merke sich den Normierungsfaktor, und verrechne ihn am Ende [74].

Analog zu dem *forward - procedure* läßt sich das *backward - procedure* definieren (2. Teil E-Step):

$$\beta_t(i) = p(y_{t+1}, \dots, y_N, x_t = i | \theta)$$

- Initialisierung:

$$\beta_N(i) = 1, \quad (\text{Konvention, wird durch Normierung zu transientem Effekt})$$

- Rekursion

$$\beta_t(i) = \sum_j a_{ij} \beta_{t+1}(j) p_j(y_{t+1})$$

FOLIE Erklärendes Bild zu $\beta_t(i)$ a la Rabiner, 1989, fig. 5

(Gemeinsame) Wahrscheinlichkeit zum Zeitpunkt t in Zustand i und zu $t + 1$ im Zustand j zu sein, gegeben die Beobachtungen ist:

$$\psi_{ij}(t) = \frac{\alpha_t(i) a_{ij} p_j(y_{t+1}) \beta_{t+1}(j)}{\sum_i \sum_j \alpha_t(i) a_{ij} p_j(y_{t+1}) \beta_{t+1}(j)}$$

FOLIE Erklärendes Bild zu $\psi_{ij}(t)$ a la Rabiner, 1989, fig. 6

Wahrscheinlichkeit zum Zeitpunkt t in Zustand i zu sein, gegeben die Beobachtungen, ergibt sich durch Summation über j :

$$\Psi_i(t) = \sum_j \psi_{ij}(t)$$

$\psi_{ij}(t)$ und $\Psi_i(t)$ sind die Analoga zu n_{ij} aus Gl.(6.16) und $\sum_j n_{ij}$ aus Gl.(6.17) dar, nun gewichtet mit den entsprechenden Wahrscheinlichkeiten.

- Im allgemeinen Ansatz des EM - Algorithmus stellen die $\psi_{ij}(t)$ die entscheidende Größe dar.
- In ihnen ist die Information über den versteckten Prozeß, gegeben die Daten und eine Schätzung der Modellparameter aus dem vorangegangenen Iterationsschritt, zusammengefaßt.

Erinnere EM Algorithmus und:

$$\begin{aligned}
 A(\theta, \theta^{(i)}) &= \langle \log p(y_1, \dots, y_N, x_1, \dots, x_N | \theta) \rangle_{p(x_1, \dots, x_N | y_1, \dots, y_N, \theta^{(i)})} \\
 &= \sum_{\text{alle Pfade}} p(x_1, \dots, x_N | y_1, \dots, y_N, \theta^{(i)}) \log p(y_1, \dots, y_N, x_1, \dots, x_N | \theta) \\
 &= \sum_{\text{alle Pfade}} p(x_1, \dots, x_N | y_1, \dots, y_N, \theta^{(i)}) (\log p(x_1, \dots, x_N | \theta) + \log p(y_1, \dots, y_N | x_1, \dots, x_N, \theta))
 \end{aligned}$$

läßt sich nun explizit ausschreiben:

$$A(\theta, \theta^{(i)}) = \sum_i \Psi_i(1) (\log(\pi_i) + \log(p(y_1 | x_1 = i))) + \sum_{ij} \sum_t \psi_{ij}(t) (\log a_{ij} + \log p(y_{t+1} | x_{t+1} = j))$$

Die Ausgaben sind "bedingt unabhängig".

- Wir vernachlässigen den zeitlich ersten Term wiederum und definieren zur Vereinfachung :

$$Q_1(a_{ij}) = \sum_{ij} \sum_t \psi_{ij}(t) \log a_{ij}$$

und

$$Q_2(\mu_i, \sigma_i) = \sum_{ij} \sum_t \psi_{ij}(t) \log p(y_{t+1} | x_{t+1} = j)$$

- Ausführen der Summation über i und Einsetzen der Gaußverteilung in der letzten Gleichung führt zu

$$\begin{aligned}
 Q_2(\mu_i, \sigma_i) &= \sum_j \sum_t \Psi_j(t) \log p(y_t | x_t = j) \\
 &= \sum_j \sum_t \Psi_j(t) \log \left(\frac{1}{\sqrt{2\pi}\sigma_j} e^{-\frac{(y_t - \mu_j)^2}{2\sigma_j^2}} \right)
 \end{aligned}$$

- Damit ist der E - Schritt des EM - Algorithmus abgeschlossen.
- Der M - Schritt läßt sich analytisch vollziehen.

Ableiten nach den Parametern liefert:

$$\hat{\mu}_j = \frac{\sum_t \Psi_j(t) y_t}{\sum_t \Psi_j(t)}$$

$$\hat{\sigma}_j^2 = \frac{\sum_t \Psi_j(t) (y_t - \hat{\mu}_j)^2}{\sum_t \Psi_j(t)}$$

mit der anschaulichen Interpretation, daß die Daten zu jedem Zeitpunkt mit der entsprechenden Wahrscheinlichkeit, daß der Prozeß im jeweiligen Zustand war, in die Schätzung eingehen.

- Die Schätzung für die Übergangswahrscheinlichkeiten muß die Nebenbedingung Gl.(6.15) $\sum_j a_{ij} = 1$ berücksichtigen. Diese geschieht wiederum durch Lagrange Multiplier :

$$\tilde{Q}_1(a_{ij}) = Q_1(a_{ij}) + \sum_i \lambda_i \sum_j (a_{ij} - 1) \quad (6.25)$$

- Nullsetzen der Ableitung nach den Parametern a_{ij} und den Lagrange Multipliern λ_i führt zu :

$$\hat{a}_{ij} = \frac{\sum_t \psi_{ij}(t)}{\sum_t \Psi_i(t)}$$

mit der gleichen anschaulichen Begründung wie zuvor.

Mit den erhaltenen Parametern wird im nächsten E – Schritt wiederum mit Hilfe der *forward - backward procedure* die auf diese Parameter bedingten Wahrscheinlichkeiten berechnet und das Verfahren iteriert, bis sich keine wesentlichen Änderungen in den Parametern mehr zeigen.

Die Nebenbedingungen Gl. (6.14,6.15) werden automatisch erfüllt.

Schätzung der Ratenkonstanten durch Parametrisierung der a_{ij}

$$a_{ij} = a_{ij}(q_{ij})$$

und numerisches Lösen von Gl.(6.25), siehe [196].

Für Anwendung siehe [19]

Meditation, warum man das Ganze überhaupt braucht, auch ZRM.

- Auswirkung auf geschätzte Raten beim HMM:
 - Zustand nicht mehr festzustellen:
 - $a_{ij} = 1/s$
 - White noise Prozeß
- Parameter beim ZRM:
 - Error-in-Variables-Problem
 - $x(t) = ax(t-1) + \epsilon(t)$ geht mit $\hat{a} \rightarrow 0$ gegen WN-Prozeß $x(t) = \epsilon(t)$

6.7.2 Viterbi Algorithmus

Eine sinnvolle Frage:

In welchem Zustand x_t war ich zur Zeit t am wahrscheinlichsten ?

$$\gamma_t(i) = P(x_t = i | y_1, \dots, y_N, \theta) = \frac{\alpha_t(i)\beta_t(i)}{\sum_{i=1}^s \alpha_t(i)\beta_t(i)}$$

$$x_t = \arg \max_i (\gamma_t(i))$$

Andere Frage:

Welches ist der wahrscheinlichste Pfad $\{x_1^*, x_2^*, \dots, x_N^*\}$?

Lösung:

Viterbi-Algorithmus, ein Spezialfall der Dynamischen Programmierung [2]:

- Lösung rückführbar auf Teillösungen
- (Vorwärts-)Berechnung der Teillösungen
- (Rückwärts-)Bildung der Lösung aus den Teillösungen

Definiere:

$$\delta_t(i) = \max_{x_1, x_2, \dots, x_{t-1}} P(x_1, x_2, \dots, x_{t-1}, x_t = i | y_1, \dots, y_t, \theta)$$

Offensichtlich:

$$\delta_{t+1}(j) = \left[\max_i (\delta_t(i) a_{ij}) \right] p_j(y_{t+1}) \quad (6.26)$$

Vergleiche mit der $\alpha_{t+1}(j)$ -Rekursion, Gl. (6.23) im *forward*-Algorithmus:

$$\alpha_{t+1}(j) = \sum_i \alpha_t(i) a_{ij} p_j(y_{t+1})$$

Definiere:

$$\psi_{t+1}(j) = \arg \max_i (\delta_t(i) a_{ij})$$

um den Gewinner aus Gl. (6.26) zu tracen.

Damit :

- Initialisierung:

$$\begin{aligned} \delta_1(i) &= \pi_i p_i(y_1), \quad i = 1, \dots, s \\ \psi_1(i) &= 0 \end{aligned}$$

- Rekursion:

$$\begin{aligned} \delta_{t+1}(j) &= \left[\max_i (\delta_t(i) a_{ij}) \right] p_j(y_{t+1}) \\ \psi_{t+1}(j) &= \arg \max_i (\delta_t(i) a_{ij}) \end{aligned}$$

- Abschluß:

$$x_N^* = \arg \max_i (\delta_N(i))$$

- Pfad backtracking:

$$x_{t-1}^* = \psi_t(x_t^*)$$

ergibt wahrscheinlichsten Pfad

$$\{x_1^*, x_2^*, \dots, x_N^*\}$$

- Komplexität: $O(Ns^2)$ statt naiv: $O(s^N)$.

Die Größe

$$-\log p(x_{1,\dots,N})$$

wird selbsterklärend auch als Pfadlänge bezeichnet.

Für Anwendung siehe [82]

Beachte:

- Parallelität und Unterschied zum Kalman-Filter.
- Beide Verfahren produzieren keine typischen Pfade

6.7.3 HMM Erweiterungen

Bisherige HMM Annahmen:

- Rauschen ist weiß
- Daten werden nicht gefiltert

Bei Ionenkanälen:

- Kapazitive Effekte der Zellmembran
- Frequenzgang des Verstärkers
- Anti-aliasing Filter (hier besonders wichtig)

ZEICHNUNG blaues Spektrum der Daten, Anti-aliasing Filter

Mathematisch:

Nicht weißes Rauschen und Filtern der Daten zerstören bedingte Unabhängigkeit $p(y_t|x_t)$.

”Einfachster Fall”:

Rauschen und Filter seien linear, i.e. ARMA-Prozesse.

Ausgabe des ursprünglichen HMMs: $\mu(x_t)$, $\sigma(x_t)$.
Das ergibt (z.B.) den Prozeß Z_t :

$$Z_t = \sum_{i=0}^b \chi_i \mu(x_{t-i}) + \eta_t$$

$$\eta_t = \sum_{i=1}^p a_i \eta_{t-i} + \sum_{i=0}^q b_i \sigma(x_{t-i}) \epsilon_{t-i}$$

Behandelt alle obigen Fälle

Anderes Setting: Regime switching [102]:

Zeitdiskretes 2 Zustands-AR[1] Modell mit Übergangsmatrix:

$$\begin{pmatrix} p_{11} & p_{21} \\ p_{21} & p_{22} \end{pmatrix}$$

und Ausgabe:

$$(y_t - \mu(x_t)) = a(y_{t-1} - \mu(x_{t-1})) + \epsilon_t$$

$$\mu(x_t = 1) = \mu_1$$

$$\mu(x_t = 2) = \mu_2$$

Geht analog auch mit $a = a(x_t)$

Parameterschätzung

Betrachte 2. Beispiel:

Grundidee: Aufblähen des Zustandsraums.

Definiere Zustände:

$$s_t = 1 \quad \text{if} \quad x_t = 1 \text{ und } x_{t-1} = 1$$

$$s_t = 2 \quad \text{if} \quad x_t = 2 \text{ und } x_{t-1} = 1$$

$$s_t = 3 \quad \text{if} \quad x_t = 1 \text{ und } x_{t-1} = 2$$

$$s_t = 4 \quad \text{if} \quad x_t = 2 \text{ und } x_{t-1} = 2$$

mit:

$$p(y_t | y_{t-1}, s_t = 1, a) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left\{ -\frac{[(y_t - \mu_1) - a(y_{t-1} - \mu_1)]^2}{2\sigma^2} \right\}$$

$$p(y_t | y_{t-1}, s_t = 2, a) = \frac{1}{\sqrt{2\pi\sigma}} \exp \left\{ -\frac{[(y_t - \mu_2) - a(y_{t-1} - \mu_1)]^2}{2\sigma^2} \right\}$$

usw...

Dieses 4 Zustandsmodell: "normales" HMM mit parametrisierter Übergangsmatrix:

$$\begin{pmatrix} p_{11} & p_{12} & 0 & 0 \\ 0 & 0 & p_{21} & p_{22} \\ p_{11} & p_{12} & 0 & 0 \\ 0 & 0 & p_{21} & p_{22} \end{pmatrix}$$

MA-Prozesse schwieriger.

Für approximativen MLE, siehe [197]

Besprechung der Anwendung und Modellselektion: [195]

- Vergleich HMM vs. MA gefiltertes HMM zur Behandlung des Anti-Aliasing Filter bei Ionenkanaldaten
- FOLIE Ströme: Fig. 6.4
- Modellselektion durch LRTs:

FOLIE LRTs: Tabelle 6.1

- Ergibt
 - ungefiltertes HMM: 2 O-Zustände
 - gefiltertes HMM: 1 O-Zustände
- Probleme:
 - Multiples Testen
 - Mögliche Modellmisspezifikation
- Darum : Unabhängige Checks

1. Vergleich: Theoretischer mittleren Zeitverlauf der beiden Modelle mit empirischem:

FOLIE Summenströme

2. Simulation quasi-kontinuierlicher Daten aus gefiltertem HMM
 - * Anti-Aliasing Filterung, Sampling
 - * Reproduktion der Modellsektion und Parameterschätzung für beide HMMs
- Interpretation:
Ungefiltertes HMM "braucht" artifiziellen 2. Offenzustand, um die Rauschkorrelation zu erklären.

Merke:

Modellsektionsergebnisse statistischer Tests sollten durch unabhängige Evidenz erhärtet werden.

Zum Merken:

Liste der Nichtidentifizierbarkeiten:

- Aggregierte Zustände
- Äquivalente Modelle
- Eventuell Parameter
- Detailed balance Modelle, wenn Aufenthaltszeiten gleich.
- (Zeitkontinuierliche) Topologie, wenn zeitdiskrete Modelle verwendet werden.
- Zustände im HMM, wenn SNR niedrig

6.8 Stochastische Differentialgleichungen

Integration SDEs revisited:

Ito und Stratonovich, Literatur: siehe [85] Kap.4

Wir hatten oben für additives Rauschen in der Langevin Gleichung:

$$\dot{x} = f(x) + \sigma\epsilon, \quad \epsilon \sim N(0, 1) \quad (6.27)$$

gesehen:

$$x(t + \delta t) = x(t) + \delta t f(x(t)) + \int_t^{t+\delta t} dt' \sigma \epsilon(t')$$

und

$$\int_t^{t+\delta t} dt' \sigma \epsilon(t') = \sigma \sqrt{\delta t} \epsilon(t)$$

Formal ist $\sqrt{\delta t} \epsilon(t)$ das Inkrement eines "Wiener Prozesses" $W(t)$, i.e. Brownian Motion:

$$\sqrt{\delta t} \epsilon(t) = W(t + \delta t) - W(t)$$

Formal korrekt ist SDE in Gl.(6.27):

$$dx = f(x)dt + \sigma dW$$

Bei multiplikativen Rauschen wird's komplizierter

$$\dot{x} = f(x) + g(x)\epsilon, \quad \epsilon \sim N(0, 1)$$

$$x(t + \delta t) = x(t) + \delta t f(x(t)) + \int_t^{t+\delta t} dt' g(x(t')) \epsilon(t')$$

Freiheit:

Wie/Wo $g(x(t'))$ in $[t, t + \delta t]$ auswerten ?

- Ito Definition:

Werte an $t' = t$ aus.

$$x(t + \delta t) = x(t) + \delta t f(x(t)) + g(x(t))[W(t + \delta t) - W(t)]$$

Wichtig $[W(t + \delta t) - W(t)]$ "unabhängig" von $g(x(t))$: Stichwort "Martingal"

- Stratonovich Definition:

$$\int_t^{t+\delta t} dt' g(x(t')) \epsilon(t') := \frac{g(x(t+\delta t)) + g(x(t))}{2} [W(t+\delta t) - W(t)]$$

(in manchen Büchern fälschlicher Weise: Werte an $t' = t + 1/2\delta t$) aus.)
 Anyway, wichtig: $[W(t+\delta t) - W(t)]$ und $g(x(t+\delta t)) + g(x(t))$ "interferieren".

Umrechnungen

- Gegeben Ito SDE:

$$dx = a(x)dt + b(x)dW$$

mit

- $a(x)$: Drift-Term
- $b(x)$: Diffusions-Term

ergibt sich (mit Taylor-Entwicklung in richtiger Ordnung) Stratonovich SDE:

$$dx = \left(a(x) - \frac{1}{2} b(x) \partial_x b(x) \right) dt + b(x) dW$$

ein zusätzlicher Drift-Term.

- Entsprechend:

Stratonovich SDE:

$$dx = a(x)dt + b(x)dW$$

ergibt Ito SDE:

$$dx = \left(a(x) + \frac{1}{2} b(x) \partial_x b(x) \right) dt + b(x) dW$$

Stratonovich:

- Physikalisch natürlich für kontinuierliche Prozesse (mit nicht-weißem Rauschen)
- Natürlich unter Variablen Transformationen
- Dem Martingal-Kalkül nicht zugänglich
- Für mathematische Beweise schwer

Ito:

- Physikalisch natürlich für zeitdiskrete Prozesse
- Unnatürlich unter Variablen Transformationen (*Ito's formula*)
- Dem Martingal-Kalkül zugänglich
- Für mathematische Beweise geeignet

Langevin Gleichung beschreibt die Zeitentwicklung der Observablen.

Fokker-Planck-Gleichung die Zeitentwicklung die Dichte $p(x, t)$ des Prozesses.

$$\frac{\partial}{\partial t} p(x, t) = \left[-\frac{\partial}{\partial x} a(x) + \frac{1}{2} \frac{\partial^2}{\partial x^2} b^2(x) \right] p(x, t)$$

... eine PDE.

Integrations-Schrittweiten Wahl:

- Fokker-Planck-Gleichung:
Siehe PDE Integrations-Theorie, e.g. [7, 274].
- Langevin-Gleichung:
Wähle Integrationssschritt δt so klein, daß sich bedingte Dichte

$$p(x(t + \Delta t) | x(t))$$

(auf Sampling-Niveau Δt) nicht mehr ändert. Es folgt : $\delta t \ll \Delta t$

Beispiel: Stochastischer van der Pol:

2 FOLIEN aus STOCHDGL [286]

6.8.1 Diskret vs. kontinuierlich Teil VIII

Gaußverteilung des treibenden Rauschens in SDE geht nach Nicht-Gaußverteilung das treibenden Rauschens in Differenzgleichung: Nichtlineare Differenzgleichungen mit Gaußschen Rauschen sind nicht natürlich. Effekt kann schwach sein

6.8.2 Maximum likelihood Schätzung von Parametern in SDEs

- Gegeben:

N Samples mit Samplingintervall Δt ($T = N\Delta t$) der SDE:

$$\dot{x} = f(x, \theta) + g(x, \theta)\epsilon \quad (6.28)$$

- Aufgabe:

Schätze θ .

Zwei $N \rightarrow \infty$ Asymptotiken:

- Δt fixed, $T \rightarrow \infty$
- $\Delta t \rightarrow 0$, T fixed

Erste ist praktisch interessanter.

Beachte:

- Auf Integrationsintervall gilt

$$x(t + \delta t) = x(t) + \delta t f(x(t), \theta) + g(x(t), \theta)\sqrt{\delta t}\epsilon$$

d.h.: θ läßt sich trivial durch least squares bestimmen.

- Auf Sampling-Intervall gilt immer noch:

$$x(t + \Delta t) = h(x(t), \theta) + \nu(x, \theta, \epsilon)$$

aber von $h(.,.)$ führt kein Weg zu θ zurück.

Likelihood:

$$\mathcal{L}_N(x_1, x_2, \dots, x_N | \theta) = \pi(x_1) \prod_{i=1}^{N-1} p(x_{i+1} | x_i, \theta)$$

resp.

$$L_N(x_2, \dots, x_N | \theta) = \sum_{i=1}^{N-1} \log p(x_{i+1} | x_i, \theta)$$

Wie immer: θ_{MLE} durch:

$$\frac{d}{d\theta} L_N(x_2, \dots, x_N | \theta) = 0 \quad (6.29)$$

Erinnere:

$\frac{d}{d\theta} L_N(\cdot)$ heißt Score function.

Eine Möglichkeit: Optimierte iterativ:

- Initial guess für θ
- Iteriere:
 - Bestimmung von $p(x_{i+1} | x_i, \theta)$ durch:
 - * häufiges Simulieren von Trajektorien & Kernschätzung von $p(x_{i+1} | x_i, \theta)$
 - * lösen der Fokker-Planck-Gleichung
 Beides spaßfrei
 - Update-Schritt für θ
- Abbruch, wenn sich nix mehr tut

Praktisch nicht machbar.

6.8.3 Quasi Maximum Likelihood Schätzer

Martingale für Physiker:

Ein Prozeß ist ein Martingal, wenn i.w. gilt:

$$E(x_i | x_{i-1}) = x_{i-1}$$

- Alles was ich jetzt über gleich deterministisch wissen kann, ist jetzt.
- "Unabhängige Zuwächse"
- Verallgemeinerung von WN für Prozesse
- Aktien sind i.w. Martingale
- Grundlage vieler mathematischer Beweise

Für AR (Markov-) Prozesse gilt:

$$\epsilon_i = x_i - E(x_i | x_{i-1})$$

ϵ_i ist ein Martingal Differenz Prozeß

Kompensatoren für Physiker:

Ist ein Prozeß kein Martingale, so heißt der Prozeß, den man von ihm abziehen muß, damit er einer wird, der Kompensator²

QMLE

- Betrachte formal

$$G_N(\theta) = \frac{d}{d\theta} L_N(x_2, \dots, x_N | \theta) \quad (6.30)$$

in Abhängigkeit von N als stochastischen Prozeß

- Der Schätzer folgt aus

$$G_N(\hat{\theta}) = 0 \quad (6.31)$$

- Nun muß $G_N(\theta)$ nicht der Score $\frac{d}{d\theta} L_N(x_2, \dots, x_N | \theta)$ sein, sondern kann eine andere Funktion der Daten sein.

- Aber:

Für das wahre θ sei $G_N(\theta)$ ein zero-mean Martingal:

$$E(G_{N+1} | G_N) = 0$$

²den wir schon aus Filmen kennen wie "Der Rabiator", "Der Kontrollator", ...

- Gl.(6.31) definiert die "Estimating Function" [91, 117, 90] und einen QMLE.
- Eigenschaften QMLE
 - Unverzerrt
 - Konfidenzintervalle \geq als bei MLE, also konservativ

Quasi Maximum Likelihood Schätzer für SDEs

$$dX_t = b(X_t, \theta)dt + \sigma(X_t, \theta)dW_t$$

F.s.o.c: Annahme: Additives Rauschen: $\sigma(X_t, \theta) = \sigma$.

Ansatz:

Das Inkrement $X_i - X_{i-1}$ sei normalverteilt mit Mittelwert $b(X_i, \theta)\Delta t$ und Varianz $\sigma^2\Delta t$

$$X_i - X_{i-1} \sim N((X_i - X_{i-1}) - b(X_i, \theta)\Delta t, \sigma^2\Delta t)$$

Dann ist Beitrag eines Punktes zur log-likelihood:

$$-\frac{((X_i - X_{i-1}) - b(X_i, \theta)\Delta t)^2}{2\sigma^2\Delta t}$$

Ergibt abgeleitet den Beitrag zum Score:

$$\frac{[(X_i - X_{i-1}) - b(X_i, \theta)\Delta t] b'(X_i, \theta)\Delta t}{\sigma^2\Delta t}$$

oder

$$\frac{b'(X_i, \theta)}{\sigma^2}(X_i - X_{i-1}) - \frac{b(X_i, \theta)b'(X_i, \theta)\Delta t}{\sigma^2}$$

Damit ist (wäre) der Score:

$$L'_N(\theta) = \sum_i \frac{b'(X_i, \theta)}{\sigma^2}(X_i - X_{i-1}) - \Delta t \sum_i \frac{b(X_i, \theta)b'(X_i, \theta)}{\sigma^2}$$

- Dieses klappt sicher, wenn Δt eine zulässige Integrationszeit δt ist.
- Ist Δt nicht klein, führt dies i.a. zu verzerrten Schätzern.
- Andersrum: $L'_N(\theta)$ ist kein zero-mean Martingal.
- Idee: Subtrahiere Kompensator, ergibt QMLE.

Der Kompensator ist (längere Rechnung):

$$\text{Komp}_N(\theta) = \sum_i \frac{b'(X_i, \theta)}{\sigma^2} (E(X_i | X_{i-1}, \theta) - X_{i-1}) - \Delta t \sum_i \frac{b(X_i, \theta)b'(X_i, \theta)}{\sigma^2}$$

Ergibt:

$$G_N(\theta) = \sum_i \frac{b'(X_i, \theta)}{\sigma^2} (X_i - E(X_i | X_{i-1}, \theta))$$

- Im Gegensatz zur bedingten Dichte $p(x(t + \Delta t) | x(t))$ für den MLE ist $E(X_i | X_{i-1}, \theta)$ einfach gut auszurechnen.

Für Beispiele siehe: [24, 286].

6.8.4 Diskret vs. Kontinuierlich Part IX

Nimmt man an, das Sampling-Intervall Δt sei ein zulässiges Integrationsintervall δt [31, 32, 268, 93], kann das zu arg verzerrten Schätzern führen [58, 77, 265, 286].

FOLIE aus STOCHDGL [286]

Oder bei Fokker-Planck frei nach [167]:

$$\begin{aligned} D_1 &= \lim_{\Delta t} \dots \\ D_2 &= \lim_{\Delta t} \dots \end{aligned}$$

6.9 Deterministische Systeme

Für deterministische Prozesse ist vieles besser als für die bisherigen stochastischen, da man eine Trajektorie hat. Das Problem des Beobachtungsrauschens bleibt bestehen.

Einzubauen, grade wegen nicht Effizienz des MLE: [221]

6.9.1 Parametrische Methoden

Gegeben

$$\begin{aligned}\dot{\vec{x}} &= \vec{f}(\vec{p}, \vec{x}) \\ y(t_i) &= g(\vec{x}(t_i)) + \epsilon(t_i)\end{aligned}$$

schätze \vec{p} .

3 Ansätze

6.9.1.1 Regressionsansatz

1. Ignoriere das Rauschen
2. Transformiere:

$$\begin{aligned}\dot{\vec{x}} &= \vec{f}(\vec{p}, \vec{x}) \\ y(t_i) &= g(\vec{x}(t_i)) + \epsilon(t_i)\end{aligned}$$

in

$$y^{(n)} = h(y^{(n-1)}, y^{(n-2)}, \dots, y, \vec{p}) \quad (*)$$

Daß das immer geht, sichert das Takens'sche Theorem [277].

3. Schätze Zeitableitungen $y^{(j)}$ aus den Daten, siehe unten.
4. Plug in (*).
5. Schätze \vec{p} durch nichtlineare Regression:

$$\Gamma(\vec{p}) = \sum_i (\hat{y}^{(n)}(t_i) - h(\hat{y}^{(n-1)}(t_i), \hat{y}^{(n-2)}(t_i), \dots, y(t_i), \vec{p}))^2$$

Probleme:

- Benötigt hohes Signal-zu-Rausch Verhältnis
- Beschränkt auf 2-3 dimensionale Systeme.
- $\Gamma(\vec{p})$ ist keine Likelihood. Darum: Schwer, Konfidenz Intervalle für Parameter zu bekommen

Für Anwendungen, siehe [55, 54, 112]

Ableitungen aus Daten schätzen

- 1.Ordnung:

$$\begin{aligned}x(t + \Delta t) &= x(t) + \dot{x}\Delta t + O(\Delta t^2) \\ \dot{x} &= \frac{x(t + \Delta t) - x(t)}{\Delta t} + O(\Delta t)\end{aligned}$$

Da $x(t + \Delta t) \approx x(t)$ wird bei Rauschen auf den Daten dessen relativer Anteil verstärkt.

- 2.Ordnung:

$$\begin{aligned}x(t + \Delta t) &= x(t) + \dot{x}\Delta t + \frac{1}{2}\ddot{x}\Delta t^2 + O(\Delta t^3) \\ x(t - \Delta t) &= x(t) - \dot{x}\Delta t + \frac{1}{2}\ddot{x}\Delta t^2 + O(\Delta t^3) \\ \dot{x} &= \frac{x(t + \Delta t) - x(t - \Delta t)}{2\Delta t} + O(\Delta t^2)\end{aligned}$$

- "Alle" Ordnungen:

$$\begin{aligned}\text{Fourier-Darstellung: } f(\omega_k) &= \sum_t e^{-i\omega_k t} x(t) \\ x(t) &= \sum_{\omega_k} e^{i\omega_k t} f(\omega_k) \quad \text{ableiten} \\ \text{ergibt: } \dot{x}(t) &= \sum_{\omega_k} i\omega_k e^{i\omega_k t} f(\omega_k)\end{aligned}$$

Zeigt das Problem:

- Hohe Frequenzen (=Rauschen) werden verstärkt.
- Daher: Rücktrafo nur bis $\omega_{cut-off}$.
- $\omega_{cut-off} >$ als typische Frequenzen des Prozesses.

FOLIE PLA plot

- Splines, müssen nicht schlecht sein.

Zweite Ableitung in 1. Ordnung:

$$\begin{aligned}
 x(t + \Delta t) &= x(t) + \dot{x}\Delta t + \frac{1}{2}\ddot{x}\Delta t^2 + O(\Delta t^3) \\
 x(t - \Delta t) &= x(t) - \dot{x}\Delta t + \frac{1}{2}\ddot{x}\Delta t^2 + O(\Delta t^3) \\
 &\text{addieren und auflösen} \\
 \ddot{x} &= \frac{x(t + \Delta t) - 2x(t) + x(t - \Delta t)}{\Delta t^2} + O(\Delta t)
 \end{aligned}$$

Verallgemeinerungen analog zu oben.

FOLIE van der Pol Ergebnisse

6.9.1.2 Anfangswertansatz

$$\begin{aligned}
 \dot{\vec{x}} &= \vec{f}(\vec{p}, \vec{x}) \quad (\ddagger) \\
 y(t_i) &= g(\vec{x}(t_i)) + \epsilon(t_i)
 \end{aligned}$$

1. Wähle Startschätzung \vec{p}_0 und Anfangswert $\vec{x}_0(t_0)$.
2. Integriere (\ddagger) , gibt $\hat{\vec{x}}(t_i)$.
3. Berechne $\hat{y}(t_i)$.
3. Schätze \vec{p} und $\vec{x}(t_0)$ durch Minimierung von :

$$\chi^2(\vec{p}, \vec{x}(t_0)) = \sum_i \left(\frac{y(t_i) - \hat{y}(t_i, \vec{p}, \hat{\vec{x}}(t_0))}{\sigma_i} \right)^2$$

Vorteil:

- Keine Zeitableitungsschätzung nötig
- Maximum likelihood Schätzer
"Übliches" χ^2 , aber mit Nebenbedingung "Trajektorie" nach (‡)
- Konfidenz Intervalle aus:

$$\frac{\partial^2 \chi^2(\hat{\vec{p}}, \hat{\vec{x}}(t_0))}{\partial p_i^2}$$

Problem:

- Konvergenz in lokale Minima

Beispiel:

Lotka-Volterra (Jäger-Beute) - System:

$$\dot{x}_1 = k_1 x_1 - k_2 x_1 x_2$$

$$\dot{x}_2 = k_2 x_1 x_2 - k_3 x_2$$

$$y(t_i) = x_1(t_i) + \epsilon(t_i)$$

FOLIE Lotka Volterra

Begründen.

Für Anwendungen, siehe [71, 258, 235]

6.9.1.3 Bocks multiple shooting Algorithmus

[27, 28]

$$\begin{aligned} \dot{\vec{x}} &= \vec{f}(\vec{p}, \vec{x}) & (\ddagger) \\ y(t_i) &= g(\vec{x}(t_i)) + \epsilon(t_i) \end{aligned}$$

1. Wähle Startschätzungen \vec{p}_0 .
2. Teile $[0, T]$ in viele Segmente.
3. Wähle Anfangswerte $\vec{x}_0(t_j)$ für jedes Segment.
4. Integriere (\ddagger) in jedem Segment.
5. Berücksichtige Kontinuitäts Constraint "sanft".
6. Schätze \vec{p} und $\vec{x}(t_0)$ durch Minimierung von

$$\chi^2(\vec{p}, \vec{x}(t_0)) = \sum_i \left(\frac{y(t_i) - \hat{y}(t_i)}{\sigma_i} \right)^2 + \text{"soft constraint"}$$

Etweas Formaler:

- Verwende Gauß-Newton für Optimierung
- Berücksichtige linearisierte Constraints im up-date Schritt ?? ausbauen ??

Nach Konvergenz

- Constraints sind hart erfüllt
- Verfahren ist dann effektiv Anfangswert-Ansatz

FOLIE Lotka Volterra

Vorteil:

- Größerer Konvergenzradius als Anfangswertansatz

Problem:

- Aufwendige Numerik

Für Anwendungen, siehe [12, 295, 129]

Verallgemeinerung auf PDEs [204], auf DDEs [130]

All models are wrong, but some are useful (G.E.P. Box):

⇒

Lange Zeitreihen lassen sich nicht fiten

Cure:

- Zerschneide Zeitreihen in Stücke
- Fitte Parameter aus Stücken simultan

Good news:

- Verwende maximal fitbare Länge als Maß der Modellgüte
- Bei verschiedenen Modellen zur Modellselektion
- Anwendung siehe unten

”Shadowing”-Problem [96].

⇒

Chaotische Daten sind i.d.R. nicht beliebig lange fitbar.

Daher wiederum:

- Zerschneide Zeitreihen in Stücke
- Fitte Parameter aus Stücken simultan

6.9.1.4 Anomale Fehlerkonvergenz bei deterministischen Systemen

Literatur : [128]

Erinnere:

- Schätzung des Mittelwertes von $N(\mu, \sigma^2)$ aus N Daten:

$$\sqrt{N}(\hat{\mu} - \mu) = N(0, \sigma^2) \quad \text{oder} \quad \hat{\mu} = N\left(\mu, \frac{\sigma^2}{N}\right)$$

- "Standard error of the mean" skaliert mit \sqrt{N}

\sqrt{N} - Skalierung gilt auch für

- Regression
- Autoregression
- HMM [198]
- Stochastische DGI

Grund: Im wesentlichen ist das alles $y_i = ax_i + \epsilon_i$ und jeder Datenpunkt trägt gleichviel bei.

Vorfaktor der Skalierung fallspezifisch.

Betrachte deterministisches System, o.B.d.A. Differenzgleichung:

$$\begin{aligned} x_i &= f(x_{i-1}, p) \\ y_i &= x_i(p) + \eta_i, \quad \eta_i \sim N(0, \sigma^2) \end{aligned}$$

- Parameterschätzer aus:

$$\chi^2(p) = \sum_{i=1}^N \left(\frac{y_i - x_i(p)}{\sigma} \right)^2 \stackrel{!}{=} \min. \quad (6.32)$$

- Schätzfehler $\Delta p = \hat{p} - p_0$ sei klein.
- Entwickle $x_i(p)$

$$x_i(\hat{p}) = x_i(p_0) + g_i \cdot \Delta p, \quad g_i = \left. \frac{dx_i}{dp} \right|_{p=p_0}$$

g_i : Sensitivitäten

- Eingesetzt:

$$\begin{aligned}\chi^2(p) &= \sigma^{-2} \sum_{i=1}^N [y_i - x_i(p)]^2 \\ &= \sigma^{-2} \sum_{i=1}^N [x_i(p_0) + \eta_i - x_i(p_0) - g_i \cdot \Delta p]^2 \\ &= \sigma^{-2} \sum_{i=1}^N [\eta_i - g_i \cdot \Delta p]^2 \stackrel{!}{=} \min\end{aligned}$$

- Minimum aus Ableitung
- Ergibt Normalen-Gleichungen

$$A\Delta p = \sum_{i=1}^N \eta_i g_i \quad \text{mit} \quad A = \sum_{i=1}^N g_i g_i^t$$

- Kovarianzmatrix :

$$\langle \Delta p \Delta p^t \rangle = \sigma_{data}^2 A^{-1}$$

Anschauliche Idee: Fehlerfortpflanzung :

$$\sigma^2(p_i) = \sum_{i=1}^N \sigma_{data}^2 \left(\frac{\partial p_i}{\partial y_i} \right)^2$$

Wichtig: Skaliert $A \propto N$, so ergibt sich \sqrt{N} Gesetz.

- Für g_i gilt (jetzt speziell Determinismus) wegen:

$$x_i = f(x_{i-1}, p)$$

die Rekursion

$$g_i = \left. \frac{dx_i}{dp} \right|_{p=p_0} = \frac{\partial f(x_{i-1}, p)}{\partial p} + \frac{\partial f(x_{i-1}, p)}{\partial x} g_{i-1}$$

und per Induktion:

$$g_i = \sum_{k=0}^{i-1} \frac{\partial f(x_k, p)}{\partial p} \prod_{j=k+1}^{i-1} \frac{\partial f(x_j, p)}{\partial x}$$

- Damit trägt nicht jeder Datenpunkt gleichviel bei. Später tragen mehr bei.

2 Spezialfälle

- 1D chaotische Abbildung mit einem Parameter
 - Denke z.B. an Logistische Abbildung
 $x(i) = 1 - rx^2(i-1)$, $r \in [3.67..., 4]$
 - Wichtige Größe: Lyapunov-Exponent: Mißt (lokale) exponentielle Divergenz benachbarter Trajektorien bei chaotischen Systemen:

$$\lambda = \lim_{i \rightarrow \infty} \frac{1}{i} \log \prod_i \left| \frac{\partial f(x_i)}{\partial x} \right| = \lim_{i \rightarrow \infty} \frac{1}{i} \sum_i \log \left| \frac{\partial f(x_i)}{\partial x} \right|$$

- Gl. (6.9.1.4) umgeschrieben:

$$g_i = z_i u_i$$

mit

$$z_i = \prod_{j=0}^{i-1} f_x(x_j) \quad \text{related to } \lambda : \lambda = \lim_{i \rightarrow \infty} \frac{1}{i} \log |z_i|$$

und

$$u_i = \sum_{k=0}^{i-1} f_p(x_k) z_{k+1}^{-1}.$$

- u_i ist "rauschige Potenzreihe"
 - * $f_p(x_k)$ ist beschränkt
 - * z_{k+1}^{-1} zerfällt exponentiell, da chaotisch
 - * Grenzwert der Potenzreihe: u_∞
- Asymptotisch

$$|z_i| = \exp(\lambda i)$$

und es folgt:

$$\begin{aligned}
A &= \sum_i g_i^2 \approx u_\infty u_\infty^t \sum_{i=1}^N z_i^2 \approx u_\infty u_\infty^t \sum_{i=1}^N (e^{2\lambda})^i \\
&= u_\infty u_\infty^t e^{2\lambda N} \sum_{i=1}^N (e^{-2\lambda})^i = u_\infty u_\infty^t \underbrace{\frac{1}{1 - e^{-2\lambda}}}_{const} e^{2\lambda N}
\end{aligned}$$

A skaliert nicht $\propto N$!

Das heißt konkret:

$$e^{2\lambda N}(\hat{p} - p_0) = N(0, \sigma^2) \quad \text{oder} \quad \hat{p} = N\left(p_0, \frac{\sigma^2}{e^{2\lambda N}}\right)$$

- Geht für > 1 Parameter nicht mehr gut.
 - * Approximation von A in Gl. (6.33) wird singulär, da $u_\infty u_\infty^t$ Rang 1 hat
 - * Empirische Skalierungen: $N, N^{3/2}$

- Periodische Systeme

- Betrachte periodisches System mit einem Parameter p . Periode: $\tau(p)$.
- Dann gilt:

$$x(t, p) = f(\omega(p)t, p), \quad \text{mit } f(\phi, p) \text{ } 2\pi\text{-periodisch in } \phi.$$

- Annahmen (o.B.d.A.):
 - * Samplingintervall: $\Delta t = \tau(p_0)/M, M \in N$
 - * Beobachtungslänge: $K\tau(p_0), K \in N$
- Dann: $i = kM + m, k = 1, \dots, K, m = 1, \dots, M$
- Sensitivitäten g_i :

$$\begin{aligned}
g_i &= \left. \frac{d}{dp} f(\omega(p) i \Delta t, p) \right|_{p=p_0} & (6.33) \\
&= \frac{\partial}{\partial \phi} f(\omega(p_0) i \Delta t, p_0) i \Delta t \frac{d\omega}{dp} + \frac{\partial}{\partial p} f(\omega(p_0) i \Delta t, p_0) & (6.34)
\end{aligned}$$

- 1. Fall: Parameter p hat keinen Einfluß auf Periode: $\frac{d\omega}{dp} = 0$
 * Dann gilt wegen 2π Periodizität:

$$\begin{aligned} g_i &= \frac{\partial}{\partial p} f(\omega(p_0)(kM + m)\Delta t, p_0) \\ &= \frac{\partial}{\partial p} f(\omega(p_0) m \Delta t, p_0) \end{aligned}$$

- * Für A folgt:

$$A = \sum_{i=1}^{KM} g_i^2 = \sum_{k=1}^K A_1 = K A_1$$

mit

$$A_1 = \sum_{m=1}^M \left[\frac{\partial}{\partial p} f(\omega(p_0) m \Delta t, p_0) \right]^2.$$

- * A_1 ist unabhängig von k , A ist $\propto K$ und es folgt " \sqrt{N} Gesetz".

- 2. Fall: ω hängt von p ab: $\frac{d\omega}{dp} \neq 0$

- * 1. Term in Gl. (6.34) ist $\mathcal{O}(i)$ ($\mathcal{O}(k)$)
 * 2. Term in Gl. (6.34) beschränkt, da $\frac{df}{dp}$ periodisch: $\mathcal{O}(1)$
 * Entwicklung von g_i in Potenzen von k

$$g_i = \frac{\partial}{\partial \phi} f(\omega(p_0) m \Delta t, p_0) k M \Delta t \frac{d\omega}{dp} + \mathcal{O}(1).$$

Restterm enthält auch m -Term aus $(kM + m)$.

- * Es folgt mit $A = \sum_i g_i^2$:

$$A = \sum_{k=1}^K k^2 A_2 + \mathcal{O}(k) = \frac{1}{3} K^3 A_2 + \mathcal{O}(K^2)$$

mit

$$A_2 = \sum_{m=1}^M \left[\frac{\partial}{\partial \phi} f(\omega(p_0) m \Delta t, p_0) M \Delta t \frac{d\omega}{dp} \right]^2.$$

unabhängig von k (und i)

- * Damit ist $A \propto K^3$ und es folgt " $N^{3/2}$ " Skalierungsgesetz.

6.9.1.5 Optimale Versuchsplanung

Literatur: [177, 333, 75]

Motiation:

Gegeben

- Ein lineares Regressionsmodell $y_i = ax_i + \epsilon$
- die Möglichkeit im Intervall $[-1;1]$ viermal zu messen

Frage:

Wo mißt man, wenn man möglichst geringe Schätzfehler für a erhalten will ?

Übertrag auf ODEs:

- Betrachte System mit Input u

$$\begin{aligned}\dot{x} &= f(x, p, u) \\ y(t_i) &= h(x(t_i, p, u)) + \eta(t_i), \quad \eta(t_i) \sim N(0, \sigma^2)\end{aligned}$$

- Analog zu oben Sensitivitäten g_i :

$$h(x(t_i, \hat{p}, u)) = h(x(t_i, p_0, u)) + g_i \cdot \Delta p, \quad g_i = \left. \frac{dh_i}{dp} \right|_{p=p_0}$$

- Wie oben:

$$A = \sum_{i=1}^N g_i g_i^t$$

- Kovarianzmatrix :

$$\langle \Delta p \Delta p^t \rangle = \sigma_{data}^2 A^{-1} \stackrel{!}{=} \min \quad (6.35)$$

- Wann messen ? t_i
- Was messen ? $h(\cdot)$
- Wie das System stören ? $u(t)$

- Gl. (6.35) numerisch optimieren

6.9.1.6 Schätzung der Varianz des Beobachtungsrauschens

Zur Ermittlung der Vertrauensintervalle für die Parameter ist es notwendig, die Varianz des Beobachtungsrauschens zu kennen. Diese läßt sich durch

$$\hat{\sigma}^2 = \frac{2}{3(N-2)} \sum_{i=2}^{N-1} \left(y_i - \frac{y_{i-1} + y_{i+1}}{2} \right)^2$$

schätzen [234].

Das Rice-Verfahren:

- x_i wahrer Wert, y_i beobachteter Wert $y_i = x_i + \eta_i$ $\eta_i \sim N(0, \sigma^2)$
- Bilde $\tilde{y}_i = \frac{y_{i+1} + y_{i-1}}{2}$
- $\tilde{y}_i = x_i + \tilde{\eta}_i$ $\tilde{\eta}_i \sim N(0, \frac{\sigma^2}{2})$
- Bilde $\bar{\eta}_i = \tilde{y}_i - y_i = \tilde{\eta}_i - \eta_i$
- Da Varianzen additiv:

$$\hat{\sigma}_i^2 = 2/3(\bar{\eta}_i)^2$$
- Mittel über i

Ist Beobachtungsrauschen nicht unkorreliert, liefert dies verzerrte Schätzer.

- Positiv korreliert: Unterschätzung (schlimm)
- Negativ korreliert: Überschätzung (nicht so schlimm, aber ärgerlich)

Cure:

Verwende Punkte y_i mit größerem Abstand, y_{i-k}, y_i, y_{i+k}

Dann sind die Fehler wieder weiß.

FOLIE Henning DA

Alternative: Schaue im hochfrequenten Bereich des Spektrums

Vorteil des Rice-Verfahrens: Man kann z.B. Fehlermodelle á la $\sigma^2 \propto x$, Poisson-Rauschen bei Intensitätsdaten, ermitteln

6.9.1.7 Identifizierbarkeit

Literatur: [302], Überblick: [46]

(I) Identifizierbarkeit des Zustandes

- System mit bekannten Parametern, kann man den Zustand \vec{x} aus den Daten y ermitteln ?
- Beispiel: Lineare Systeme:

$$\begin{aligned}\dot{\vec{x}} &= A x, & \vec{x} &\in R^m \\ y &= C \vec{x}, & y &\in R\end{aligned}$$

- Betrachte:

$$\begin{aligned}\dot{y} &= C A x \\ \ddot{y} &= C A A x \\ &\dots \\ y^{m-1} &= C A \dots A x\end{aligned}$$

$$x = \begin{pmatrix} C \\ C A \\ C A A \\ \dots \\ C A \dots A \end{pmatrix}^{-1} \begin{pmatrix} y \\ \dot{y} \\ \ddot{y} \\ \dots \\ y^{m-1} \end{pmatrix}$$

- Zustand identifizierbar, wenn Matrix invertierbar.
- Bei nichtlinearen Systemen kompliziert

(II) Identifizierbarkeit der ParameterParameter (θ) strukturell identifizierbar, wenn gilt:

$$\forall \theta_1 \theta_2, \theta_1 \neq \theta_2 \implies \exists t, \text{ so da\ss } g(x(t, \theta_1)) \neq g(x(t, \theta_2))$$

Lokale und praktische Identifizierbarkeit diskutieren

Lineare Systeme: Ähnlich wie oben [49]

Nichtlineare Systeme:

- Taylor-Entwicklung [222]
 - Betrachte:

$$y(t, \theta) = \sum_k \frac{(t - t')^k}{k!} \frac{d^k}{dt^k} y(t', \theta)$$
 - Parameter eindeutig, wenn Lösung der Gleichung eindeutig
 - Probleme:
 - * \nexists obere Grenze der zu untersuchenden Ordnung
 - * Sehr komplizierte nichtlineare Gleichungen
- Characteristic sets [329]
- Vajda [303]
- Differentialalgebraisch [176]
- Alle analytischen Verfahren:
 - Numerisch aufwendig
 - Nur für 2-3 D praktikabel
- Empirisch [203]
 - Schätze Parameter
 - Simuliere Trajektorien
 - Addiere Rauschen
 - Mache dies für immer feiner gesampelte Daten
 - Berechne Konditionszahl κ der Kovarianzmatrix der Parameterschätzfehler
 - 2 Fälle
 - * $\lim_{N \rightarrow \infty} \kappa(N) = \infty \implies$ Parameter nicht identifizierbar
 - * $\lim_{N \rightarrow \infty} \kappa(N) = \text{const} \implies$ Parameter identifizierbar

FOLIE Thorsten

6.9.1.8 Gag am Rande: Universal ODEs

Literatur: [246]

$$n^2 y^{(4)} - 3n^2 y^{(3)} \ddot{y} + 2n(n-1) \dot{y}^2 = 0 \quad n > 3$$

Gegeben $\phi(t)$, $R \rightarrow R$, stetig, $\epsilon(t)$, $R \rightarrow R^+$, stetig.Es existiert Lösung $y(t)$ mit

$$|\phi(t) - y(t)| < \epsilon(t)$$

6.9.1.9 A worked example

Ziel/Nutzen mathematischer Modelle in Biologie:

- Annahmen explizit machen
- Informationen zusammenführen
- Verständnis essentieller Eigenschaften
- Verständnis für die Rolle dynamischer Prozesse, z.B. Rückkopplung
- Komplexität handhaben
- Vorhersage und Kontrolle
- Allgemeine Prinzipien erkennen

Ira paper [276] besprechen.

FOLIE PATHWAY

"You have to put numbers to the arrows" [41]

FOLIEN Daten

Erst das einfache Modell Identifizierbarkeit diskutieren

Wahre Modelle können nicht identifizierbar sein (kennen wir schon von den aggregierten Markov Modellen)

$$\begin{aligned}\dot{x}_1 &= -p_1 x_1 E p o R_A + 2 p_4 x_3^\tau \\ \dot{x}_2 &= -p_2 x_2^2 + p_1 x_1 E p o R_A \\ \dot{x}_3 &= -p_3 x_3 + \frac{1}{2} p_2 x_2^2 \\ \dot{x}_4 &= -p_4 x_3^\tau + p_3 x_3\end{aligned}$$

Letzte Gleichung effektiv, sonst nicht-identifizierbar, weil Prozesse

- Genankopplung
- Deaktivierung

nicht trennbar.

FOLIE Out of sample

FOLIE Hidden components, Sensitivity

Nur ein Zyklus: 45 % der Ausbeute

Take home message:

Sehr viele Systeme sind Populationsdynamik:

- Laser
- Chemie
- Bakterien

Deterministische Modellierung gut, wenn

- Populationen groß, Fluktuationen $\propto 1/\sqrt{N}$
- Transiente Dynamik

6.9.2 Nichtparametrische Methoden

Alles bisherige, z.B.

$$\begin{aligned}\dot{\vec{x}} &= \vec{f}(\vec{x}, \vec{p}) \\ \text{oder} \\ x(i+1) &= f(x(i), x(i-1), \vec{p}) + \epsilon\end{aligned}$$

setzte parametrische Ansätze voraus.

Nichtparametrische Ansätze:

Versuch, $f(\cdot)$ aus Daten zu schätzen.

6.9.2.1 Kernschätzer und Lokal polynomiale Modelle

Konzept stammt ursprünglich aus der nichtlinearen Regression.

ZEICHUNG $f(x)$, $f(x)+\epsilon$, Glättung

Zuerst univariat:

Sei $K(\cdot)$ ein Kern, erinnere Spektralschätzung, im einfachsten Falle:

- $K(\cdot) > 0$ in $[-h, h]$, 0 sonst
- $K(x) = K(-x)$

ZEICHNUNG Drei verschiedenen Kerne

Sei

$$y = f(x)$$

Dann liefert:

$$\hat{f}(x) = \frac{\sum_{i=1}^N K(x - x_i) y_i}{\sum_{i=1}^N K(x - x_i)}$$

einen Kernschätzer für $f(x)$.

Eigenschaften:

- h klein: Bias klein, Varianz groß
- h groß: Bias groß, Varianz klein

Ausführlich in [107, 257]:

- Wahl von h durch Kreuzvalidierung
- Konfidenz-Intervalle für $f(x)$ durch Bootstrap

Zusammenhang: Nichtparametrik und lokalen Polynome-Fits

- Kernschätzer kann formuliert werden als punktweise Auswertung eines lokal angepaßten Polynoms [107].

ZEICHNUNG dazu

- Zu jedem lokalen Polynomfit gehört ein Kern
 - der i.d.R. nicht positiv definit ist.
 - Historische "Angst" vor indefiniten Kernen, weil Anwendung auf Dichte- und Spektralschätzung.
 - Für Regression indefinite Kerne besser.

Merke:

Nichtparametrisch = viele Parameter

Anwendungen auf Zeitreihen: [299, 300]

Verallgemeinert natürlich auf:

$$y = f(x_1, x_2) + \epsilon, \text{ entsprechend } x_i = f(x_{i-1}, x_{i-2}) + \epsilon$$

aber:

Problem:

Kernschätzung in hohen Dimensionen schwierig "Curse of dimensionality"

Bescheidener:

Betrachte nichtlinear additive Modelle der Form [299] :

$$y = f_1(x_1) + f_2(x_2) + \dots + f_m(x_m) + \epsilon \quad (\diamond)$$

Allgemeinere nichtlineare Modelle gibts durch

$$x_i = h_i(\tilde{x}_1, \dots, \tilde{x}_n)$$

Hierfür muß $h_i(\cdot)$ aber parametrisch vorgegeben werden.

Nichtparametrische Schätzung der f_i durch "backfitting-Algorithmus"

- Beobachtung:

Wenn Modell stimmt, gilt

$$f_i(x_i) = E[y - \sum_{j \neq i} f_j(x_j) | x_i]$$

- Schätze $f_i(x_i)$ durch Kernschätzer
- Da $f_j(x_j)$ unbekannt, iteriere das Verfahren:

- Startschätzung:

$$f_i^0(x_i) = 0$$

- Iteriere

$$f_i^{k+1}(x_i) = E[y - \sum_{j \neq i} f_j^k(x_j) | x_i]$$

- Bis sich nichts mehr tut

Eleganter sind die :

6.9.2.2 Optimale Transformationen

Originalpaper: [34]

Erinnere:

Für

$$y = ax + \epsilon$$

quantifiziert der Korrelationskoeffizient:

$$R(y, x) = \frac{\langle yx \rangle}{\sqrt{\langle y^2 \rangle \langle x^2 \rangle}}$$

die Güte der linearen Abhängigkeit.

Betrachte statt:

$$y = ax + \epsilon$$

nun

$$g(y) = f(x) + \epsilon$$

Hier quantifiziert die Maximalkorrelation

$$\Psi(y, x) = |R(g(y), f(x))|, \quad \Psi(y, x) \in [0, 1]$$

die allgemeine Abhängigkeit zwischen x und y .

Gegeben Daten, heißen die Transformationen Φ_0, Φ_1 , für die gilt

$$\Psi(y, x) = \sup_{\Phi_0^*, \Phi_1^*} |R(\Phi_0^*(y), \Phi_1^*(x))| \quad (\ddagger)$$

optimale Transformationen.

Vorteil gegen Gl. (\diamond):

Linearität von Φ_0 ist Modellcheck.

Schätzung von Φ_0, Φ_1 durch

Alternating conditional estimation (ACE) Algorithmus (auch ein "backfitting" Algorithmus)

Andere Formulierung für (\ddagger):

$$E [(\Phi_0^*(y) - \Phi_1^*(x))^2] \stackrel{!}{=} \min \quad (\ddagger)$$

Beobachtung:

- Sei der Erwartungswert von $\Phi_1(x)$ bedingt auf y : $E[\Phi_1(x)|y]$
- Sei $\bar{\Phi}_0(y) = E[\Phi_1(x)|y]$
- Dann minimiert $\bar{\Phi}_0(y)$ Gl. (†) bezüglich $\Phi_0^*(y)$ für gegebenes $\Phi_1(x)$

Motiviert die Iteration:

- Starte mit

$$\Phi_1^{(1)}(x) = E[y|x]$$

- Iteriere

$$\Phi_0^{(i)}(y) = E[\Phi_1^{(i-1)}(x)|y]$$

und

$$\Phi_1^{(i)}(x) = E[\Phi_0^{(i)}(y)|x]/\|E[\Phi_0^{(i)}(y)|x]\|$$

Normierung durch $\|E[\Phi_0^{(i)}(y)|x]\|$, um triviale Lösung $\Phi_0(y) = \Phi_1(x) = 0$ auszuschließen

- Abbruch, wenn sich $E\left[\left(\Phi_0^{(i)}(y) - \Phi_1^{(i)}(x)\right)^2\right]$ nicht mehr ändert.

Die bedingten Erwartungswerte, z.B. $\Phi_1^{(1)}(x) = E[y|x]$, werden geschätzt durch:

- Sortiere x_i aufsteigend
- Permutiere y_i entsprechend
- (z.B. naiver) Kernschätzer:

$$\hat{E}[y|x_i] = \frac{1}{2h+1} \sum_{j=-h}^h y_{i+j}$$

Konvergenzbeweis, siehe [34].

Überträgt sich natürlich auf multivariate nichtlinear additive Modelle der Form:

$$g(y) = f_1(x_1) + f_2(x_2) + \dots + f_m(x_m) + \epsilon$$

Allgemein nichtlineare Modelle gibts durch

$$x_i = h_i(\tilde{x}_1, \dots, \tilde{x}_n)$$

Hierfür muß $h_i(\cdot)$ aber (wieder) parametrisch vorgegeben werden.

Zeitreihenspezifisches:

- Ursprünglich entwickelt für nichtlineare zeitdiskrete Modelle

$$x(i) = f_1(x(i-1)) + f_2(x(i-2)) + \dots + f_m(x(i-m)) + \epsilon(i)$$

oder entsprechend:

$$x(i) = f_1(h_1(x(i-1), \dots, x(i-m))) + \dots + f_k(h_k(x(i-1), \dots, x(i-m))) + \epsilon(i)$$

- Zeitdiskrete Modelle mit Beobachtungsrauschen:
Error-in-Variables Problem (erinnere Zustandsraummodell)
- Differentialgleichungen: Ableitungen müssen aus den Daten geschätzt werden (erinnere "Regressionsansatz")
- Erfahrung: Gut als explorative Methode

Bemerkung:

Im Prinzip gehören auch "Neuronale Netze", Radial Basis Functions und Verwandtes zu den Nichtparametrischen Methoden, siehe Vorlesung "Neuronale Netze".

"Neuronale Netze", Abteilung "Backpropagation" [248] :

$$x(t) = \sum_{i=1}^L w_i f \left(\sum_{j=1}^m w_{ij} x(t-j) \right)$$

mit $f(\cdot)$ sigmoid :

$$\begin{aligned} \lim_{x \rightarrow -\infty} &= a \\ \frac{d}{dx} f(\cdot) &> 0 \\ \lim_{x \rightarrow \infty} &= b \end{aligned}$$

ZEICHNUNG sigmoide Funktion

z.B.:

$$f(x) = \arctan(x), f(x) = \frac{1}{1 + e^{-x}} \text{ oder } f(x) = \tanh(x)$$

ZEICHNUNG NN mit Eingabe- hidden und Ausgabe Schicht

Radial Basis Functions [37]:

Mit $\vec{x}(t-1) = (x(t-1), x(t-2), \dots, x(t-m))$ und Zentren \vec{c}_i

$$x(t) = \sum_{i=1}^L \alpha_i \Phi (|\vec{x}(t-1) - \vec{c}_i|)$$

mit z.B.:

$$\Phi(x) = \frac{1}{1 + x^2/r^2} \text{ oder } \Phi(x) = e^{-x^2/r^2}$$

Pflasterung des Phasenraums mit "Bumps".

- Beachte:

"Neuronale Netze" ($f(x)$) & Radiale Basis Functions ($\Phi(x)$) haben prinzipiell lokalierte Basisfunktionen (wie Kernschätzer ihre äquivalenten lokalen Polynome).

- Daher:

- Ist L klein, ist dieses parametrisch.
- Ist L groß & das ist die Regel, ist dieses nicht-parametrisch.

- Ist L groß:
 - Aus Bias- vs. Varianz- Problem der Kernschätzer wird Überparametrisierungs- (Overfitting-, "Auswendiglernen"-) vs. Überglättungs- Problem.

6.9.2.3 Anwendungen

Mackey-Glass System [179]:

$$\dot{y}(t) = -by(t) + \frac{ay(t-\tau)}{1-y^c(t-\tau)}$$

Ansatz:

$$h(\hat{y}_t) = f_0(y_t) + f_1(y_{t-\tau})$$

FOLIE aus [311]

Weitere Anwendungen in [151, 111]

Besprechung von [295]

Kapitel 7

Punktprozesse

Literatur: [52]

Definition Punktprozess (Physiker):
Stochastischer Prozeß für Ereignisse in der Zeit.

Halbmathematisch:
Punktprozeß $dN(t)$:

$$dN(t) = \sum_i \delta(t - T_i) dt$$

Ganz mathematisch:
Zählprozeß:

$$N(t) = \#(T_j, T_0 < T_j < t) = \int_{T_0}^t dN(t') = \int_{T_0}^t \delta(t' - T_i) dt'$$

Punktprozeß als Inkrement des Zählprozesses

Erinnere: Zusammenhang Wienerprozeß/Weiße Rauschen (als Inkrement des ersteren)

Fundamentalfall:
Poisson Prozeß:

$$\lim_{\delta t \rightarrow 0} \text{prob}(\text{Event in}(t, t + \delta t)) = \rho \delta t, \quad (7.1)$$

ρ : Intensität des Prozesses.

Interevent ($T_j - T_{j-1} = \tau$)-Verteilung:

$$p(T_j - T_{j-1}) = \rho e^{-\rho \tau}$$

$$\langle T_j - T_{j-1} \rangle = 1/\rho$$

2 × Stochastik:

- Stochastisches Dynamisches System :
Stochastik beeinflusst Zeitverlauf in jedem Moment
 - Punktprozeß:
Stochastik beeinflusst, wann "Ereignis" stattfindet.
 - Verbindung (z.B, siehe auch unten):
 - Definiere Schwellwert für stochastisches Dynamisches System
 - Ereignis := Überschreiten der Schwelle
- ZEICHNUNG dazu
- Gibt hübschen Punktprozeß

Allgemeinere Punktprozesse:

- Abhängigkeiten aufeinanderfolgender events, speziell:
- Prozesse mit Refraktärzeit. Beispiel: spikende Neuronen.

ZEICHNUNG Neuron mit Dendrit und Axon

$$H(t) = \int_{t_0}^t e^{-\kappa t'} \text{input}(t')$$

Fire if $H(t) > \text{Schwelle}, t_0 = t$

ZEICHNUNG Integrate and Fire Neuron Zeitverlauf

Ergibt relative Totzeit

Absolute Totzeit durch :

$$H(t) = \int_{t_0+t_{tot}}^t e^{-\kappa t'} \text{input}(t')$$

- Mathematisch: Erneuerungsprozesse, z.B.: $(T_j - T_{j-1} = \tau)$:

– Gamma-Verteilung:

$$p(\tau) = \frac{\rho(\rho\tau)^{\alpha-1} e^{-\rho\tau}}{\Gamma(\alpha)}$$

Weibull Verteilung:

$$p(\tau) = a\rho(\rho\tau)^{a-1} e^{-(\rho\tau)^a}$$

α, a geben mehr Flexibilität.

Analysemethoden, Überblick: [35]:

- Univariate Analyse im Zeitraum [218]

– Oft: Reiz, dann feuert Neuron.

Untersuche bedingte Aktivität.

Betrachte:

Post Stimulus Time Histogram

ZEICHUNG PSTH

FOLIE 2D plot von Ad Aertsen [213]

– Fortlaufender Prozeß:

Ist $SD(\tau) \ll \langle \tau \rangle$, gehe von

T_j über nach $(T_j - T_{j-1})(j) = \tau(j)$

Zeit-Index nach Interbeat-Intervall – Index

FOLIE EKG

Sonst:

Intensität:

$$P_N = \lim_{\Delta t \rightarrow 0} \text{Prob}(\text{Event in } (t, t + \Delta t)) / \Delta t$$

2. Ordnung Intensität:

$$P_{NN}(\tau) = \lim_{\Delta t \rightarrow 0} P(\text{Event für } N(t) \text{ in } (t, t + \Delta t) \text{ und Event für } N(t) \text{ in } (t + \tau, t + \tau + \Delta t)) / \Delta t^2$$

Problem $P_{NN}(\tau)$ divergiert an $\tau = 0$. Grund: $P_N = P_{NN}$ nach Voraussetzung. Stetige Fortsetzung.

Für well behaved Prozesse:

$$\lim_{\tau \rightarrow \infty} P_{NN} = P_N P_N$$

Damit: Autokovarianz:

$$ACF(\tau) = P_{NN}(\tau) - P_N^2$$

Vergleiche Zeitreihen-ACF: $P_{NN}(\tau)$ immer positiv semidefinit.

- Bivariate Analyse im Zeitraum [219, 213, 301]

Analogie zur Kreuzkovarianzfunktion:

$$CCF(\tau) = \langle y(t)x(t - \tau) \rangle$$

nun 2. Ordnung Kreuz-Intensität:

$$P_{NM}(\tau) = \lim_{\Delta t \rightarrow 0} P(\text{Event für } N(t) \text{ in } (t, t + \Delta t) \text{ und Event für } M(t) \text{ in } (t + \tau, t + \tau + \Delta t)) / \Delta t^2$$

Kreuzkovarianzfunktion für Punktprozesse $N(t)$, $M(t)$:

$$CCF(\tau) = P_{NM}(\tau) - P_N P_M$$

- Spektralanalyse [83, 236]

Entweder direkt:

$$f(\omega) = \int e^{-i\omega t} dN(t) = \sum_j e^{-i\omega T_j}$$

Oder Umwandlung in Zeitreihe :

$$x(i) = \begin{cases} 1 & \text{if } T_j \in [i\Delta t, (i+1)\Delta t] \\ 0 & \text{sonst} \end{cases}$$

mit hinreichend feinem Δt (z.B. Sampling Intervall) und normale (F)FT.

Dann jeweils Spektralschätzung wie gehabt.

Bei direkten Verfahren gibt es Divergenz bei $\omega = 0$, ungeeignet.

Oder:

basierend auf:

$$S(\omega) = P_N + \int e^{-i\omega t} (P_{NN}(t) - P_N^2) dt$$

P_N Term aus stetiger Fortsetzung, siehe oben

- Kreuzspektralanalyse [242, 101]

Analog

7.1 Shot noise

Häufig:

Punktprozess triggert (determinischen) Transienten.

Beispiele:

- EKG [152]

2 FOLIEN EKG

- Apnoen [237] [1, 42]

FOLIE aus Rigney: Apnoen, ECG

- Quantum bumps in Photorezeptoren [143]
- Röntgenvariabilität von Schwarz-Loch Kandidaten [296]

FOLIE Cygnus Daten

Fundamentalfall:

Poisson Prozeß (T_j) triggert exponentiellen Zerfall:

$$x(t_i) = \sum_j M \Theta(t_i - T_j) e^{-(t_i - T_j)/\tau} \quad (7.2)$$

Allgemeinere Fälle:

- Ersetze M durch $p(M)$
- Statt Poisson-Prozeß allgemeinerer Punkt-Prozeß
- Kompliziertere Transienten

Zeitskalen:

- $1/\rho > \tau$: Transient beendet vor neuem event, z.B. EKG und Apnoen
- $1/\rho < \tau$: Überlagerung mehrerer Transienten z.B. Quantum bumps und X-ray variability

Interessante Fragen:

1. (Dynamik der) Statistik des Punktprozesses ?

FOLIE : ECG von K/S

2. Dynamik des Transienten ?

FOLIE EKG

PQRST-Welle : Projektion der 3D Dynamik der elektrischen Ausbreitung im Herzen in 1D.

3. Für $1/\rho < \tau$: Ist es shotnoise oder Dynamisches System ?

Ad [3.] Besprechung von [296]

Die Daten:

- Leuchtstärkevariabilität von Cygnus X-1
- "Low State": 400.000 Punkte
- "Intermediate State": 650.000 Punkte

FOLIE FIG. 1 intermediate Daten

- Erfahrung sagt:
 - Sieht aus wie AR[1] Prozeß
 - Signal zu Rausch Verhältnis nicht groß
- Fitte Zustandsraummodell an die Daten
 - FOLIE Per/Spec
- Ergebnis :
 - Spektrum des Modells paßt sehr gut zu Periodogramm der Daten.
 - Spektrum ZRM[1] bis ZRM[4] auf Strichdicke identisch.
 - ⇒ Es ist kein nichtlineares System.
- Cygnus X-1 ein AR[1] Prozeß [225] ?

Erinnere Spektrum AR[1] Prozeß

$$S(\omega) = \frac{\sigma^2}{|1 - a e^{-i\omega}|^2} = \frac{\sigma^2}{1 + a^2 - 2a \cos(\omega)}. \quad (7.3)$$

Betrachte Shotnoise Prozeß aus Gl. (7.2)

$$x(t_i) = \sum_j M \Theta(t_i - T_j) e^{-(t_i - T_j)/\tau} \quad (7.4)$$

Läßt sich formal schreiben als (Δt : Samplingzeit):

$$x(t) = e^{-\Delta t/\tau} x(t - \Delta t) + \epsilon(t) = ax(t - \Delta t) + \epsilon(t), \quad (7.5)$$

mit $\epsilon(t)$ der Form:

$$\epsilon(t) = \begin{cases} 0 & \text{with probability } 1 - \rho\Delta t \\ M & \text{with probability } \rho\Delta t \end{cases}. \quad (7.6)$$

Da das Spektrum eines linearen Prozesses nicht von der Verteilung des Rauschens abhängt, folgt:

AR[1]-Prozeß und klassischer Shot Noise Prozeß haben das selbe Spektrum.

oder:

Am Spektrum kann man nicht zwischen AR[1] und (klassischem) shot noise unterscheiden.

Nebenbemerkung: Für

- $\rho\Delta t \approx 1$
- $M \sim N(0, \sigma^2)$

geht klassischer Shot noise in AR[1] Prozeß über.

Aber:

Für $\rho\Delta t \approx 1$ ist der Prozeß stark untersampelt, es feuert ständig.

Inverses Problem:

Kann man aus Daten zwischen AR[1] und shot noise unterscheiden?

- Erinnere:
Nur (gaußsche) ARMA-Prozesse sind zeitumkehrinvariant [320]. Argument war ACF-Symmetrie
- Maß für Zeitumkehrinvarianz [281]:

$$Q(m) = \frac{\langle (x(t+m) - x(t))^3 \rangle}{\langle (x(t+m) - x(t))^2 \rangle}. \quad (7.7)$$

Misst die (A)symmetrie aufsteigender und abfallender Flanken.

- Problem: Verteilung von $Q(m)$ im Endlichen für lineare Prozesse nicht bekannt.
- Lösung: Bootstrap, hier sogenannte surrogate data [281]
 - Berechnen $Q(m)$ für die Daten
 - Fourier-transformiere die Daten
 - Ersetze die Phasen der FT durch Realisierungen von $U[0, 2\pi]$.
 - Inverse FT, ergibt Realisierung eines Prozesses, der nur die linearen Eigenschaften des Ausgangsprozesses hat. Warum ?
 - Berechne $Q(m)$ für diese surrogate data.
 - Mache dies häufig.
 - Berechne $\langle Q(m) \rangle_{surr}$ und $\sigma(Q(m))_{surr}$
 - Bilde

$$S(m) = \frac{|Q(m) - \langle Q(m) \rangle_{surr}|}{\sigma(Q(m))_{surr}}, \quad (7.8)$$
 - $S > 2.6$: signifikant zum 0.01 % Niveau

FOLIE Fig. 2 Simus

- Problem: Beobachtungsrauschen \implies Unterschätzung von $S(m)$ speziell für kleine m . Warum ? Auf kurzen Skalen SNR besonders schlecht.
- Lösung: Benutze Kalman/Smoothing-Filter zur Rauschreduktion.

FOLIE FIG. 6 Ergebnisse

Artefakte des Kalman-Filters ?
Simulation unter H_0

FOLIE FIG. 7 SIMUS

- Problem : Multiples Testen

FOLIE Fig. 4

- Lösung: Es gibt viele Daten.

Berechne aus Segmenten:

- $\langle S(m) \rangle$
- $\sigma(S(m))$

- FINALES:

FOLIE FIG. 8,9 Results

- Astrophysikalische Interpretation:
 - Instabilitäten der Aggregationsscheibe erzeugen X-ray δ -Pulse
 - Multiples Compton-Streuen an umgebenen Elektronen macht daraus Transienten
 - Mathematisch: Projektion von stochastischer raumzeitlicher Reaktions (mit e^-)-Diffusions (mit c) Dynamik auf die Zeitachse.
- Next step: Was für ein shot noise Prozeß macht so ein $S(m)$ wie der intermediate state (ein Direktes Problem :-) ?

Kapitel 8

Modellselektion

Bisher: Modellklasse vorausgesetzt.

Im allgemeinen:

- Gegeben Daten und m Modelle: Welches ist das wahre ?
- Oder zumindest das "beste".
- Ein alter Traum: "Equations of motion from data series" [55]
- But: "All models are wrong but some are useful"

Unterschied:

- Vorhersage

Mean Square Prediction Error = Bias² + Varianz

Falls: Wenige Daten & viele Parameter \implies Parameterschätzer große Varianz

Oft sinkt Mean Square Prediction Error durch:

$$\hat{\theta}_S = c \hat{\theta}_{MLE}, \quad c \in [0, 1)$$

Produziert bias, verringert Varianz

– Shrinkage [283, 97]

- James Stein Schätzer [271, 140]
- Ridge Regression [124]
- Bayes'sche Methoden, Prior
- Regularisierung [178]
- Modellierung
 - Verzerrungsfrei
 - Möglichst kleine Varianz

Nur MLE, resp. QMLE

Konsistentes Modellselektionsverfahren:

Für $N \rightarrow \infty$ wird das wahre Modell mit Wahrscheinlichkeit = 1 gefunden

Formale Modellselektionsverfahren:

- AR/MA Prozesse
 - ACF/PACF
 - FOLIE PACF
 - Test auf Parameter=0 $a_i \sim N(0, \tilde{\sigma}^2)$ für $i > p$
 - FOLIE AR-Parameter-Konfidenzen
- Mutter aller Modellselektions-Test: **F-Test**

Gegeben

 - Gauß'sche Fehler, Least-Squares-Schätzproblem
 - Modell M_1 mit p_1 Parametern, $\chi^2(M_1)$, DoFs: $N - p_1$
 - Modell M_2 mit p_2 Parametern, $\chi^2(M_1)$, DoFs: $N - p_2$
 - Modelle geschachtelt, $p_1 > p_2$
 - H_0 : M_2 ist zulässige Vereinfachung von M_1

Unter H_0 :

$$F = \frac{\chi^2(M_2) - \chi^2(M_1)/(p_1 - p_2)}{\chi^2(M_1)/(N - p_1 - 1)}$$

ist F -verteilt mit $p_1 - p_2$ und $N - p_2 - 1$ Freiheitsgraden.

Erhöht man kritisches Niveau mit N , ergibt sich konsistentes Selektionsverfahren [16, 209].

- **Likelihood ratio tests**

Allgemein für MLE:

Selbes Setting wie oben ($L(\cdot)$ log-likelihood):

Unter H_0 :

$$2(L(M_1) - L(M_2)) \sim \chi_{p_1 - p_2}^2$$

Erinnere: Nichtstandard Test-Situationen

Erhöht man kritisches Niveau mit N , ergibt sich konsistentes Selektionsverfahren.

LRT für Regression = F-Test, siehe z.B. [262]

- **8.1 Akaike Information Criterion (AIC)**

¹ [5, 6] schön dargestellt in [169]

Ansatz: Vereinigung von Parameterschätzung und Modellselektion

- Likelihood des gefitteten Modells ist größer/gleich der des wahren Modells.
- Schätze Ausmaß der Überschätzung
- Formal analog zu Cross-Validation [273]
- Führt auf

$$AIC(p) = -2 \log(\text{Likelihood}) + 2p$$

¹Akaike selber nannte es in [6] An Information Criterion (AIC :-)

- Zur Modellwahl, wähle Modell mit kleinstem AIC.

Kommentare:

- Auf Grund der Einfachheit sehr beliebt.
- Aber:
Betrachte:
Geschachtelte Modelle mit $\Delta p = 1$

$$\begin{aligned} AIC(p_1) &= -2(L(M_1)) + 2p_1 \\ AIC(p_2) &= -2(L(M_2)) + 2p_2 \\ AIC(p_1) - AIC(p_2) &= -2(L(M_1) - L(M_2)) + 2 \end{aligned}$$

Erinnere LRT:

Unter H_0

$$2(L(M_1) - L(M_2)) \sim \chi_1^2$$

Ergo: AIC ist LRT mit kritischem Wert α

$$\chi^2(2) = \alpha, \quad \text{ergibt } \alpha = 15.7$$

Im testtheoretischen Sinne: 15.6 % Fehler 1. Art
Führt systematisch zu zu komplexen Modellen

- Kein konsistentes Modellselektionsverfahren
- Verhalten für Parameter auf dem Rand und bei nicht-identifizierbaren Parametern unklar

Literatur: [264]

• 8.2 Bayes'sches Information Criterion (BIC)

[261]

- Annahme: Schwächste Bayes'sche Priors

- Berücksichtigt Datenanzahl

$$BIC = -2 \log(\text{Likelihood}) + p \log(N)$$

Signifikanzniveau ist für einen Parameter Unterschied [280]

$$\text{Prob}(\chi_1^2 > \log(N))$$

- Ist wahres Modell in der Testmenge, ergibt Wahl des kleinsten BIC konsistentes Modellselektionsverfahren
- Vergleiche AIC vs. BIC, siehe [23, 10, 280, 164, 187]

• 8.3 Minimum Description Length (MDL)

[238]

Anzahl der bits, um

- Model
- Parameter
- Residuen

zu beschreiben.

Asymptotisch äquivalent zu BIC

Für Diskussion, siehe [239, 318]

Speziell für Vorhersage:

• 8.4 Mallows' C_p

[180, 181]

$$C_p = \text{PredictionError} - n + 2p$$

• 8.5 Final Prediction Error (FPE)

[4]

$$FPE = PredictionError \frac{n+p}{n-p}$$

Selektionsstrategien für F-Test, LRTs:

- Forward Selection
 - Aufsteigend komplizierte Modelle
 - Nachteile:
 - * Falsch Negative. Early stopping
 - * Es gibt mit unter keine natürliche Ordnung im Nichtlinearen
- Backward Selection
 - Vom generellsten Modell absteigend
 - Nachteil
 - * Was ist das generellste Modell ?
 - * Existenz des Obermodells, siehe HMMs
- Stepwise Selection

Nach jedem Forward-Schritt, wieder Backward-Schritte

Ist empfohlen
- Für Bayesianer:

Reversible jump Markov Chain Monte Carlo [43, 241]

Siehe auch Hinde [119]

Noch irgendwo richtig einzubauen: Cox [50] baut allgemeines Obermodell
Für Modellselektion wichtig, ob Einsicht oder Voraussage das Ziel ist. AIC
gut für Voraussage, da es da schlimm ist, wenn Modell zu klein, aber nicht
viel ausmacht, wenn ein bißchen zu groß.

AIC und BIC gehen nicht in hierarchischen Modellen (Höhenried)

Formale Selektionsverfahren können arg unphysikalisch sein :

- AIC und sunspots
 - AR Modelle auf Sonnenfleckendaten führt auf $a_{11} \neq 0$, Rest =0,
 - macht copy and paste, nicht Dynamik.
- Speziell bei misspezifizierten Modellen:
 Physiologischer Tremor mit AR, AIC führt auf Ordnung 52

FOLIE AR-Residuen bei phys. Tremor

Ergo:

Traue nie blind der Statistik:

Möglichst viele (unabhängige) Checks.

Worked examples:

- Phys.Tremor:
 - Knick in Residualvarianz (formal F-Test)
 - Residuen weiß ?
 - Residuen gauß'sch ?
- Vergleich von Vorhersagen des angepassten Modells mit empirischen Größen
 - Vergleiche Spektrum des Prozesses mit Periodogramm
 Aber Warnung: Cygnus X-1
 - Attraktoren, siehe Schaltkreis.
 - Offenzeitverteilung, siehe HMM und Ionenkanäle
- Reproduktion der Ergebnisse auf simulierten Daten, siehe HMM
- Länge vorhersagbarer Trajektorien bei deterministischen Systemen
- Vorhersage unabhängiger Experimente

Es gibt noch viel zu tun !

Kapitel 9

Prozesse der Finanzmathematik

Geometrisch Brown'sche Bewegung

Black Scholes [25]

qq-plots

[297]

Galla/St.Johann

[215]

Zwei Annahmen für Zuwächse, beide nicht erfüllt:

- Homeoskadistisch
- Gauß'sch

GARCH Modelle [30]

Kapitel 10

Neuronale Netze

”Neuronale Netze bilden ihre Begriffe selber”

”Neural nets are an inefficient implementation of well-known algorithms”

(NAG)

Pawelczik gucken

10.1 Ein bißchen Neurobiologie

Menschlicher Cortex 10^{10} Neurone, 10^4 Kopplungen pro Neuron, Kopplungen über Synapsen,

Dendriten: Inputseitig

Axone: Output

Hodgkin-Huxley [122, 123]

Cayal

Ruhepotential -80 mV

Depolarisieren, hyperpolarisieren

Pyramidenzellen, Interneurone

Integrate-and-fire Mechanismus

One neuron, one neurotransmitter, i.e. entweder excitatorisch (senkt Potential) oder inhibitorisch (erhöht Potential)

$$Pot(t) = \int dt' e^{-(t-t')/\tau} input(t-t'), \quad \tau : \text{Leak-Parameter}$$

Fire, if

$$Pot(t) > S, \text{ danach } Pot(t) = 0.$$

ZEICHUNGEN Integrate and fire, τ groß, τ klein

Zwei Fälle:

- τ groß, Ratenkodierung, zeitliche Summation, Anatomische Verbindung
- τ klein, Raumzeitliche Kodierung, räumliche Summation, Koinzidenzdetektor, Funktionelle Verbindung

Hebb'sches Lernen: Wenn Verbindung genutzt, dann wird sie verstärkt [208]

Sigmoidale Funktion: Arbeitspunkt läßt sich gut schieben.
Zitate aus Eckmiller Buch

10.2 Hopfield

[126, 127]

10.3 Back-Propagation

Historisches [188, 189]

"Principles of Neurodynamics" (1959) [243]

Killerbuch: [200]

Revival: [322]

KI als Alternative

[248]

es gibt nicht "DAS neuronale Netz"

Universelle Approximierbarkeit mit Rückführung aus sin/cos [131, 56]

Fischer'sche Diskriminanzanalyse [69, 104]

ROC Kurven

Sensitive und spezifische Tests

10.4 Clustering

Allgemein: [139]

[231]

adjusted Rand-Index: [132, 199]
 Erst PC, dann Clustering [45] [331]

10.5 Kohonen

[166] [175]

[240]

Tessellation Voronoi-Diagramme (Buch von Informatiker Algorithmen zitieren)

10.6 Reinforcement learning

Irgendwo: Bindung nach van der Malsburg [305]

Neurosprachen Krams

Gegenüberstellung: Neurosprache vs. konventionelle Sprache

Neurosprache	konventionelle Sprache
lernen, Selbstorganisation	schätzen, fitten
Back Propagation	Kettenregel
Back Propagation	Steepest descent
verallgemeinern	glätten
auswendig lernen	interpolieren
Neurone	Basisfunktionen
Gewichte	Parameter
pruning	Variablenselektion
erkennen	klassifizieren
# versteckte Neurons	Modellselektion
Muster	Beobachtungen
supervised learning	Regression / Diskriminanzanalyse
competitive learning	Clusteranalyse
unsupervised learning (Kohonen)	Optimierung der Mutual Information
higher order neurons	Wechselwirkungen

Häufig: "Manchmal mögen neuronale Netze ja helfen."

Aber nur, weil eine Einlauf manchmal hilft, macht man ihn ja auch nicht bei jedem Problem.

"Neuronale Netze" suggeriert mehr als sind sind. Man könnte den Ansatz

$$y_i = ax_i^2 + \epsilon$$

"Den alleinseelig machenden Ansatz" nennen, und dann immer fragen, warum man nicht "Den alleinseelig machenden Ansatz" verwendet anstatt eines Neuronalen Netzes.

Kapitel 11

Jobs, careers, the future and all that

Was fehlt:

- PCA und ICA [173]
- EOF [310]

Bei Interesse an einem HiWi-Job,
einer Diplom- oder Doktorarbeit:

Send me e-mail: jeti@fdm.uni-freiburg.de

Einladung zur Vorlesung im WS:
Konzepte der Nichtlinearen Dynamik

Literaturverzeichnis

- [1] L.A. Aguirre, V.C. Barror, and A.V.P Souza. Nonlinear modeling and analysis of sleep apnea time series. *preprint*, 1998.
- [2] M. Aigner. *Diskrete Mathematik*. Vieweg, Braunschweig, 1999.
- [3] H. Akaike. Some problems in the application of the cross-spectral method. In B. Harris, editor, *Spectral Analysis of Time Series*, New York, 1967. Wiley.
- [4] H. Akaike. Fitting autoregressive models for prediction. *Ann. Inst. Stat. Math.*, 21:243–247, 1969.
- [5] H. Akaike. Information theory and an extension of the maximum likelihood principle. In B.N. Petrov and F. Csaki, editors, *2nd International Symposium on Information Theory*, pages 267–281, Budapest, 1973. Akademiai Kiado.
- [6] H. Akaike. A new look at the statistical model identification. *IEEE Trans. Automatic Control*, AC-19:716–723, 1974.
- [7] W.F. Ames. *Numerical Methods for Partial Differential Equations*. Academic Press, Boston, 1992.
- [8] J.S. Andrade Jr., I. Wainer, J.M. Filho, and J.E. Moreira. Self-organized criticality in the El Nino southern oscillation. *Phys. Lett. A*, 215:331–338, 1995.
- [9] D.W.K. Andrews and W. Ploberger. Admissibility of the likelihood ratio test when a nuisance parameter is present only under the alternative. *Ann. Stat.*, 23:1609–1629, 1995.
- [10] A.C. Atkinson. Likelihood ratios, posterior odds and information criteria. *J. Econometrics*, 16:15–20, 1981.
- [11] M. Ausloos and K. Ivanova. Power-law correlations in the Southern-Oscillation-Index fluctuations characterizing El Niño. *Phys. Rev. E*, 63(4), 2001.

- [12] E. Baake and J.P. Schlöder. Modeling the fast fluorescence rise of photosynthesis. *Bull. Math. Biol.*, 54:999–1021, 1992.
- [13] L.A. Baccalá and K. Sameshima. Partial directed coherence: a new concept in neural structure determination. *Biol. Cybern.*, 84:463–474, 2001.
- [14] P. Bak, C. Tang, and K. Wiesenfeld. Self-organized criticality: an explanation for $1/f$ noise. *Physical Review Letters*, 59:381–384, 1987.
- [15] M.S. Bartlett. *Stochastic Processes*. Cambridge University Press, Cambridge, 1978.
- [16] P. Bauer, B.B. Pötscher, and P. Hackl. Model selection by multiple test procedures. *Statistics*, 1:39–44, 1988.
- [17] L.E. Baum, T. Petrie, G. Soules, and N. Wiess. A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *Ann. Math. Stat.*, 41:164–171, 1970.
- [18] J. Bechhoefer. Feedback for physicists: a tutorial essay on control. *Rev. Mod. Phys.*, 77:783–836, 2005.
- [19] J.D. Becker, J. Honerkamp, J. Hirsch, U. Fröbe, E. Schlatter, and R. Greger. Analysing ion channels. *Pflüg. Arch*, 426:328–332, 1994.
- [20] Y. Benjamini and Y. Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Roy. Stat. Soc. B*, 57:289–300, 1995.
- [21] J. Beran. Statistical methods for data with long-range dependence. *Statistical Sciences*, 7:404–427, 1992.
- [22] J. Beran. *Statistics for Long-Memory Processes*. Monographs on Statistics and Applied Probability. Chapman & Hall, 1994.
- [23] R.J. Bhansali and D.Y. Downham. Some properties of the order of an autoregressive model selected by a generalization of Akaike’s FPE criterion. *Biometrika*, 64:547–551, 1977.
- [24] B.M. Bibby and M. Sørensen. Martingale estimation functions for discretely observed diffusion processes. *Bernoulli*, 1:17–39, 1995.
- [25] F. Black and M. Scholes. The pricing of options and corporate liabilities. *J. Pol. Econ.*, 81:637–654, 1973.

- [26] P. Bloomfield, H.L. Hurd, and R.B. Lund. Periodic correlation in stratospheric ozone data. *J. Time Ser. Anal.*, 15:127–150, 1992.
- [27] H.G. Bock. Numerical treatment of inverse problems in chemical reaction kinetics. In K.H. Ebert, P. Deuffhard, and W. Jäger, editors, *Modelling of Chemical Reaction Systems*, volume 18, pages 102–125, New York, 1981. Springer.
- [28] H.G. Bock. Recent advances in parameter identification for ordinary differential equations. In P. Deuffhard and E. Hairer, editors, *Progress in Scientific Computing*, volume 2, pages 95–121, Boston, 1983. Birkhäuser.
- [29] P. Boeijinga and F.D. da Silva. A new method to estimate time delays between EEG signals applied to beta activity of the olfactory cortical areas. *Electroenceph. clin. Neurophys.*, 73:198–205, 1988.
- [30] T. Bollerslev. Generalized autoregressive conditional heteroskedasticity. *J. Econometrics*, 31:307–327, 1986.
- [31] L. Borland and H. Haken. Unbiased determination of forces causing observed processes. *Z. Phys. B - Cond. Matt.*, 88:95–103, 1992.
- [32] L. Borland and H. Haken. Unbiased estimate of forces from measured correlation functions, including the case of strong multiplicative noise. *Ann. Phys.*, 1:452–459, 1992.
- [33] G.E.P. Box and G.M. Jenkins. *Time Series Analysis: Forecasting and Control*. Holden-Day, San Francisco, 1970.
- [34] L. Breiman and J.H. Friedman. Estimating optimal transformations for multiple regression and correlation. *J. Am. Stat. Assoc.*, 80:580–598, 1985.
- [35] D.R. Brillinger. Nerve cell spike train data analysis: a progression of technique. *J. Am. Stat. Ass.*, 87:260–271, 1992.
- [36] P.J. Brockwell and R.A. Davis. *Time Series: Theory and Methods*. Springer, New York, 1998.
- [37] D.S. Broomhead and D. Lowe. Multivariate functional interpolation and adaptive networks. *Complex Systems*, 2:321–355, 1988.
- [38] R. Brown. Calculating Lyapunov exponents for short and/or noisy data sets. *Phys. Rev. E*, 47:3962–3969, 1993.

- [39] P. Bühlmann. Locally adaptive lag-window spectral estimation. *J. Time Series Anal.*, 17:247–270, 1996.
- [40] A. Buse. The likelihood ratio, Wald and Lagrange multiplier test: an expository note. *The American Statistician*, 36:153–157, 1982.
- [41] P. Campbell. Can physics deliver another biological revolution? *Nature*, 397:89, 1999.
- [42] L. Cao and A. Mees. Determinism in multi-channel physiological data. *preprint*, 1998.
- [43] B.P. Carlin and S. Chib. Bayesian model choice via Markov chain Monte Carlo methods. *J. R. Stat. Soc. B*, 57:473–484, 1995.
- [44] R.J. Carroll, D. Ruppert, and L.A. Stefanski. *Measurement Error in Nonlinear Models*. Chapman and Hall, London, 1995.
- [45] W.C. Chang. On using principal components before separating a mixture of two multivariate normal distributions. *Appl. Statist.*, 32:267–275, 1983.
- [46] M.J. Chappell, K.R. Godfrey, and S. Vajda. Global identifiability of the parameters of nonlinear systems with specified inputs: a comparison of methods. *Math. Biosciences*, 102:41–73, 1990.
- [47] D.G. Childers. *Modern Spectral Analysis*. IEEE Press, New York, 1978.
- [48] P.-S. Chu and R.W. Katz. Modeling and forecasting the southern oscillation: A time domain approach. *Monthly Weather Review*, 113:1876–1888, 1985.
- [49] C. Cobelli, A. Lepschy, and G.R. Jacur. Identifiability of compartmental systems and related structural properties. *Math. Biosci.*, 44:1–18, 1979.
- [50] D.R. Cox. Tests of separate families of hypotheses. In *Proc. Fourth Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, pages 105–123. University of California Press, 1961.
- [51] D.R. Cox and D.V. Hinkley. *Theoretical Statistics*. Chapman & Hall, London, 1994.
- [52] D.R. Cox and V. Isham. *Point Processes*. Chapman and Hall, London, 1980.
- [53] D.R. Cox and N. Wermuth. *Multivariate Dependencies*. Chapman and Hall, London, 1996.

- [54] J. Cremers and A. Hübler. Construction of differential equations from experimental data. *Z. Naturforsch.*, 42a:797–802, 1987.
- [55] J.P. Crutchfield and B.S. McNamara. Equations of motion from data series. *Complex Systems*, 1:417, 1987.
- [56] G. Cybenko. Approximation by superpositions of a sigmoidal function. *Mathematics of Control Signals Systems*, 2:303–314, 1989.
- [57] F.D. da Silva, J.P. Pijn, and P. Boeijinga. Interdependence of EEG signals: linear vs. nonlinear associations and the significance of time delays and phase shifts. *Brain Topography*, 2:9–18, 1989.
- [58] D. Dacunha-Castelle and D. Florens-Zmirou. Estimation of the coefficients of a diffusion from discrete observations. *Stochastics*, 19:263–284, 1986.
- [59] R. Dahlhaus. Spectral analysis with tapered data. *J. Time Ser. Anal.*, 4:163–175, 1983.
- [60] R. Dahlhaus. Graphical interaction models for multivariate time series. *Metrika*, 51:157–172, 2000.
- [61] R. Dahlhaus, M. Eichler, and J. Sandkühler. Identification of synaptic connections in neural ensembles by graphical models. *J. Neuroscience Methods*, 77:93–107, 1997.
- [62] R. Dahlhaus and L. Giraitis. On the optimal segment length for parameter estimates for locally stationary time series. *J. Time Ser. Anal.*, 19:629–655, 1998.
- [63] R. Dahlhaus and D. Janas. A frequency domain bootstrap for ratio statistics in time series analysis. *Annals Stat.*, 24:1934–1963, 1995.
- [64] I. Daubechies. *Ten lectures on wavelets*. SIAM, Philadelphia, 1992.
- [65] A.C. Davison and D.V. Hinkley. *Bootstrap Methods and their Application*. Cambridge University Press, Cambridge, 1997.
- [66] A.P. Dempster, N.M. Laird, and D.B. Rubin. Maximum likelihood from incomplete data via EM algorithm. *J. Roy. Stat. Soc.*, 39:1–38, 1977.
- [67] D.L. Donoho and I.M. Johnstone. Adapting to unknown smoothness via wavelet shrinkage. *J. Am. Stat. Ass.*, 90:1200–1224, 1995.
- [68] I.M. Dremlin, O.V. Ivanov, and V.A. Nechitailo. Wavelets and their use. arXiv:hep-ph/0101182, 2001.

- [69] O. Duda and P.E. Hart. *Pattern classification and Scene Analysis*. Wiley, New York, 1973.
- [70] R. Durbin, S. Eddy, A. Krogh, and Mitchson G. *Biological Sequences Analysis*. Cambridge University Press, Cambridge, 1998.
- [71] L. Edsberg and P. Wedin. Numerical tools for parameter estimation in ODE-systems. *Opt. Meth. Software*, 6:193–217, 1995.
- [72] B. Efron. Introduction to: James and Stein: Estimation with quadratic loss. In S. Kotz and N.L. Johnson, editors, *Breakthroughs in Statistics 1*, pages 437 – 443. Springer, 1992.
- [73] B. Efron and R.J. Tibshirani. *An Introduction to the Bootstrap*. Chapman & Hall, New York, 1998.
- [74] R.J. Elliott, L. Aggoun, and J.B. Moore. *Hidden Markov models : Estimation and Control*. Springer, New York, 1995.
- [75] D. Faller, U. Klingmüller, and J. Timmer. Simulation methods for optimal experimental design in systems biology. *Simulation: Trans. Soc. Modeling Computer Simulation*, 79:717–725, 2003.
- [76] R.A. Fischer. Tests of significance in harmonic analysis. *Proc. Roy. Soc. Ser. A*, 125:54–59, 1929.
- [77] D. Florens-Zmirou. Approximate discrete-time schemes for statistics of diffusion processes. *Statistics*, 4:547–557, 1989.
- [78] G.E. Forsythe. Generation and use of orthogonal polynomials for data – fitting with a digital computer. *J. Soc. Indust. Appl. Math.*, 5:74–88, 1957.
- [79] J. Franke and W. Härdle. On bootstrapping kernel spectral estimates. *Annals of Statistics*, 2:121–145, 1992.
- [80] D.R. Fredkin, M. Montal, and J.A. Rice. Identification of aggregated Markovian models: application to the nicotine acetylcholine receptor. In L.M. Le Cam and R.A. Olshen, editors, *Proceedings of the Berkeley Conference in honor of Jerzy Neyman and Jack Kiefer*, volume I, pages 269–289, Belmont, 1985. Wadsworth, Inc.
- [81] D.R. Fredkin and J.A. Rice. On aggregated Markov processes. *J. Appl. Prob.*, 23:208–214, 1986.

- [82] D.R. Fredkin and J.A. Rice. Bayesian restoration of single-channel patch clamp recordings. *Biometrics*, 48:427–448, 1992.
- [83] A.S. French and A.V. Holden. Alias-free sampling of neuronal spike trains. *Kybernetik*, 8:165–171, 1971.
- [84] W.A. Fuller. *Measurement error models*. John Wiley, New York, 1987.
- [85] C.W. Gardiner. *Handbook of Stochastic Methods*. Springer, Berlin, 1997.
- [86] W.A. Gardner. *Introduction to Random Processes with Application to Signals and Systems*. MacGraw-Hill, New York, 1990.
- [87] A. Gelb. *Applied Optimal Estimation*. MIT – Press, Cambridge, 1989.
- [88] W.R. Gilks, S. Richardson, and D.J. Spiegelhalter. *Markov chain Monte Carlo in practice*. Chapman & Hall, London, 1997.
- [89] M. Giona, F. Lentini, and V. Cimiagalli. Functional reconstruction and local prediction of chaotic time series. *Phys. Rev. A*, 44:3496–3502, 1991.
- [90] V.P. Godambe. *Estimating Functions*. Clarendon Press, Oxford, 1991.
- [91] V.P. Godambe and C.C. Heyde. Quasi-likelihood and optimal estimation. *Int. Stat. Rev.*, 55:231–244, 1987.
- [92] N.R. Goodman. Statistical tests for stationarity within the framework of harmonizable processes. Technical report, Rocketdyne Research, 1965.
- [93] J. Gradisek, S. Siegert, R. Frierich, and I. Grabec. Analysis of time series from stochastic processes. *Phys. Rev. E*, 62:3146–3155, 2000.
- [94] M.D. Graham et al. Effects of boundaries on pattern formation: Catalytic oxidation of co on platinum. *Science*, **264**:80–82, 1994.
- [95] J. Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* 37, 37:424–438, 1969.
- [96] C. Grebogi, S.M. Hammel, J.A. Yorke, and T. Sauer. Shadowing of physical trajectories in chaotic dynamics: Containment and refinement. *Phys. Rev. Lett.*, 65:1527–1530, 1990.
- [97] M.H.J. Gruber. *Improving Efficiency by Shrinkage: The James-Stein and Ridge Regression Estimators*. Marcel Dekker, New York, 1998.

- [98] L. Györfi, W. Härdle, P. Sarda, and P. Vieu. *Nonparametric Curve Estimation from Time Series*. Number 60 in Lecture Notes in Statistics. Springer, New York, 1989.
- [99] P. Hall and S.R. Wilson. Two guidelines for bootstrap hypothesis testing. *Biometrics*, 47:757–762, 1991.
- [100] J.D. Halley, A.C. Warden, S. Sadedin, and W. Li. Rapid self-organized criticality: Fractal evolution in extreme environments. *Phys. Rev. E*, 70:036118, 2004.
- [101] D.M. Halliday, J.R. Rosenberg, A.M. Amjad, P. Breeze, B.A. Conway, and S.F. Farmer. A framework for the analysis of mixed time series/point process data – Theory and application to the study of physiological tremor, single motor unit discharges and electromyograms. *Prog. Biophys. molec. Biol.*, 64:237–278, 1995.
- [102] J.D. Hamilton. *Time Series Analysis*. Princeton University Press, Princeton, NJ, 1994.
- [103] B. V. Hamon and E. J. Hannan. Spectral estimation of time delay for dispersive and non-dispersive systems. *Appl. Statist.*, 23:134–142, 1974.
- [104] D.J. Hand. *Discrimination and Classification*. Wiley, New York, 1992.
- [105] E.J. Hannan and P.J. Thomson. The estimation of coherency and group delay. *Biometrika*, 58:469–481, 1971.
- [106] B. E. Hansen. The likelihood ratio test under nonstandard conditions: testing the Markov switching model of GNP. *J. Appl. Econom.*, 7:61–82, 1992.
- [107] W. Härdle. *Applied Nonparametric Regression*. Cambridge University Press, Cambridge, 1989.
- [108] J. Hartung. *Statistik*. Oldenbourg, München, 1989.
- [109] A.C. Harvey. *Forecasting Structural Time Series Models and the Kalman Filter*. Cambridge University Press, Cambridge, 1994.
- [110] J.M. Hausdorff and C.-K. Peng. Multiscaled randomness: a possible source of 1/f noise in biology. *Phys. Rev. E*, 54:2154–2157, 1996.
- [111] R. Hegger, M. Bünner, H. Kantz, and A. Giaquinta. Identifying and modeling delay feedback systems. *Phys. Rev. Lett.*, 81:558–601, 1998.

- [112] R. Hegger, H. Kantz, F. Schmüser, M. Diestelhorst, R.-P. Kapsch, and H. Beige. Dynamical properties of a ferroelectric capacitor observed through nonlinear time series analysis. *Chaos*, 8:727–736, 1998.
- [113] C. Heitz. Optimum time-frequency representations for the classification and detection of signals. *Applied Sig. Process.*, 3:124–143, 1995.
- [114] C. Heitz and J. Timmer. Using optimized time-frequency representations for acoustic quality control of motors. *Acta Acoustica*, 83:1053–1064, 1997.
- [115] B. Hellwig, S. Häußler, B. Schelter, M. Lauk, B. Guschlbauer, J. Timmer, and C.H. Lücking. Tremor-correlated cortical activity in essential tremor. *The Lancet*, 357:519–523, 2001.
- [116] A. Hense. On the possible existence of a strange attractor for the southern oscillation. *Beitr. Phys. Atmosph.*, 60:34–47, 1987.
- [117] C.C. Heyde. *Quasi-likelihood and its Application*. Springer, New York, 1997.
- [118] B. Hille. *Ionic Channels of Excitable Membranes*. Sinauer, Sunderland, Mass., 1992.
- [119] J. Hinde. Choosing between non-nested models: a simulation approach. In L. Fahrmeir et al., editor, *Advances in GLIM and Statistical Modeling*, number 78 in Lecture Notes in Statistics, pages 119–124. Springer, New York, 1992.
- [120] M.J. Hinich. Testing for gaussianity and linearity of stationary time series. *J. Time Series Anal.*, 3:169–176, 1982.
- [121] K.W. Hipel and A.I. McLoed. Preservation of the rescaled adjusted range analysis 2. simulation studies using Box-Jenkins models. *Water Resources Research*, 14:509–516, 1978.
- [122] A.L. Hodgkin. The local electric changes associated with repetitive action in a non-medullated axon. *J. Physiol.*, 107:165–181, 1948.
- [123] A.L. Hodgkin and A.F. Huxley. A quantitative description of ion currents and its application to conduction and excitation in nerve membranes. *J. Physiol.*, 117:500–544, 1952.
- [124] A.E. Hoerl and R.W. Kennard. Ridge regression: Biased estimation for non-orthogonal problems. *Technometrics*, 12:55–67, 1970.
- [125] J. Honerkamp. *Stochastic Dynamical Systems*. VCH, New York, 1993.

- [126] J.J. Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Science*, 79:2554–2558, 1982.
- [127] J.J. Hopfield and D.W. Tank. Computing with neural circuits: a model. *Science*, 233:625–633, 1986.
- [128] W. Horbelt and J. Timmer. Asymptotic scaling laws for precision of parameter estimates in dynamical systems. *Phys. Lett. A*, 310:269–280, 2003.
- [129] W. Horbelt, J. Timmer, M. Bünner, R. Meucci, and M. Ciofini. Identifying physical properties of a CO₂ laser by dynamical modeling of measured time series. *Phys. Rev. E*, 64:016222, 2001.
- [130] W. Horbelt, H.U. Voss, and J. Timmer. Parameter estimation in nonlinear delayed feedback systems from noisy data. *Phys. Lett. A*, 299:513–521, 2002.
- [131] K. Hornick, M. Stinchcombe, and H. White. Multilayer feedforward networks are universal approximators. *Neural Networks*, 2:359–366, 1989.
- [132] L. Hubert and P. Arabie. Comparing partitions. *J. Classifikation*, 2:193–218, 1985.
- [133] H.L. Hurd and N.L. Gerr. Graphical methods for the determining the presence of periodic correlation. *J. Time Ser. Anal.*, 12:337–350, 1991.
- [134] H.E. Hurst. Long-term storage capacity of reservoirs. *Trans. Amer. Soc. Civil Eng.*, 116:770–808, 1951.
- [135] M. Hürzeler. *Statistical Methods for General State-Space Models*. PhD thesis, ETH Zürich, 1998.
- [136] C.E. Hutchinson. The Kalman filter applied to aerospace and electronic systems. *IEEE Trans. Aerospace Elec. Syst.*, 20:500–504, 1984.
- [137] H. Isliker and J. Kurths. A test for stationarity: finding parts in time series apt for correlation dimension estimates. *Int. J. Bif. Chaos*, 3:1573–1579, 1993.
- [138] H. Ito, S.-I. Amari, and K. Kobayashi. Identifiability of hidden Markov information sources and their minimum degree of freedom. *IEEE Trans. Inf. Theory*, 38:324–333, 1992.
- [139] A.K. Jain and R.C. Dubes. *Algorithms for Clustering Data*. Prentice-Hall, Englewood Cliffs, NJ, 1988.

- [140] W. James and C.M. Stein. Estimation with quadratic loss. *Proceedings of the 4.th Berkeley Symposium on Mathematical Statistics and Probability*, 1:197–206, 1961.
- [141] M. Jamshidian and R.I. Jennrich. Conjugate gradient acceleration of the EM-algorithm. *A. Am. Stat. Ass.*, 88:221–228, 1993.
- [142] M. Jamshidian and R.I. Jennrich. Acceleration of the EM-algorithm by using quasi-Newton methods. *J. Roy. Stat. Soc B*, 59:569–587, 1997.
- [143] S. Joeken and J. Schnackenberg. Monte Carlo simulation of shot noise analysis for reconstructing elementary events: quantum bumps in photoreceptor cells. *Z. Naturforsch.*, 48c:519–528, 1992.
- [144] J.B. Johnson. The Schottky effect in low frequency circuits. *Phys. Rev.*, 26:71–85, 1925.
- [145] S. Julier, J. Uhlmann, and H. F. Durrant-Whyte. A new method for the nonlinear transformation of means and covariances in filters and estimators. *IEEE Trans. Automat. Contr.*, 45:477–482, 2000.
- [146] S. J. Julier and J. K. Uhlmann. A general method for approximating nonlinear transformations of probability distributions. Technical report, Department of Engineering Science, University of Oxford, 1996.
- [147] S. Kakutani. On equivalence of infinite product measures. *Ann. Math.*, 49:214–224, 1948.
- [148] R.E. Kalman. A new approach to linear filtering and prediction problems. *Trans. ASME J. Basic Eng. Series D*, 82:35–46, 1960.
- [149] M.J. Kamiński and K.J. Blinowska. A new method of the description of the information flow in the brain structures. *Biological Cybernetics*, 65:203–210, 1991.
- [150] M.J. Kamiński, M. Ding, W.A. Truccolo, and S.L. Bressler. Evaluating causal relations in neural systems: Granger causality, directed transfer function and statistical assessment of significance. *Biol. Cybern.*, 85:145–157, 2001.
- [151] H. Kantz and T. Schreiber. *Nonlinear Time Series Analysis*. Cambridge University Press, Cambridge, 1997.
- [152] H. Kantz and T. Schreiber. Human ECG: nonlinear deterministic versus stochastic aspects. *IEE Proc. Science, Measurement and Technology*, 145:279, 1998.

- [153] R.E. Kass, B.P. Carlin, A. Gelman, and R.M. Neal. Markov Chain Monte Carlo in practice: A roundtable discussion. *The American Statistician*, 52:93–100, 1998.
- [154] B. Kaulakys. Autoregressive model of 1/f noise. *Phys. Lett. A*, 257:37–42, 1999.
- [155] B. Kaulakys and J. Ruseckas. Stochastic nonlinear differential equation generating 1/f noise. *Phys. Rev. E*, 70:020101(R), 2004.
- [156] M.S. Keshner. 1/f noise. *Proc. IEEE*, 70:212–218, 1982.
- [157] P. Kienker. Equivalence of aggregated markov models of ion-channel gating. *Proc. Roy. Soc. Lond. B*, 236:269–309, 1989.
- [158] K.I. Kim and E.J. Powers. A digital method of modeling quadratically nonlinear systems with a general random input. *IEEE Trans. Acoustics, Speech, Sign. Proc.*, 36:1758–1769, 1988.
- [159] K.I. Kim, E.J. Powers, C.P. Ritz, R.W. Miksad, and F.J. Fischer. Modeling of the nonlinear drift oscillations of moored vessels subject to non-Gaussian random sea-wave excitation. *IEEE J. Oceanic Eng.*, 12:568–575, 1987.
- [160] G. Kitagawa. A nonstationary time series model and its fitting by a recursive filter. *J. Time Series Analysis*, 2:103–116, 1981.
- [161] G. Kitagawa and W. Gersch. *Smoothness priors analysis of time series*. Number 116 in Lecture Notes in Statistics. Springer, New York, 1996.
- [162] P.E. Kloeden and E. Platen. *Numerical solution of stochastic differential equations*, volume 23 of *Applications of Mathematics*. Springer, New York, 1992.
- [163] P.E. Kloeden, E. Platen, and H. Schurz. *The numerical solution of SDE through computer experiments*. Springer, New York, 1994.
- [164] A.B. Koehler and E.S. Murphee. A comparison of the Akaike and Schwarz criteria for selcting model order. *Appl. Statist.*, 37:187–195, 1988.
- [165] T. Koh and E.J. Powers. Second-order Volterra filtering and its application to nonlinear system identification. *IEEE Trans. Acoust. Speech Sign. Proc.*, 33:1445–1455, 1985.
- [166] T. Kohonen. *Self-organizing maps*. Springer, Berlin, 1995.

- [167] A.N. Kolmogorov. Über die analytischen Methoden in der Wahrscheinlichkeitsrechnung. *Math. Ann.*, 140:415–458, 1931.
- [168] B. Köster, M. Lauk, J. Timmer, B. Guschlbauer, T. Winter, F.X. Glocker, G. Deuschl, and C.H. Lücking. Central mechanisms in enhanced physiological tremor. *Neuroscience Letters*, 241:135–138, 1998.
- [169] S. Kotz and N.L. Johnson. *Breakthroughs in Statistics 1 & 2*. Springer, New York, 1992.
- [170] A. Lasota and M. Mackay. *Probabilistic properties of deterministic systems*. Cambridge University Press, Cambridge, 1985.
- [171] S. Lauer, J. Timmer, D. von Calker, D. Maier, and J. Honerkamp. Optimal weighted Bayesian design applied to dose-response-curve analysis. *Communications in Statistics - Theory and Methods*, 26:2879–2903, 1997.
- [172] S.L. Lauritzen. *Graphical Modelling*. Oxford University Press, Oxford, 1996.
- [173] T.W. Lee. *Independent Component Analysis*. Kluwer, Dordrecht, 1998.
- [174] E.L. Lehmann. *Theory of Point Estimation*. Wadsworth Inc., New York, 1991.
- [175] R. Linsker. How to generate ordered maps by maximizing the mutual information between input and output signals. *Neural Computation*, 1:402–411, 1989.
- [176] L. Ljung and T. Glad. On global identifiability for arbitrary model parametrizations. *Automatica*, 30:265–276, 1994.
- [177] T.W. Lohmann, H.G. Bock, and J.P. Schlöder. Numerical methods for parameter estimation and optimal experimental design in chemical reaction systems. *Ind. Eng. Chem. Res.*, 31:54–57, 1992.
- [178] A.K. Louis. *Inverse und schlecht gestellte Probleme*. Teubner, Stuttgart, 1989.
- [179] M.C. Mackey and L. Glass. Oscillation and chaos in physiological control systems. *Science*, 197:287–289, 1977.
- [180] C.L. Mallows. Some comments on C_p . *Technometrics*, 15:661–675, 1973.
- [181] C.L. Mallows. More comments on C_p . *Technometrics*, 37:362–372, 1995.

- [182] E. Mammen. *When does bootstrap work? : asymptotic results and simulations*. Number 77 in Lecture notes in statistics. Springer, New York, 1992.
- [183] B.B. Mandelbrot and J.W. van Ness. Fractional brownian motions, fractional noises and applications. *SIAM Review*, 10:422–437, 1968.
- [184] B.B. Mandelbrot and J.R. Wallis. Noah, joseph and operational hydrology. *Water Resources Research*, 4:909–918, 1968.
- [185] D. Maraun, H. Rust, and J. Timmer. Tempting long-memory - on the interpretation of DFA results. *Nonlinear Processes in Geophysics*, 11:495–503, 2004.
- [186] V.Z. Marmarelis, editor. *Advanced methods of physiological system modeling*, volume 1-3, New York, 1988. Plenum Press.
- [187] P. McCullagh and J.A. Nelder. *Generalized linear models*. Chapman and Hall, London, 1995.
- [188] W. McCulloch and W. Pitts. A logical calculus of the ideas immanent in nervous activity. *Bull. Math. Biophys.*, 5:115–133, 1943.
- [189] W. McCulloch and W. Pitts. How we know universals: The perception of auditory and visul forms. *Bull. Math. Biophys.*, 9:127–147, 1947.
- [190] A.I. McLoed and K.W. Hipel. Preservation of the rescaled adjusted range analysis 1. a reassessment of the Hurst phenomenon. *Water Resources Research*, 14:491–508, 1978.
- [191] R.J. Meinhold and N.D. Singpurwalla. Understanding the Kalman filter. *Am. Stat.*, 37:123–127, 1983.
- [192] J.M. Mendel. *Lessons in Estimation Theory for Signal Processing, Communications and Control*. Prentice Hall, New Jersey, 1995.
- [193] X. Meng and D. van Dyk. The EM algorithm – an old folk-song sung to a fast tune. *J. Roy. Stat. Soc B*, 59:511–567, 1997.
- [194] R. Metzler. Comment on "Power-law correlations in the southern-oscillation-index fluctuations characterizing El Niño". *Phys. Rev. E*, 67:018201, 2003.
- [195] S. Michalek, H. Lerche, M. Wagner, N. Mitrovic, M. Schiebe, F. Lehmann-Horn, and J. Timmer. On identification of sodium channel gating schemes using moving-average filtered hidden Markov models. *Euro. Biophys. J.*, 28:605–609, 1999.

- [196] S. Michalek and J. Timmer. Estimating rate constants in hidden Markov models by the EM algorithm. *IEEE Trans. Signal Proc.*, 47:226–228, 1999.
- [197] S. Michalek, M. Wagner, and J. Timmer. A new approximate likelihood estimator for ARMA-filtered hidden Markov models. *IEEE Trans. Signal Proc.*, 48:1537–1547, 2000.
- [198] S. Michalek, M. Wagner, W. Vach, and J. Timmer. Finite sample properties of the maximum likelihood estimator and likelihood ratio tests in hidden Markov models. *Biometrical J.*, 43:863–879, 2001.
- [199] G.W. Milligan and M.C. Cooper. A study of the comparability of external criteria for hierarchical cluster analysis. *Multivariate Behav. Res.*, 21:441–458, 1986.
- [200] M. Minsky and S. Papert. *Perceptrons*. MIT Press, Cambridge, 1969.
- [201] C. Moler and C. van Loan. Nineteen dubious ways to compute the exponential of a matrix. *SIAM Review*, 20:801–836, 1978.
- [202] C. Moler and C. van Loan. Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later. *SIAM Review*, 45:3–49, 2003.
- [203] T.G. Müller, N. Noykova, M. Gyllenberg, and J. Timmer. Parameter identification in a dynamical model of anaerobic waste water treatment processes. *Math. Biosciences*, 177-178:147–160, 2002.
- [204] T.G. Müller and J. Timmer. Parameter estimation in partial differential equations from partially observed noisy data. *Physica D*, 171:1–7, 2002.
- [205] I. Najfeld and T.F. Havel. Derivatives of the matrix exponential and their computation. *Adv. Appl. Math.*, 16:321–375, 1995.
- [206] J. Nakano and S. Tagami. Delay estimation by a Hilbert transform method. *Austral. J. Statist.*, 30:217–227, 1988.
- [207] A.H. Nayfeh and D.T. Mook. *Nonlinear Oscillations*. J. Wiley, New York, 1979.
- [208] D.O. Hebb: A neuropsychological theory. *The Organization of Behaviour*. John Wiley, New York, 1949.
- [209] L. Neyman and E.S. Pearson. On the problem of the most efficient tests of statistical hypotheses. *Phil. Trans. Roy. Soc. A*, 231:289–337, 1933.

- [210] C.L. Nikias and M.R. Raghuveer. Bispectrum estimation: A digital signal processing framework. *Proc. IEEE*, 75:869–891, 1987.
- [211] A.V. Oppenheim and R.W. Schaffer. *Digital signal processing*. Prentice-Hall, Englewood Cliffs, NJ, 1975.
- [212] J.J.K. ÓRuanaidh and W. Fitzgerald. *Numerical Bayesian methods applied to signal processing*. Springer, New York, 1996.
- [213] G. Palm, A. Aertsen, and G.L. Gerstein. On the significance of correlations among neuronal spike trains. *Biol. Cybern.*, 59:1–11, 1988.
- [214] A. Papoulis. *Probability, random variables and stochastic processes*. McGraw – Hill, New York, 1984.
- [215] W. Paul and J. Baschnagel. *Stochastic Processes: From Physics to Finance*. Springer, Heidelberg, 2000.
- [216] M. Peifer, B. Schelter, M. Winterhalder, and J. Timmer. Mixing properties of the Rössler system and consequences for coherence and synchronization analysis. *Phys. Rev. E.*, 72:026213, 2005.
- [217] J. Pelt, J. Brooke, P.J. Pulkkinen, and I. Tuominen. A new interpretation of the solar magnetic cycle. *Astron. & Astrophys.*, 2000.
- [218] D.H. Perkel and G.L. Gerstein and G.P. Moore. Neuronal spike trains and stochastic processes I. The single spike trains. *Biophys. J.*, 7:391–418, 1967.
- [219] D.H. Perkel and G.L. Gerstein and G.P. Moore. Neuronal spike trains and stochastic processes II. Simultaneous spike trains. *Biophys. J.*, 7:419–440, 1967.
- [220] M.S. Phadke and S.M. Wu. Modeling of continuous stochastic processes from discrete observations with application to sunspots data. *J. Am. Stat. Ass.*, 69:325–329, 1974.
- [221] V.F. Pisarenko and D. Sornette. Statistical methods of parameter estimation for deterministically chaotic time series. *Phys. Rev. E*, 69:036122, 2004.
- [222] H. Pohjanpalo. System identifiability based on power series expansion of the solution. *Math. Biosciences*, 41:21–33, 1978.
- [223] D.S.G. Pollock. *A Handbook of Time-Series Analysis, Signal Processing and Dynamics*. Academic Press, San Diego, 1999.

- [224] A.B. Poritz. Hidden Markov models: A guided tour. *Proc. ICASSO*, 1:7–13, 1988.
- [225] K. Pottschmidt, M. König, J. Wilms, and R. Staubert. Analyzing short-term X-ray variability of Cygnus X-1 with linear state space models. *Astronomy Astrophysics*, 334:201–207, 1998.
- [226] W.H. Press, B.P. Flannery, S.A. Saul, and W.T. Vetterling. *Numerical Recipes*. Cambridge University Press, Cambridge, 1992.
- [227] M.B. Priestley. *Spectral Analysis and Time Series*. Academic Press, London, 1989.
- [228] M.B. Priestley and T. Subba Rao. A test for nonstationarity of time series. *J. Roy. Stat. Soc. B*, 31:140–149, 1969.
- [229] L.R. Rabiner. A tutorial on hidden Markov models and selected applications in speech recognition. *Proc. IEEE*, 77:257–285, 1989.
- [230] L.R. Rabiner and B. Gold. *Theory and Application of Digital Signal Processing*. Prentice – Hall, Englewood Cliffs, NJ, 1975.
- [231] W.M. Rand. Objective criteria for the evaluation of clustering methods. *J. Amer. Stat. Soc.*, 66:846–850, 1971.
- [232] T.S. Subba Rao and M.M. Gabr. A test for linearity of stationary time series. *J. Time Series Anal.*, 1:145–158, 1980.
- [233] W.J. Reed and B.D. Hughes. From gene families and genera to incomes and internet file sizes: Why power laws are so common in nature. *Phys. Rev. E*, 66:067103, 2002.
- [234] J. Rice. Bandwidth choice for nonparametric regression. *Annals of Statistics*, 12:1215–1230, 1984.
- [235] O. Richter, P. Nörtersheuser, and W. Pestemer. Non-linear parameter estimation in pesticide degradation. *Science Total Env.*, 123/124:435–450, 1992.
- [236] A.G. Rigas. Spectral analysis of stationary point processes using the fast Fourier transform algorithm. *J. Time. Ser. Anal.*, 13:441–450, 1991.
- [237] D.R. Rigney, A.L. Goldberger, W.C. Ocasio, Y. Ichimaru, G.B. Moody, and R.G. Mark. Multi-channel physiological data: description and analysis. In A.S. Weigend and N. Gerschenfeld, editors, *Time Series Prediction*, volume XV of *SFI Studies in the Sciences of Complexity*, pages 105–129. Addison-Wesley, New York, 1993.

- [238] J. Rissanen. Modeling by shortest data description. *Automatica*, 14:465–471, 1978.
- [239] J. Rissanen. Stochastic complexity. *J. Roy. Stat. Soc B*, 49:223–239 and 252–265, 1987.
- [240] H. Ritter, T. Martinetz, and K. Schulten. *Neuronale Netze*. Addison-Wesley, New York, 1990.
- [241] C.P. Robert, T. Rydén, and D.M. Titterton. Bayesian inference in hidden Markov models by the reversible jump Markov Chain Monte Carlo method. *J. Roy. Stat. Soc. B*, 62:57–75, 2000.
- [242] J.R. Rosenberg, A.M. Amjad, P. Breeze, D.R. Brillinger, and D.M. Halliday. The Fourier approach to the identification of functional coupling between neuronal spike trains. *Prog. Biophys. molec. Biol.*, 53:1–31, 1989.
- [243] F. Rosenblatt. *Principles of Neurodynamics: Perceptrons and the theory of brain mechanisms*. Spartan Books, New York, 1959.
- [244] O.E. Rössler. An equation for continuous chaos. *Phys. Lett. A*, 57:397–381, 1976.
- [245] H. H. Rotermund, G. Haas, R.U. Franz, R.M. Tromp, and G. Ertl. Imaging pattern formation in surface reactions from ultra-high vacuum to atmospheric pressures. *Science*, 270:608–610, 1995.
- [246] L.A. Rubel. A universal differential equation. *Bull. Am. Math. Soc.*, 127:345–349, 1981.
- [247] N.F. Rulkov, A.R. Volkovskii, A. Rodrigues-Lozano, E. Del Rio, and M.G. Velarde. Mutual synchronization of chaotic self-oscillators with dissipative coupling. *Int. J. Bif. Chaos*, 2:669, 1992.
- [248] D. Rumelhart and J. McClelland. *Parallel Distributed Processing*. MIT Press, Cambridge, 1986.
- [249] L. Sachs. *Applied Statistics*. Springer, New York, 1984.
- [250] B. Sakmann and E. Neher. *Single-Channel Recording*. Plenum Press, New York, 1995.
- [251] K. Sameshima and L.A. Baccala. Using partial directed coherence to describe neuronal ensemble interactions. *J. Neuroscience Methods*, 94:93–103, 1999.

- [252] G. Samorodnitsky and M.S. Taqu. *Stable Non-Gaussian Random Processes*. Chapman & Hall, New York, 1994.
- [253] A.-H. Sato. Explanation of power law behavior of autoregressive duration processes based on the random multiplicative process. *Phys. Rev. E*, 69:047101, 2004.
- [254] B. Schelter, M. Winterhalder, M. Eichler, M. Peifer, B. Hellwig, B. Guschlbauer, C.H. Lücking, R. Dahlhaus, and J. Timmer. Testing for directed influences in neuroscience using partial directed coherence. *J. Neuroscience Methods*, 152:210–219, 2006.
- [255] B. Schelter, M. Winterhalder, K. Schwab, L. Leistritz, W. Hesse, R. Bauer, H. Witte, and J. Timmer. Quantification of directed signal transfer within neural networks by partial directed coherence: A novel approach to infer causal time-depending influences in noisy, multivariate time series. *J. Neurosci. Methods*.
- [256] M. Schetzen. *The Volterra and Wiener Theorems of Nonlinear Systems*. R.E. Krieger Pub. Group, Malabar, Florida, 1989.
- [257] M.G. Schimek. *Smoothing and Regression*. Wiley, New York, 2000.
- [258] K. Schittkowski. Parameter estimation in differential equations. In R. Agarwal, editor, *Recent trends in optimization theory and applications*, pages 353–370, Singapore, 1995. World Scientific.
- [259] W. Schottky. Small-shot effect and flicker effect. *Phys. Rev.*, 28:74–103, 1926.
- [260] A. Schuster. On the investigation of hidden periodicities with application to a supposed 26-day period of meteorological phenomenon. *Terr. Mag. Atmos. Elect.*, 3:13–41, 1898.
- [261] G. Schwarz. Estimating the dimension of a model. *Annals Statistics*, 6:461–464, 1978.
- [262] G.A.F. Seber and C.J. Wild. *Nonlinear regression*. Wiley, New York, 1989.
- [263] S. G. Self and K. Y. Liang. Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *J. Am. Stat. Ass.*, 82:605–610, 1987.
- [264] J. Shao. An asymptotic theory for linear model selection. *Statistica Sinica*, 7:221–264, 1997.

- [265] I. Shoji and T. Ozaki. Comparative study of estimation methods for continuous time stochastic processes. *J. Time Series Analysis*, 18:487–506, 1997.
- [266] R.H. Shumway and D.S. Stoffer. An approach to time series smoothing and forecasting using the EM algorithm. *J. Time Ser. Anal.*, 3:253–264, 1982.
- [267] R.H. Shumway and D.S. Stoffer. *Time Series Analysis and Its Application*. Springer, New York, 2000.
- [268] S. Siegert, R. Friedrich, and J. Peinke. Analysis of data sets of stochastic systems. *Phys. Lett. A*, 243:275–280, 1998.
- [269] D. Sigeti. Survival of deterministic dynamics in the presence of noise and the exponential decay of power spectra at high frequency. *Phys. Rev. E*, 52:2443–2457, 1995.
- [270] D. Sigeti and W. Horsthemke. High-frequency power spectra for systems subject to noise. *Phys. Rev. A*, 35:2276–2282, 1987.
- [271] C.M. Stein. Inadmissibility of the usual estimator of the mean of a multivariate normal distribution. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, pages 197–206, Berkeley, 1956. University of California Press.
- [272] M. Stix. *The Sun*. Springer, Berlin, 1989.
- [273] M. Stone. An asymptotic equivalence of choice of model by cross-validation and Akaike’s criterion. *J. Roy. Stat. Soc. B*, 39:44–47, 1977.
- [274] J.C. Strikwerda. *Finite Difference Schemes and Partial Differential Equations*. Wadsworth & Brooks, New York, 1989.
- [275] T. Subba Rao and M.M. Gabr. *An Introduction to bispectral Analysis and bilinear Time Series Models*. Springer, New York, 1984.
- [276] I. Swameye, T. Müller, J. Timmer, O. Sandra, and U. Klingmüller. Identification of nucleocytoplasmic cycling as a remote sensor in cellular signaling by data-based modeling. *Proc. Natl. Acad. Sci.*, 100:1028–1033, 2003.
- [277] F. Takens. Detecting strange attractors in turbulence. In D.A. Rand and L.S. Young, editors, *Dynamical Systems and Turbulence*, volume 898 of *Lecture Notes in Mathematics*, pages 366–381, Berlin, 1981. Springer.
- [278] M.A. Tanner. *Tools for statistical inference : methods for the exploration of posterior distributions and likelihood functions*. Springer, New York, 1996.

- [279] M. S. Taqqu, V. Teverovsky, and W. Willinger. Estimators for long-range dependence: An empirical study. *Fractals*, 3(4):785–798, 1995.
- [280] T. Teräsvirta and I. Mellin. Model selection criteria and model selection tests in regression models. *Scandinavian Journal of Statistics*, 13:159–171, 1986.
- [281] J. Theiler, S. Eubank, A. Longtin, B. Galdrikian, and J.D. Farmer. Testing for nonlinearity in time series: the method of surrogate data. *Physica D*, 58:77–94, 1992.
- [282] D.J. Thomson. Spectrum estimation and harmonic analysis. *Proc. IEEE*, 70:1055–1096, 1982.
- [283] R. Tibshirani. Regression shrinkage and selection via the lasso. *J. Roy. Stat. Soc. B*, 58:267–288, 1996.
- [284] J. Timmer. Modeling noisy time series: Physiological tremor. *Int. J. Bif. Chaos*, 8:1505–1516, 1998.
- [285] J. Timmer. Power of surrogate data testing with respect to nonstationarity. *Phys. Rev. E*, 58:5153–5156, 1998.
- [286] J. Timmer. Estimating parameters in nonlinear stochastic differential equations. *Chaos, Solitons & Fractals*, 11:2571–2578, 2000.
- [287] J. Timmer, S. Häußler, M. Lauk, and C.H. Lücking. Pathological tremors: Deterministic chaos or nonlinear stochastic oscillators ? *Chaos*, 10:278–288, 2000.
- [288] J. Timmer and S. Klein. Testing the Markov condition in ion channel recordings. *Phys. Rev. E*, 55:3306–3310, 1997.
- [289] J. Timmer, M. Lauk, and G. Deuschl. Quantitative analysis of tremor. *Electroenceph. clin. Neurophys.*, 101:461–468, 1996.
- [290] J. Timmer, M. Lauk, and C.H. Lücking. Confidence regions for spectral peak frequencies. *Biometrical J.*, 39:849–861, 1997.
- [291] J. Timmer, M. Lauk, and C.H. Lücking. A test for a difference between spectral peak frequencies. *Comp. Stat. Data Anal.*, 30:45–55, 1999.
- [292] J. Timmer, M. Lauk, W. Pfleger, and G. Deuschl. Cross-spectral analysis of physiological tremor and muscle activity. I. Theory and application to unsynchronized EMG. *Biol. Cybern.*, 78:349–357, 1998.

- [293] J. Timmer, M. Lauk, W. Pflieger, and G. Deuschl. Cross-spectral analysis of physiological tremor and muscle activity. II. Application to synchronized EMG. *Biol. Cybern.*, 78:359–368, 1998.
- [294] J. Timmer, Lauk M., Häußler S., Radt V., Köster B., Hellwig B., Guschlbauer B., Lücking C.H., Eichler M., and Deuschl G. Cross-spectral analysis of tremor time series. *Int. J. Bif. Chaos*, 10:2595–2610, 2000.
- [295] J. Timmer, H. Rust, W. Horbelt, and H.U. Voss. Parametric, nonparametric and parametric modelling of a chaotic circuit time series. *Phys. Lett. A*, 274:123–134, 2000.
- [296] J. Timmer, U. Schwarz, H.U. Voss, I. Wardinski, T. Belloni, G. Hasinger, M. van der Klis, and J. Kurths. Linear and non-linear time series analysis of the black hole candidate Cygnus X-1. *Phys. Rev. E*, 61:1342–1352, 2000.
- [297] J. Timmer and A. Weigend. Modeling volatility using state space models. *Int. J. Neur. Syst.*, 8:385–398, 1997.
- [298] A. Timmermann, H.U. Voss, and RA Pasmanter. Empirical dynamical system modeling of enso using nonlinear inverse techniques. *Journal of Physical Oceanography*, 2000.
- [299] D. Tjostheim and B.H. Auestad. Nonparametric identification of nonlinear time series: Projections. *J. Am. Stat. Ass.*, 89:1398–1409, 1994.
- [300] D. Tjostheim and B.H. Auestad. Nonparametric identification of nonlinear time series: Selecting significant lags. *J. Am. Stat. Ass.*, 89:1410–1419, 1994.
- [301] K.J. Utikal. A new method for detecting neural interconnectivity. *Biol. Cybern.*, 76:459–470, 1997.
- [302] S. Vajda, K.R. Godfrey, and H. Rabitz. Similarity transformation approach to identifiability analysis of nonlinear compartmental models. *Mathematical Biosciences*, 93:217–248, 1989.
- [303] S. Vajda and H. Rabitz. State Isomorphism Approach to Global Identifiability of Nonlinear Systems. *IEEE Trans. Auto. Control*, 34(2):220–223, 1989.
- [304] G.K. Vallis. El nino: A chaotic dynamical system? *Science*, 232:243–245, 1986.
- [305] C. van der Malsburg and E. Singer. A neural cocktail-party processor. *Biol. Cybern.*, 54:29–40, 1986.

- [306] R. van der Merwe, N. de Freitas, A. Doucet, and E. Wan. The unscented particle filter. Technical Report CUED/F-INFENG/TR 380, Cambridge University Engineering Department, Cambridge, England, 2000.
- [307] B. van der Pol. On oscillation-hysteresis in a simple triode generator. *Phil. Mag.*, 43:700–719, 1922.
- [308] N.G. van Kampen. *Stochastic Processes in Physics and Chemistry*. Springer, Berlin, 1997.
- [309] R. von Sachs. Detecting periodic components in a stationary time series by an improved nonparametric procedure. In O. Lessi, editor, *Applications of Time Series Analysis in Astronomy and Meteorology*, pages 115–118, Padova, 1993. SM Legatori.
- [310] H. von Storch and F.W. Zwiers. *Statistical Analysis in Climate Research*. Cambridge University Press, Cambridge, 1999.
- [311] H. Voss and J. Kurths. Reconstruction of nonlinear time delay models from data by the use of optimal transformations. *Phys. Lett. A*, 234:336–344, 1997.
- [312] Q. H. Vuong. Likelihood ratio tests for modelselection and non-nested hypotheses. *Econometrica*, 57(2):307–333, 1989.
- [313] M. Wagner, S. Michalek, and J. Timmer. Estimating rate constants in hidden Markov models with loops and with nearly equal dwell times. *Proc. Roy. Soc. B*, 266:1919–1926, 1999.
- [314] M. Wagner, S. Michalek, and J. Timmer. Testing for the number of states in hidden Markov models with application to ion channel data. In W. Gaul and H. Locarek-Junge, editors, *Studies in Classification, Data Analysis and Knowledge Organization*, pages 260–267, Heidelberg, 1999. Springer.
- [315] M. Wagner and J. Timmer. The effects of non-identifiability on testing for detailed balance in aggregated Markov models of ion-channel gating. *Biophys. J.*, 79:2918–2924, 2000.
- [316] M. Wagner and J. Timmer. Model selection in non-nested Markov models for ion channel gating. *J. Theo. Biol.*, 208:439–450, 2001.
- [317] A.M. Walker. Some results concerning the asymptotic distribution of sample Fourier transforms and periodograms for a discrete-time stationary process with a continuous spectrum. *J. Time Ser. Anal.*, 21:95–109, 2000.

- [318] C.S. Wallace and P.R. Freeman. Estimation and inference by compact coding. *J. Roy. Stat. Soc. B*, 49:240–265, 1987.
- [319] E. A. Wan, R. van der Merwe, and A. T. Nelson. Dual estimation and the unscented transformation. In S. A. Solla, T. K. Leen, and K.-R. Muller, editors, *Advances in Neural Information Processing Systems 12*, pages 666–672. MIT Press, Cambridge, 2000.
- [320] G. Weiss. Time-reversibility of linear stochastic processes. *J. Appl. Prob.*, 12:831–836, 1975.
- [321] M.B. Weissman. 1/f noise and other slow, non-exponential kinetics in condensed matter. *Rev. Mod. Phys.*, 60:537–572, 1988.
- [322] P. Werbos. *Beyond regression: New tools for prediction and analysis in the behavioral sciences*. PhD thesis, Harvard University, 1974.
- [323] Westfall and Young. *Resampling based Multiple Testing: Examples and Methods for p-value Adjustment*. Wiley, New York, 1993.
- [324] J. Whittaker. *Graphical Models in Applied Multivariate Statistics*. John Wiley, Chichester, 1995.
- [325] P. Whittle. Estimation and information in stationary time series. *Ark. Mat.*, 2:423–434, 1953.
- [326] M. Winterhalder, B. Schelter, W. Hesse, K. Schwab, L. Leistritz, D. Klan, R. Bauer, J. Timmer, and H. Witte. Comparison of time series analysis techniques to detect direct and time-varying interrelations in multivariate, neural systems. *Signal Processing Journal*, 85:2137–2160, 2005.
- [327] A. Witt, J. Kurths, and A. Pikovsky. Testing stationarity in time series. *Phys. Rev. E*, 58:1800–1810, 1998.
- [328] C.F. Wu. On the convergence properties of the EM algorithm. *Ann. Stat.*, 11:95–103, 1983.
- [329] W.T. Wu. On zeros of algebraic equations - an application of Ritt principle. *Kexue Tongbao*, 31:1–5, 1986.
- [330] X. Yang, S. Du, and J. Ma. Do earthquakes exhibit self-organized criticality? *Phys. Rev. Lett.*, 92:228501, 2004.
- [331] K.Y. Yeung and W.L. Ruzzo. Principal component analysis for clustering gene expression data. *Bioinformatics*, 17:763–774, 2001.

- [332] G.U. Yule. On a method of investigating periodicities in disturbed series, with special reference to wolfer's sunspot numbers. *Phil. Trans. R. Soc. A*, 226:267–298, 1927.
- [333] M.A. Zarrop. *Optimal Experimental Design for Dynamic System Identification*. Number 21 in Lecture Notes in Control and Information Sciences. Springer, berlin, 1979.