

Skript zur Vorlesung

STOCHASTISCHE PROZESSE

Ludwig Fahrmeir, Günter Raßer, Thomas Kneib

Dieses Skript beruht zu einem großen Teil auf Auszügen aus Fahrmeir, Kaufmann & Ost (1981). Ergänzungen betreffen Teile der Kapitel 2 und 3, sowie die Kapitel 6 und 7.

Unser Dank gilt Rudi Eichholz, der den größten Teil des \LaTeX -Skripts erstellte.

Inhaltsverzeichnis

Literatur	5
1 Einführung und Beispiele	6
1.1 Einführende Beispiele und erste Definition	6
1.2 Spezielle stochastische Prozesse	7
1.2.1 Irrfahrten (Random walks)	7
1.2.2 Wiener-Prozess	12
1.2.3 Zählprozesse und Poisson-Prozess	17
2 Grundbegriffe der allgemeinen Theorie	27
2.1 Definitionen stochastischer Prozesse	27
2.1.1 Klassische Definition: SP als Familie von Zufallsvariablen	27
2.1.2 Stochastischer Prozess als Produktabbildung	30
2.1.3 Stochastischer Prozess als Abbildung in Funktionenraum	31
2.2 Existenzsatz von Kolmogorov	31
2.3 Äquivalenz- und Stetigkeitsbegriffe	34
2.3.1 Äquivalente stochastische Prozesse	34
2.3.2 Stetigkeitsbegriffe	36
2.4 Stationäre und nichtstationäre stochastische Prozesse	39
3 Markov-Ketten	44
3.1 Grundlegende Eigenschaften, Beispiele	44
3.2 Klassifizierung von Zuständen und Rückkehrverhalten	53
3.3 Das Grenzverhalten von homogenen MK	62
3.4 Instationäre und inhomogene MK	66

3.4.1	Instationäre und inhomogene binäre MK 1. Ordnung	67
3.4.2	Weitere inhomogene MK	68
3.5	Statistische Inferenz für MK	69
3.5.1	Likelihood-Inferenz für SP	69
3.5.2	Inferenz bei homogenen Markov-Ketten	71
3.5.3	Fallstudie: Niederschlag bei den Snoqualmie-Wasserfällen	73
3.6	Hidden Markov Modelle	75
3.7	Allgemeine Markov-Ketten und MCMC-Methoden	77
3.7.1	Allgemeine Markovketten	78
3.7.2	Metropolis-Hastings-Algorithmus und MCMC	81
3.8	Der Metropolis-Hastings-Algorithmus	81
4	Diskrete Markov-Prozesse	85
4.1	Definition und elementare Eigenschaften	85
4.2	Geburts- und Todesprozesse	89
4.2.1	Geburtsprozesse	89
4.2.2	Geburts- und Todesprozesse	92
4.3	Diskrete MP: Skizze der allgemeinen Theorie	95
4.4	Zur statistischen Analyse	107
5	Erneuerungs- und Semi-Markov-Prozesse	112
5.1	Erneuerungsprozesse	112
5.1.1	Definition und grundlegende Begriffe	113
5.1.2	Zur Theorie der Erneuerungsprozesse	116
5.2	Semi-Markov-Prozesse	118
5.2.1	Definition und grundlegende Eigenschaften	118
5.2.2	Grenzwertsätze	122
5.3	Bemerkung zur statistischen Inferenz	123
6	Martingale	124
6.1	Martingale in diskreter Zeit	124
6.1.1	Definition und Beispiele	124

6.1.2	Spielsysteme und das Optional Stopping Theorem	128
6.1.3	Doob-Meyer-Zerlegung in diskreter Zeit	132
6.2	Martingale in stetiger Zeit	133
6.2.1	Definition und Beispiele	133
6.2.2	Doob-Meyer-Zerlegung in stetiger Zeit	134
7	Punkt- und Zählprozesse	137
7.1	Definition und einige Eigenschaften	139
7.2	Spezielle Zählprozesse und Beispiele	144
7.2.1	MP, SMP und EP als spezielle Zählprozesse	144
7.2.2	Lebensdauern und Survivalanalyse	149
7.3	Ein allgemeines Zählprozess-Modell	153
7.4	Statistische Inferenz für Survival- und Ereignisdaten	155
7.4.1	Nelson-Aalen-Schätzer	155
7.4.2	Parametrische und semiparametrische Likelihood-basierte Inferenz	157
8	MP mit stetigem Zustands- und Parameterraum	159
8.1	Modellierung von Aktienpreisen	159
8.1.1	Bonds (risikofreie Anlagen)	159
8.1.2	Die geometrische Brownsche Bewegung als Aktienkursmodell	160
8.2	Markov-Prozesse mit kontinuierlichem Zustandsraum	161
8.3	Diffusionsprozesse und stochastische Differentialgleichungen	163

Literatur

- AMMANN, M. (2001). *Credit Risk Valuation* (2nd ed.). Springer.
- ANDERSEN, P.K., BORGAN, O. (1985). Counting process models for life history data: A review (with discussion). *Scand. J. Statist.* **12**, 97–158.
- ANDERSEN, P.K., BORGAN, O., GILL, R., KEIDING, N. (1993). *Statistical Models Based on Counting Processes*. Springer.
- BAUER, H. (2001). *Wahrscheinlichkeitstheorie* (5. Aufl.). De Gruyter.
- BHATTACHARYA, R., WAYMIRE, E. (1990). *Stochastic Processes with applications*. Wiley.
- BILLINGSLEY, P. (1995). *Probability and measure* (3rd ed.). Wiley.
- BINGHAM, N.H., KIESEL, R. (2004). *Risk-Neutral Valuation* (2nd ed.). Springer.
- BRÉMAUD, P. (1981). *Point Processes and Queues: Martingale Dynamics*. Springer.
- FAHRMEIR, L., KAUFMANN, H., OST, F. (1981). *Stochastische Prozesse*. Hanser.
- FLEMING, T.R., HARRINGTON, D.P. (1991). *Counting Processes and Survival Analysis*. Wiley.
- FRANKE, J., HÄRDLE, W., HAFNER, C. (2001). *Einführung in die Statistik der Finanzmärkte*. Springer.
- GÄNSSLER, P., STUTE, W. (1977). *Wahrscheinlichkeitstheorie*. Springer.
- GUTTORP, P. (1995). *Stochastic Modelling of Scientific Data*. Chapman & Hall.
- GRIMMETT, G., STIRZAKER, D. (2001). *Probability and Random Processes*. Oxford University Press.
- JACOD, J. (1979). *Calcul stochastique et problèmes de martingales*. Lectures Notes in Mathematics 714. Springer.
- KARLIN, S., TAYLOR, H.M. (1975). *A first course in stochastic processes* (2nd ed.). Academic Press.
- KOLLER, M. (2000). *Stochastische Modelle in der Lebensversicherung*. Springer.
- KORN, R., KORN, E. (1999). *Optionsbewertung und Portfolio-Optimierung*. Vieweg.
- LAWLER, G.F. (2006). *Introduction to Stochastic Processes* (2nd ed.). Chapman & Hall.
- LIPTSER, R.S., SHIRYAYEV, A.N. (1989). *Theory of Martingales*. Kluwer.
- ØKSENDAL, B. (2000). *Stochastic Differential Equations* (5th ed.). Springer.
- PRUSCHA, H. (2000). *Statistik stochastischer Prozesse*. Vorlesung, LMU.
- ROSS, S.M. (2007). *Introduction to Probability Models* (9th ed.). Academic Press.
- TAYLOR, H.E., KARLIN, S. (1998). *An Introduction to Stochastic Modelling* (2nd ed.). Academic Press.
- TODOROVIC, P. (1992). *Introduction to Stochastic Processes and their applications*. Springer.

Kapitel 1

Einführung und Beispiele

Inhalt:

- Anwendungsbeispiele
- erste Definition eines stochastischen Prozesses
- einige spezielle stochastische Prozesse

Ziel:

- Aufzeigen der Vielfalt stochastischer Prozesse
- Einführung erster Grundbegriffe anhand von Beispielen

1.1 Einführende Beispiele und erste Definition

Beispiele:

- rein zeitlich:
 - Mikrodaten des IFO-Konjunkturtests
 - Markenwahl
 - Erwerbstätigkeit: Individuelle Verläufe, Anzahl Erwerbsloser
 - Krankheiten : Individuelle Verläufe, Anzahl Erkrankter
 - DNA-Sequenzen
 - Genetische Evolution
 - (Aktien)Kurse: Tagesdaten

- (Aktien)Kurse: Inter-Tagesdaten
- Schlafzustände
- Schadensfälle und Schadenshöhen bei KFZ-Versicherung
- Brown'sche Bewegung (ein- und mehrdimensional)
- räumlich + räumlich-zeitlich:
 - Krebsatlas (räumlich, räumlich-zeitlich)
 - Erwerbstätigkeit (räumlich, räumlich-zeitlich)
 - fMRI Daten (human brain mapping)
 - Kosten bei privater Krankenversicherung (räumlich-zeitlich)

In allen Beispielen beobachtet bzw. misst man Realisierungen von ZV $X_t, t \in T$, mit einer geeigneten Indexmenge T , z.B. mit

$$\begin{aligned}
 T &\subset \mathbb{N}_0, \mathbb{R}_+ && \text{(Zeit),} \\
 T &\subset \mathbb{Z}^2 && \text{(räumliches Gitter),} \\
 T &\subset \mathbb{R}^2 && \text{(kontinuierliche Teilmenge),} \\
 T &\subset \mathbb{Z}^2 \times \mathbb{N}_0 && \text{(Raum-Zeit).}
 \end{aligned}$$

Dies führt zur ersten

Definition 1.1 Stochastischer Prozess (SP) als Familie von ZV

Sei T Indexmenge bzw. *Parameterraum*. Die Familie $\{X_t, t \in T\}$ von ZVen heißt stochastischer Prozess. Der Wertebereich S der ZVen heißt *Zustandsraum*.

Bemerkung:

Genauer gilt $X_t : (\Omega, \mathcal{F}, P) \rightarrow (S, \mathcal{S})$ ZV für alle t , \mathcal{S} ist σ -Algebra und (Ω, \mathcal{F}, P) ein W-Raum. Dabei ist in der Regel $S \subset \mathbb{Z}, \mathbb{R}, \mathbb{R}^2, \dots$ und für die σ -Algebra \mathcal{S} gilt $\mathcal{S} = \mathcal{P}(S)$ für S diskret bzw. $\mathcal{S} = \mathcal{B}$ Borelmengen für $S \in \mathbb{R}_+, \mathbb{R}, \dots$

Dann heißt das Quadrupel $X = \{\Omega, \mathcal{F}, P, (X_t, t \in T)\}$ stochastischer Prozess.

1.2 Spezielle stochastische Prozesse

1.2.1 Irrfahrten (Random walks)

Irrfahrten („Random Walks“): einfache stochastische Modelle für Spielsituationen; diskretisierte Idealisierung von Kursverläufen bzw. der Brown'schen Bewegung.

Diskrete einfache Irrfahrt auf der Geraden

Start in 0 (Zeit $t = 0$)

Bewegung: ($t = 1, 2, \dots$)

Ein Schritt nach rechts mit Wahrscheinlichkeit p

Ein Schritt nach links mit Wahrscheinlichkeit q

Verbleiben im Zustand mit Wahrscheinlichkeit r

$$p + q + r = 1$$

$\{Z_t, t = 1, 2, \dots\}$ i.i.d. Folge

$Z_t \in \{-1, 0, 1\}$ ZV für t -ten Schritt

$$P(Z_t = -1) = q, P(Z_t = 0) = r, P(Z_t = 1) = p$$

$X_t, t = 0, 1, 2, \dots$ Position nach dem t -ten Schritt

$$X_0 = 0, X_t = Z_1 + Z_2 + \dots + Z_t, \text{ bzw. } X_t = X_{t-1} + Z_t, t \geq 1$$

Definition 1.2 Einfache Irrfahrt

Die Folge $X = \{X_t, t \in \mathbb{N}_0\}$, mit $X_t = X_{t-1} + Z_t$, heißt (einfache) Irrfahrt auf der Geraden.

(Z_t i.i.d. mit $P(Z_t = 1) = p, P(Z_t = -1) = q, P(Z_t = 0) = r$)

Also: $\{X_t, t \in \mathbb{N}_0\}$ ist ein stochastischer Prozess mit $T = \mathbb{N}_0$, $S = \mathbb{Z}$.

Zugrundeliegender Ergebnisraum Ω , Ergebnisse ω :

ω Folge der Realisierungen von Z_1, Z_2, \dots

z.B. $\omega = (1, 1, -1, 1, 1, 0, 0, 1, -1, \dots) = (Z_1(\omega) = 1, Z_2(\omega) = 1, Z_3(\omega) = -1, \dots)$

$\Omega = \{-1, 0, 1\}^{\mathbb{N}}$

Pfad, Trajektorie, Realisierung: Treppenfunktion bzw. Polygon.

- Spezialfälle:

$$\begin{aligned} r &= 0 && \text{kein Unentschieden} \\ p = q &= \frac{1}{2} && \text{faïres Spiel, symmetrische Irrfahrt} \end{aligned}$$

- Modifikation: Absorbierende Schranken

Interpretation: Spieler A mit Anfangskapital a

Spieler B mit Anfangskapital b

X_t Gewinn von A nach t Spielen

$\{X_t = -a\}$ Ruin von A („Gambler's ruin“)

$\{X_t = b\}$ Ruin von B .

Ziel: z.B. Berechnung von Ruinwahrscheinlichkeiten bzw. Gewinnwahrscheinlichkeiten.

Ohne Beweis:

Gewinnwahrscheinlichkeit bei $p = q = \frac{1}{2}$ ist

$$\text{für } A: P_A = \frac{a}{a+b} \quad \text{und für } B: P_B = \frac{b}{a+b}$$

- Markov-Eigenschaft der diskreten Irrfahrt:

$$P(X_{t+1} = j \mid X_t = i, X_{t-1} = i_{t-1}, X_{t-2} = i_{t-2}, \dots, X_1 = i_1) = P(X_{t+1} = j \mid X_t = i),$$

für $j \in \{i+1, i, i-1\}$

Interpretation: Die Zukunft X_{t+1} ist bei bekannter Gegenwart X_t (bedingt) unabhängig von der Vergangenheit X_{t-1}, \dots, X_1 ($X_0 = 0$!) bzw. X_{t-1}, \dots, X_1, X_0 ; X_0 unabhängig von $\{Z_t\}$.

Beweis:

$$\begin{aligned}
 P(X_{t+1} = j \mid X_t = i, X_{t-1} = i_{t-1}, \dots, X_1 = i_1, X_0 = i_0) \\
 &= P(Z_{t+1} = j - i \mid Z_t = i - i_{t-1}, \dots, Z_1 = i_1 - i_0, X_0 = i_0) \\
 &\stackrel{(Z_{t+1}, Z_t, \dots, Z_1, X_0 \text{ unabhängig})}{=} P(Z_{t+1} = j - i) = P(X_{t+1} = j \mid X_t = i).
 \end{aligned}$$

□

Ungerichtete Form der Markov-Eigenschaft:

$$P(X_t = i \mid X_{s \neq t}) = P(X_t = i \mid X_{t+1} = i_{t+1}, X_{t-1} = i_{t-1})$$

Interpretation: Bei gegebenen Nachbarn X_{t+1}, X_{t-1} ist X_t von weiteren Nachbarn links und rechts unabhängig.

Beweis:

□

- Zwei-dimensionale symmetrische Irrfahrt

$$\begin{aligned}
 X_n &= (X_{1n}, X_{2n}) \in \mathbb{Z}^2 \\
 X_{n+1} &= X_n + Z_n ; \\
 Z_n &\in \left(\left(\begin{array}{c} 1 \\ 0 \end{array} \right), \left(\begin{array}{c} -1 \\ 0 \end{array} \right), \left(\begin{array}{c} 0 \\ 1 \end{array} \right), \left(\begin{array}{c} 0 \\ -1 \end{array} \right) \right)
 \end{aligned}$$

mit

$$P\left(Z_n = \left(\begin{array}{c} 1 \\ 0 \end{array}\right)\right) = \dots = P\left(Z_n = \left(\begin{array}{c} 0 \\ -1 \end{array}\right)\right) = \frac{1}{4}$$

Verallgemeinerungen

- Allgemeine Irrfahrt (Random Walk)

Bisher: $\{Z_t, t \in \mathbb{N}\}$ iid Folge mit $Z_t \in \{-1, 0, 1\}$.

Allgemein: $\{Z_t, t \in \mathbb{N}\}$ iid Folge, Z_t beliebig verteilt.

Beispiel: $\{Z_t, t \in \mathbb{N}\}$ iid $N(0, \sigma^2)$, „Gauß'sches Weißes Rauschen“,

$$X_t = X_{t-1} + Z_t \text{ Gauß-Irrfahrt.}$$

Markov-Eigenschaft gilt analog:

$$f(x_t | X_{t-1} = x_{t-1}, \dots, X_1 = x_1) = f(x_t | X_{t-1} = x_{t-1}) \sim N(x_{t-1}, \sigma^2)$$

bzw. $X_t | X_{t-1}, \dots, X_1 \sim X_t | X_{t-1} \sim N(X_{t-1}, \sigma^2)$

- Autoregressive Prozesse

Autoregressiver Prozess der Ordnung 1 ($AR(1)$)

$$X_t = \rho X_{t-1} + Z_t, \quad Z_t \text{ iid } N(0, \sigma^2)$$

Autoregressiver Prozess der Ordnung p ($AR(p)$)

$$X_t = \rho_1 X_{t-1} + \dots + \rho_p X_{t-p} + Z_t$$

Gauß-Prozess, falls Z_t iid $N(0, \sigma^2)$.

Markov-Eigenschaft

$$X_t | X_{t-1}, \dots, X_1 \sim X_t | X_{t-1} \sim N(\rho X_{t-1}, \sigma^2) \quad \text{für } AR(1)$$

$$X_t | X_{t-1}, \dots, X_1 \sim X_t | X_{t-1}, \dots, X_{t-p} \sim N(\mu_t, \sigma^2) \quad \text{für } AR(p)$$

mit $\mu_t = \rho_1 X_{t-1} + \dots + \rho_p X_{t-p}$.

Autoregressive Prozesse sind wichtiger Baustein in der Zeitreihenanalyse.

- Räumlicher Markov-Prozess

Baustein für Bildanalyse, geographische Epidemiologie, etc.

$\{X_s, s = (i, j) \in \mathbb{Z}^2\}$ heißt räumlicher (Gitter-)Prozess oder Zufallsfeld.

Beispiele:

- $X_{ij} \in \{0, 1\}$ Indikatorvariablen, z.B.

$$X_{ij} = \begin{cases} 1 & \text{archäologischer Fund in } (i, j) \\ 0 & \text{sonst} \end{cases}$$

- $X_s \in \{0, 1, 2, \dots\}$ Zählvariable, z.B. Anzahl von Krebserkrankungen in bestimmter Periode in Landkreisen; dabei ist das Gitter irregulär.
- X_{ij} stetig verteilt, z.B. Graustufe/Farbe in der Bildanalyse
Räumliche Markov-Eigenschaft

$$f(X_{ij} | X_{kl}, (k, l) \neq (i, j)) = f(X_{ij} | X_{kl}, (k, l) \sim (i, j))$$

\sim : „Nachbar von“

1.2.2 Wiener-Prozess

Historisch: Der Wiener-Prozess ist stochastisches Modell für die Brown'sche Bewegung, d.h. die (zeits-tetige) Irrfahrt kleiner Teilchen in homogener ruhender Flüssigkeit. Irrfahrt kommt durch zufälliges Zusammenstoßen mit Molekülen der Flüssigkeit zustande.

Moderne Herleitung: N. Wiener, 1923.

Parameterraum $T = \mathbb{R}_+$, Zustandsraum $S = \mathbb{R}$ (bzw. $\mathbb{R}^2, \mathbb{R}^3$).

Bezeichnung: $W = \{W_t, t \in \mathbb{R}_+\}$ oder $\{W(t), t \in \mathbb{R}_+\}$,

($W(t)$ um die „Funktion“ von t zu betonen)

Der Wiener-Prozess ist wichtiger Baustein zur Modellierung von Wertpapierpreisen mit Anwendung in der Optionsbewertung (Black-Scholes-Regel), vgl. z.B. Korn/Korn (1999):

Sei $P_0(0), P_0(t)$ der Preis eines risikolosen Wertpapiers (z.B. Sparguthaben) zu den Zeitpunkten 0 und t . Bei stetiger Verzinsung mit konstantem Zinssatz r gilt

$$\begin{aligned} P_0(t) &= P_0(0)e^{rt} && \text{bzw.} \\ \ln P_0(t) &= \ln P_0(0) + rt \end{aligned}$$

\implies loglinearer Ansatz für Aktienkurse:

$$\ln P(t) = \ln P(0) + rt + W(t),$$

$W(t)$: regelloser Fehler in stetiger Zeit $\sim N(0, \sigma^2 t)$.

Der Wiener-Prozess als Grenzfall von Irrfahrten

$X(t)$ symmetrische diskrete Irrfahrt mit Bewegungen zu den Zeitpunkten $n\Delta t$, $n = 1, 2, \dots$ um $\pm\Delta x$ nach oben bzw. unten, jeweils mit Wahrscheinlichkeit $\frac{1}{2}$.

$X(t)$ = Lage des Teilchens für $t = n\Delta t$, $X(0) = 0$ Start.

$$\implies X(t) = \sum_{k=1}^n Z_k, \quad Z_k \text{ iid mit } \begin{cases} P(Z_k = +\Delta x) = \frac{1}{2} \\ P(Z_k = -\Delta x) = \frac{1}{2} \end{cases}$$

$$E(Z_k) = 0$$

$$\text{Var}(Z_k) = (\Delta x)^2$$

$$\implies E(X(t)) = 0, \quad \text{Var}(X(t)) = (\Delta x)^2 n = \underbrace{\frac{(\Delta x)^2}{\Delta t}}_{=: \sigma^2} t$$

Grenzübergang $n \rightarrow \infty$, so dass

$$\sigma^2 := \frac{(\Delta x)^2}{\Delta t}, \quad 0 < \sigma^2 < \infty$$

konstant bleibt: Zentraler Grenzwertsatz

$$\implies X(t) \sim N(0, \sigma^2 t) \quad \text{für } n \rightarrow \infty.$$

Weiter gilt:

Die *Zuwächse* $X(t) - X(s)$, $X(v) - X(u)$, mit $u < v < s < t$ sind *unabhängig*, da sie sich aus getrennten iid Teilsommen der $\{Z_n\}$ -Folge zusammensetzen.

Die *Zuwächse* $X(t+s) - X(s)$ sind *stationär*, d.h. die Verteilung hängt nur von der Zeitdifferenz ab

$$X(t+s) - X(s) \sim X(t) - X(0) \sim N(0, \sigma^2 t)$$

Plausibel: Unabhängigkeit und Stationarität der Zuwächse überträgt sich auf Grenzprozess für $n \rightarrow \infty$. Exakter Beweis schwierig.

Deshalb: Axiomatischer Zugang, obige Eigenschaften + *Stetigkeit der Pfade* (aus „physikalischen“ Gründen) werden *postuliert*.

Die „Herleitung“ des Wiener-Prozesses als Grenzfall von Irrfahrten funktioniert auch für allgemeine „symmetrische“ Irrfahrten, z.B. mit

$$Z_k \sim N(0, \sigma^2 \Delta t), \quad t = n\Delta t$$

$$\text{Var}(X(t)) = n\sigma^2 \Delta t = \frac{t\sigma^2 \cancel{\Delta t}}{\cancel{\Delta t}} = \sigma^2 t$$

$$\implies X(t) \sim N(0, \sigma^2 t) \quad \text{für } n \rightarrow \infty$$

Axiomatische Definition und Eigenschaften des Wiener-Prozesses**Definition 1.3 Wiener-Prozess W**

Ein stochastischer Prozess $W = \{W(t), t \in \mathbb{R}_+\}$, $S = \mathbb{R}$, heißt Wiener-Prozess, wenn gilt:

(W1) *Zuwächse normalverteilt und stationär:*

$$W(s+t) - W(s) \sim N(0, \sigma^2 t) \quad \text{für alle } s, t \geq 0$$

(W2) Für alle $0 \leq t_1 < t_2 < \dots < t_n$, $n \geq 3$ sind die *Zuwächse*

$$W(t_2) - W(t_1), \dots, W(t_n) - W(t_{n-1})$$

unabhängig

(W3) $W(0) = 0$

(W4) Pfade sind stetig

Für $\sigma^2 = 1$ heißt der Wiener-Prozess normiert.

Bemerkungen:

- (a) (W1), (W2) \Rightarrow Wiener-Prozess ist Prozess mit stationären, unabhängigen und normalverteilten Zuwächsen.
- (b) (W3) ist Anfangsbedingung. Addition von c liefert $\tilde{W}(t) = W(t) + c$ mit $\tilde{W}(0) = c$.
- (c) (W1), (W2), (W3) bestimmen „endlich-dimensionale Verteilung“ von $W(t_1), W(t_2), \dots, W(t_n) \quad \forall n \geq 1, t_1, \dots, t_n \in \mathbb{R}_+$.
(W4) folgt *nicht* aus (W1), (W2), (W3). Im Gegenteil: Man müsste zeigen, dass (W4) mit (W1) – (W3) verträglich ist.

Eigenschaften des Wiener-Prozesses

(a) **Verteilungseigenschaften**

eindimensionale Verteilung:

$$W(t) \sim N(0, \sigma^2 t) \quad \forall t \in \mathbb{R}_+,$$

da $W(t) - \underbrace{W(0)}_{=0} \sim N(0, \sigma^2 t)$ nach (W1).

zweidimensionale Verteilungen:

$$\begin{pmatrix} W(s) \\ W(t) \end{pmatrix} \sim N(0, \Sigma), \quad 0 \leq s < t$$

mit

$$\Sigma = \sigma^2 \begin{pmatrix} s & s \\ s & t \end{pmatrix},$$

also $\text{Cov}(W(s), W(t)) = \sigma^2 s$. Für s, t beliebig: $\text{Cov}(W(s), W(t)) = \min(t, s)\sigma^2$.

Beweis:

$\underbrace{W(s)}_{=:U}, \underbrace{W(t) - W(s)}_{=:V}$, $0 \leq s < t$ o.B.d.A unabhängig und normalverteilt. (W1, W2)

$\implies W(s) = U, W(t) = U + V$ bivariat normalverteilt

$$\begin{aligned} \text{Cov}(W(s), W(t)) &= E(W(s) \cdot W(t)) \\ &= E[\underbrace{(W(t) - W(s))}_V \underbrace{W(s)}_U + \underbrace{(W(s))^2}_U] \\ &= E[\underbrace{(W(t) - W(s))}_V \underbrace{(W(s) - W(0))}_U] + E(\underbrace{W(s)}_U)^2 \\ &\stackrel{\text{unabhängige Zuwächse}}{=} \underbrace{E[(W(t) - W(s))]}_{E(V)=0} \underbrace{E[(W(s) - W(0))]}_{E(U)=0} + \underbrace{\text{Var}(W(s))}_{\sigma^2 s} \end{aligned}$$

□

Die bivariate Normalverteilungsdichte lässt sich schreiben als:

$$f_{s,t}(x_1, x_2) = \frac{1}{2\pi\sigma^2\sqrt{s(t-s)}} \exp\left\{-\frac{1}{2\sigma^2}\left[\frac{x_1^2}{s} + \frac{(x_2 - x_1)^2}{t-s}\right]\right\}, \quad s < t$$

Beweis: Übung

Bedingte Dichten:

$$\begin{aligned} W(s) \mid [W(t) = b] &\sim N\left(\frac{s}{t}b, \sigma^2\frac{s}{t}(t-s)\right), \quad s < t \\ W(t) \mid [W(s) = a] &\sim N(a, \sigma^2(t-s)) \end{aligned}$$

Beweis: Übung

$W(s) \mid [W(t) = b]$

Linearität des bedingten Erwartungswertes

$W(t) \mid [W(s) = a]$

„Neustart zur Zeit s in a “

Endlichdimensionale Verteilungen:

$$0 \leq t_1 < t_2 < \dots < t_n, \quad n \in \mathbb{N}$$

$$(W(t_1), \dots, W(t_n))' \sim N(0, \sigma^2 \Sigma),$$

$$\Sigma = \begin{pmatrix} t_1 & t_1 & t_1 & \dots & t_1 \\ t_1 & t_2 & t_2 & \dots & t_2 \\ t_1 & t_2 & t_3 & \dots & t_3 \\ \vdots & \vdots & \vdots & & \vdots \\ t_1 & t_2 & t_3 & \dots & t_n \end{pmatrix}$$

Dichte:

$$f_{t_1, \dots, t_n}(x_1, \dots, x_n) = \frac{\exp \left\{ -\frac{1}{2\sigma^2} \left[\frac{x_1^2}{t_1} + \frac{(x_2 - x_1)^2}{t_2 - t_1} + \dots + \frac{(x_n - x_{n-1})^2}{t_n - t_{n-1}} \right] \right\}}{(\sigma\sqrt{2\pi})^n \sqrt{t_1(t_2 - t_1) \dots (t_n - t_{n-1})}}$$

$$\stackrel{!}{=} f_{t_1}(x_1) f_{t_2|t_1}(x_2 | x_1) \dots f_{t_n|t_{n-1}}(x_n | x_{n-1})$$

Bemerkung: Dies zeigt die Markoveigenschaft von W , vgl. Kap. 8.

(b) **Pfade**

Die Pfade sind stetig, aber (m. Wkeit 1) nirgends differenzierbar und in jedem endlichen Intervall von unbeschränkter Variation

Beweis: vgl. A2, Kap 6, FKO.

(unbeschränkte Variation: $s = t_0 < t_1 < \dots < t_n = t$ Gitter auf $[s, t]$, Schrittweite $\delta_n = \frac{t-s}{n}$)

Dann gilt:

$$\sum_{k=1}^n |W(t_k) - W(t_{k-1})| \rightarrow \infty, \quad \text{für } n \rightarrow \infty$$

Plausibilitätserklärung für Nicht-Differenzierbarkeit:

Für den Diff.Quotient gilt

$$\frac{W(t+h) - W(t)}{h} \sim N\left(0, \frac{\sigma^2}{|h|}\right)$$

d.h die Varianz des Diff.quot. $\rightarrow \infty$ für $h \rightarrow 0$.

Es gilt sogar

$$P\left\{a \leq \frac{W(t+h) - W(t)}{h} \leq b\right\} \rightarrow 0, \quad [a,b] \text{ endl. Intervall}$$

\Rightarrow Diff.quot. kann nicht gegen endliche ZV konvergieren.

Weitere Eigenschaften, z.B. Markoveigenschaft, in Kap. 8.

Fazit: Trotz der harmlos erscheinenden Verteilungseigenschaften ergeben sich extrem „unglatte“ Pfade.

1.2.3 Zählprozesse und Poisson-Prozess

Zählprozesse

- $\{S_n, n \in \mathbb{N}\}$ SP von Ereigniszeitpunkten
 S_n Zeitpunkt des n -ten Ereignisses
 z.B.
 - ◊ Eintreten von Schadensfällen bei KFZ-Versicherung
 - ◊ Todesfälle in klinischer Studie
 - ◊ Ankünfte von Jobs bei Computer, von Kunden bei „Servicestelle“...
 - ◊ Kauf eines Produkts,
 - ◊ Transaktionen an Börse
- $\{T_n, n \in \mathbb{N}\}$ SP von Verweildauern, Zwischenankunftszeiten, etc. mit $T_n = S_n - S_{n-1}$
- Zählprozess

$$N(t) = \sum_{n=1}^{\infty} I_{(0,t]}(S_n) = \text{Anzahl der Ereignisse in } (0,t].$$
 (S_0 wird nicht gezählt)

Übereinkunft: $T_n > 0$ für alle $n \in \mathbb{N}$, d.h. keine gleichzeitigen Ereignisse.

Pfade = Treppenfunktion mit Sprunghöhe 1, rechtsstetig

Es gilt:

$$\begin{aligned} S_n \leq t &\Leftrightarrow N(t) \geq n \\ S_n \leq t < S_{n+1} &\Leftrightarrow N(t) = n \\ N(t) = \max_n \{S_n \leq t\} &= \min_n \{S_{n+1} > t\} \end{aligned}$$

Mehr zu Zählprozessen in Kapitel 7; hier: Poisson-Prozess als einfachster Zählprozess

Erweiterung:

Markierter (oder „bewerteter“) Zählprozess

Zu jedem Ereigniszeitpunkt S_n wird eine zweite Variable Y_n beobachtet.

z.B. Y_n Schadenshöhe bei n -ten Schaden

Y_n Preis(-veränderung) bei n -ter Transaktion

Y_n Todesart

Der Poisson-Prozess

FKO, S. (80/81) ff.

Definition 1.4 Zählprozess N mit unabhängigen und stationären Zuwächsen

N besitzt *unabhängige Zuwächse* $:\Leftrightarrow$

$$\forall n \forall 0 \leq t_0 < t_1 < \dots < t_n \quad \text{sind} \quad N(t_1) - N(t_0), \dots, N(t_n) - N(t_{n-1}) \quad \text{unabhängig}$$

N besitzt *stationäre Zuwächse* $:\Leftrightarrow$

$$\forall 0 \leq t_1 < t_2, s > 0 \quad \text{sind} \quad N(t_2) - N(t_1) \quad \text{und} \quad N(t_2 + s) - N(t_1 + s) \quad \text{identisch verteilt}$$

Definition 1.5 Poisson-Prozess

Der Zählprozess $N = \{N(t), t \geq 0\}$ heißt Poisson-Prozess $:\Leftrightarrow$

- (1) N hat unabhängige und stationäre Zuwächse
- (2) $P\{N(t+h) - N(t) \geq 2\} = o(h)$
 $P\{N(t+h) - N(t) = 1\} = \lambda h + o(h)$

Definition 1.6 $o(h)$

$o(h)$ bezeichnet eine Funktion von h , die für $h \downarrow 0$ schneller gegen 0 geht als h , d.h.

$$\lim_{h \downarrow 0} \frac{o(h)}{h} = 0$$

Bemerkung: (2) \Leftrightarrow Sprunghöhe der Pfade haben (ohne Beweis) (m. Wkeit 1) die Höhe 1
 \Leftrightarrow keine gleichzeitigen Ereignisse
 \Leftrightarrow Zählprozess nach unserer Übereinkunft
(ohne Beweis)

Satz 1.1 Poisson-Verteilung

Für Poisson-Prozess N gilt

$$P\{N(t) = n\} = \frac{(\lambda t)^n}{n!} e^{-\lambda t}, \quad n \in \mathbb{N}_0,$$

d.h. $N(t) \sim \text{Po}(\lambda t)$,

mit $\lambda \geq 0$ „Intensität“, „Rate“.

Bemerkung: (2) in Def. 1.5 kann durch $N(t) \sim \text{Po}(\lambda t)$ ersetzt werden

Beweis:

Sei $p_0(t) := P(N(t) = 0)$, $p_n(t) := P(N(t) = n)$

Wir zeigen zunächst $p_0(t) = e^{-\lambda t}$. Es gilt wegen Definition 1.5,(1)

$$\begin{aligned} p_0(t+h) &= P\{N(t+h) = 0\} \\ &= P\{N(t) = 0, N(t+h) - N(t) = 0\} \\ &= P\{N(t) = 0\} \cdot P\{N(t+h) - N(t) = 0\} \\ &= p_0(t)p_0(h) \\ \frac{p_0(t+h) - p_0(t)}{h} &= p_0(t) \frac{p_0(h) - 1}{h} \end{aligned}$$

Beachtet man noch, dass aus Def 1.5,(2)

$$p_0(h) = 1 - \lambda h + o(h)$$

folgt, so erhält man für $h \rightarrow 0$

$$p_0'(t) = -\lambda p_0(t), p_0(0) = 1.$$

Dies ist eine lineare Differentialgleichung für die unbekannte Funktion $p_0(t)$ und besitzt eine eindeutige Lösung. Einsetzen zeigt, dass $e^{-\lambda t}$ die Gleichung inklusive der Anfangsbedingung erfüllt.

Für $n > 0$ gilt

$$\begin{aligned} p_n(t+h) &= P\{N(t+h) = n\} \\ &= P\{N(t) = n, N(t+h) - N(t) = 0\} \\ &+ P\{N(t) = n-1, N(t+h) - N(t) = 1\} \\ &+ \underbrace{\sum_{k=2}^n P\{N(t) = n-k, N(t+h) - N(t) = k\}}_{o(h) \text{ wegen (2)}} \\ &= p_n(t)p_0(h) + p_{n-1}(t)p_1(h) + o(h) \\ &= p_n(t)(1-\lambda h) + p_{n-1}(t)\lambda h + o(h) \end{aligned}$$

$h \rightarrow 0$ liefert

$$p_n'(t) = -\lambda p_n(t) + \lambda p_{n-1}(t), \quad n = 1, 2, \dots$$

mit $p_0(t) = e^{-\lambda t}$, $p_n(0) = 0$.

Man verifiziert nun leicht, dass die Wahrscheinlichkeitsfunktion der Poisson-Verteilung (eindeutige) Lösung dieser Differentialgleichung ist. \square

Satz 1.2 N ist homogener Markov-Prozess

Für den Poisson-Prozess gilt mit $s_0 < \dots < s_n < s < t$

$$\begin{aligned} P(N(t) = j \mid N(s) = i, N(s_n) = i_n, \dots, N(s_0) = i_0) \\ &= P(N(t) = j \mid N(s) = i) = P(N(t) - N(s) = j - i) \\ &= P(N(t-s) = j - i) \\ &= \frac{(\lambda(t-s))^{j-i}}{(j-i)!} e^{-\lambda(t-s)} \quad j \geq i \end{aligned}$$

Beweis:

$$\begin{aligned} P\{N(t) = j \mid N(s) = i, N(s_n) = i_n, \dots, N(s_0) = i_0\} \\ &= P\{N(t) - N(s) = j - i \mid N(s) - N(s_n) = i - i_n, \dots\} \\ &\stackrel{(1)}{=} P\{N(t) - N(s) = j - i\} \\ &= P\{N(t) - N(s) = j - i \mid N(s) - N(0) = i\} \\ &= P\{N(t) = j \mid N(s) = i\} \quad (\text{Markoveigenschaft}) \end{aligned}$$

$$\begin{aligned}
P\{N(t) = j \mid N(s) = i\} &= P\{N(t) - N(s) = j - i\} \\
&\stackrel{(1)}{=} P\{N(t-s) - N(0) = j - i\} \quad (\text{Homogenität}).
\end{aligned}$$

Aus Satz 1.1 folgt damit die Formel für die Übergangswahrscheinlichkeit

$$P(N(t) = j \mid N(s) = i)$$

□

Bemerkung: Die erste Gleichung ist die *Markov-Eigenschaft* und besagt, dass bei Kenntnis des „gegenwärtigen“ Zustands $N(s) = i$ der „zukünftige“ Zustand $N(t) = j$ von der „Vergangenheit“ unabhängig ist. Wie der Beweis zeigt, wird für die Markov-Eigenschaft nur die Unabhängigkeit der Zuwächse benutzt. Also: Jeder Prozess mit unabhängigen Zuwächsen ist ein Markov-Prozess.

Satz 1.3

N ist Poisson-Prozess mit Rate $\lambda \Leftrightarrow$

Die Zwischenzeiten T_n sind iid $\text{Ex}(\lambda)$ verteilt (exponentialverteilt mit Parameter λ).

Beweis:

Wir zeigen nur die „ \Rightarrow “ Richtung. Es gilt

$$P\{T_1 > t\} = P\{N(t) = 0\} = e^{-\lambda t},$$

d.h. T_1 ist exponentialverteilt mit Parameter λ .

Weiter gilt wegen der Unabhängigkeit und Stationarität der Zuwächse sowie Satz 1.2

$$\begin{aligned}
P\{T_2 > t \mid T_1 = s\} &= P\{N(s+t) - N(s) = 0 \mid N(s) = 1\} \\
&= P\{N(s+t) - N(s) = 0\} \\
&= e^{-\lambda t}
\end{aligned}$$

unabhängig von s . Die Regel der totalen Wahrscheinlichkeit liefert deshalb

$$P\{T_2 > t\} = \int_0^\infty e^{-\lambda t} \lambda e^{-\lambda s} ds = e^{-\lambda t} = P\{T_2 > t \mid T_1 = s\},$$

also sind T_1 und T_2 unabhängig und exponentialverteilt. Die analoge Beweisführung für $n \geq 2$ liefert das allgemeine Ergebnis. □

Verallgemeinerung: **räumlicher Poisson-Prozess**

- Anzahl der Ereignisse in einem Gebiet A ist Poisson-verteilt mit E-Wert $\lambda \cdot \text{Fläche}(A)$.
- Anzahl der Ereignisse in zwei nicht überlappenden Gebieten A_1 und A_2 sind unabhängig.

Einige weitere Eigenschaften des Poisson-Prozesses

(FKO S. 85–90)

- $S_n = T_1 + \dots + T_n$ Wartezeit auf n -tes Ereignis, $S_0 = 0$.

Es gilt:

$$S_n \sim \text{Ga}(n, \lambda)$$

$$f_{S_n}(t) = \frac{e^{-\lambda t} \lambda^n t^{n-1}}{(n-1)!}$$

mit $E(S_n) = \frac{n}{\lambda}$, $\text{Var}(S_n) = \frac{n}{\lambda^2}$, $\text{Modus}(S_n) = \frac{n-1}{\lambda}$.**Beweis:**

(Zur Gammaverteilung)

$$\{N(t) \geq n\} \Leftrightarrow \{S_n \leq t\}$$

$$\Rightarrow F_{S_n}(t) = \sum_{j=n}^{\infty} e^{-\lambda t} \frac{(\lambda t)^j}{j!}$$

Differentiation liefert die Dichte der Gamma-Verteilung. □

- Vorwärts-und Rückwärtsrekurrenzenzeiten

 $V(t) = S_{N(t)+1} - t$ Vorwärtsrekurrenzenzeit

$U(t) = t - S_{N(t)}$ Rückwärtsrekurrenzzeit

$U(t) + V(t) = T_{N(t)+1} = S_{N(t)+1} - S_{N(t)}$

Bemerkung: $N(t)$ in $S_{N(t)}$ ist *zufälliger* Index;

$$T_{(N(t)+1)} \neq T_n, \text{ (n fest)}$$

Es gilt:

$$P(V(t) \leq x) = F_{V(t)}(x) = 1 - e^{-\lambda x}$$

d.h. $V(t) \sim \text{Ex}(\lambda)$

Bemerkung: $V(t) \sim \text{Ex}(\lambda)$ *unabhängig von gewähltem* t

\Leftrightarrow „Gedächtnislosigkeit“ der Exponentialverteilung:

$$X \sim \text{Ex}(\lambda) \Rightarrow P(X \leq t + x \mid X > t) = P(X \leq x)$$

$$\diamond P(U(t) = t) = e^{-\lambda t}, \quad P(U(t) \leq x) = 1 - e^{-\lambda x}, \quad 0 \leq x < t$$

(kein Ereignis vor t)

$$\diamond \text{Sampling Paradoxon: } t \text{ fest, } T_{N(t)+1} = V(t) + U(t)$$

$$E(T_{N(t)+1}) = E(U(t)) + E(V(t)) = \frac{1}{\lambda}(1 - e^{-\lambda t}) + \frac{1}{\lambda} > \frac{1}{\lambda} = E(T_n)$$

Plausibilitätserklärung: Bei fest vorgegebenem t ist $T_{N(t)+1}$ die zufällige Länge der enthaltenen Zwischenzeit. Im Schnitt werden längere Intervalle dabei favorisiert.

Überlagerung und Zerlegung von Poisson-Prozessen

$$\left. \begin{array}{l} L = \{L(t), t \geq 0\} \quad \text{PP mit Rate } \lambda \\ M = \{M(t), t \geq 0\} \quad \text{PP mit Rate } \mu \end{array} \right\} \text{unabhängig}$$

- Dann heißt $N = \{N(t), t \geq 0\}$ mit

$$N(t, \omega) = L(t, \omega) + M(t, \omega)$$

Überlagerung

Es gilt: N ist PP mit Rate $\lambda + \mu$

- **Zerlegung**

N PP mit Rate λ . Bei Eintritt eines Ereignisses wird ein binomisches Experiment $P(X = 1) = p$, $P(X = 0) = 1 - p$, unabhängig von N , durchgeführt.

$$X = 1 \quad : \quad \text{Typ-1 Ereignis} \Rightarrow \text{Zählprozess } M$$

$$X = 0 \quad : \quad \text{Typ-0 Ereignis} \Rightarrow \text{Zählprozess } L$$

Es gilt: M und L sind unabhängige PP mit Raten λp und $\lambda(1 - p)$

Beispiel: Netzwerke, etwa Straßensystem

Verallgemeinerungen des Poisson-Prozesses

Die Unabhängigkeit und *insbesondere* Stationarität der Zuwächse ist in Anwendungen oft kritisch (bereits diskutiert für Schadensfälle bei Versicherungen)

Definition 1.7 Inhomogener Poisson-Prozess

Ein Zählprozess $N = \{N(t), t \geq 0\}$ heißt *nichtstationärer (inhomogener)* PP mit Rate (Intensitätsfunktion) $\lambda(t)$, $t \geq 0 \Leftrightarrow$

(1) N hat unabhängige Zuwächse

(2) $P(N(t+h) - N(t) \geq 2) = o(h)$

$$P(N(t+h) - N(t) = 1) = \lambda(t)h + o(h)$$

Es lässt sich zeigen:

$$\begin{aligned} P(N(t+s) - N(t) = n) &= \exp\left(-\int_t^{t+s} \lambda(u) du\right) \frac{\left(\int_t^{t+s} \lambda(u) du\right)^n}{n!} \\ &= \exp(-(\Lambda(t+s) - \Lambda(t))) \frac{(\Lambda(t+s) - \Lambda(t))^n}{n!} \end{aligned}$$

mit $\Lambda(t) = \int_0^t \lambda(u) du$ als „kumulierte“ Rate, d.h.

$$N(t+s) - N(t) \sim \text{Po}\left(\int_t^{t+s} \lambda(u) du\right)$$

Definition 1.8 Bewerteter (Compound) Poisson-Prozess

N PP, $\{Y_n, n \in \mathbb{N}\}$ iid Folge, von N unabhängig.

$X = \{X(t), t \geq 0\}$ mit $X(t) = \sum_{n=1}^{N(t)} Y_n$ heißt bewerteter PP

Beispiele:

- N Schadensprozess
 $\{Y_n, n \in \mathbb{N}\}$ zugehörige Schadenshöhe
 $X(t) = \sum_{n=1}^{N(t)} Y_n$ Gesamtschaden in $[0, t]$
- Klassisches Versicherungsmodell von Cramér-Lundberg

$$\text{Risikoprozess } R(t) = c_0 + c_1 t - X(t)$$

c_1 = Prämienintensität

$X(t)$ = compound PP mit Intensität λ

Schadenhöhen Y_n iid. $\sim F$

$N(t)$ klassischer P.P.

Ziel: $P(R(t) > 0 \forall t) = ?$ bzw. $P(R(t) \leq 0) \leq ?$

Kapitel 2

Grundbegriffe der allgemeinen Theorie stochastischer Prozesse

Inhalt

- Definitionen von SP
- Verteilung eines SP: SP als W -Maß auf Funktionenraum
- Existenzsatz von Kolmogorov
- Pfadigenschaften

Hier nur „Skizze“, Details insbesondere Billingsley, Bauer.

2.1 Definitionen stochastischer Prozesse

2.1.1 Klassische Definition: SP als Familie von Zufallsvariablen

(Ω, \mathcal{F}, P) W -Raum

T Indexmenge (i.a. unendlich, z.B. $\mathbb{N}_0, \mathbb{R}_+, \dots$)

$\{X_t, t \in T\}$ Familie von ZV

$$X_t : (\Omega, \mathcal{F}, P) \longrightarrow (S, \mathcal{S})$$

mit Wertebereich S , σ -Algebra \mathcal{S} .

- S abzählbar, $\mathcal{S} = \mathcal{P}(S)$ „diskrete ZV“
- $S = \mathbb{R}$ oder Intervall $I \subset \mathbb{R}$, $\mathcal{S} = \mathcal{B}$ oder $\mathcal{B} \cap I$ „reellwertige ZV“

- $S = \mathbb{R}^p$ oder $I \subset \mathbb{R}^p$, $\mathcal{B} = \mathcal{B}^p$ oder $\mathcal{B}^p \cap I$ „vektorielle ZV“

Definition 2.1 Stochastischer Prozess

Ein stochastischer Prozess (SP) ist das Quadrupel $X = (\Omega, \mathcal{F}, P; X_t, t \in T)$, T heißt Parameterraum, S Zustandsraum.

Bemerkung:

- (a) Meist lässt man W-Raum (Ω, \mathcal{F}, P) weg. Also: SP $X = \{X_t, t \in T\}$ Familie von (i.a. abhängigen) ZV X_t .
- (b)
- Für $T = \mathbb{N}_0$ oder \mathbb{R}_+ wird t meist als diskrete oder stetige Zeit interpretiert.
 - Ist $T \subset \mathbb{Z}^2$ (Gitter) oder $T \subset \mathbb{R}^2$, heißt ein SP auch Zufallsfeld (random field) \Rightarrow „räumliche Statistik“.
 - $T \subset \mathbb{N}_0 \times \mathbb{Z}^2 \Rightarrow$ zeitlich-räumliche Statistik.

Klassifizierung von SP nach Zustands- und Parameterraum

Definition 2.2 Endlich-dimensionale Verteilungen und Verteilungsfamilien

Sei X SP und sei $\{t_1, \dots, t_n, n \in \mathbb{N}\} \subset T$ beliebig.

Dann heißen

$$P_{t_1, \dots, t_n}(B_1 \times \dots \times B_n) = P(X_{t_1} \in B_1, \dots, X_{t_n} \in B_n)$$

endlich-dimensionale Verteilungen des SP X .

Für reelle ZV heißen

$$F_{t_1, \dots, t_n}(x_1, \dots, x_n) = P(X_{t_1} \leq x_1, \dots, X_{t_n} \leq x_n)$$

endlich-dimensionale Verteilungsfunktionen des SP X .

Die Menge *aller* endlich-dimensionalen Verteilungsfunktionen heißt die *Familie der endlich-dimensionalen Verteilungsfunktionen*

Folgende *Verträglichkeitsbedingungen* („Konsistenzbedingungen“) gelten für alle $n \in \mathbb{N}$, $t_1, \dots, t_n \in T$:

- (a) $F_{t_{k_1}, \dots, t_{k_n}}(x_{k_1}, \dots, x_{k_n}) = F_{t_1, \dots, t_n}(x_1, \dots, x_n)$ für jede Permutation k_1, \dots, k_n von $1, \dots, n$
- (b) $F_{t_1, \dots, t_k}(x_1, \dots, x_k) = F_{t_1, \dots, t_n}(x_1, \dots, x_k, \infty, \dots, \infty) \quad \forall 1 \leq k < n$ und $x_1, \dots, x_k \in \mathbb{R}$

Verträglichkeitsbedingungen gelten analog für Verteilungen ($F \rightarrow P$, $x_1, \dots, x_k \rightarrow B_1, \dots, B_k$).

Definition 2.3 Konsistente Verteilungsfamilie

Eine endlich-dimensionale Verteilungsfamilie heißt *konsistent* \Leftrightarrow (a), (b) gelten.

Definition 2.4 Pfad, Trajektorie, Realisierung

Für jedes (feste) $\omega \in \Omega$ heißt die Funktion

$$X(\omega) : T \rightarrow S, \quad t \mapsto X_t(\omega),$$

Pfad, Trajektorie oder *Realisierung* des SP X .

Also: ω fest, t läuft; $X(\omega)$ übliche (reelle) Folge (T diskret) bzw. Funktion (T stetig).

2.1.2 Stochastischer Prozess als Produktabbildung

SP lässt sich auch auffassen als Funktion der *beiden* Variablen t und ω

$$\begin{aligned} X(\cdot) : T \times \Omega &\rightarrow S && \text{„Produktabbildung“} \\ (t, \omega) &\mapsto X_t(\omega) \end{aligned}$$

Hält man in der Produktabbildung t *fest*, erhält man die ZV

$$\begin{aligned} X_t : \Omega &\rightarrow S && t \text{ fest} \\ \omega &\mapsto X_t(\omega) && \omega \text{ läuft} \end{aligned}$$

zurück.

Aber: Produktabbildung $X(\cdot)$ *i.a.* nicht messbar. Muss zusätzlich gefordert werden. Dann heißt der SP X *messbar*. Die meisten praktisch vorkommenden Prozesse sind (als Produktabbildung) messbar, insbesondere alle die wir besprechen.

Alle SP mit diskretem T sind messbar.

2.1.3 Stochastischer Prozess als Abbildung in Funktionenraum

Ordnet man jedem $\omega \in \Omega$ „seinen“ Pfad $X(\omega) \in S^T$ ($S^T =$ Raum aller Funktionen $T \rightarrow S$) zu, kann man einen stochastischen Prozess auch auffassen als Abbildung in den Funktionenraum

$$X : (\Omega, \mathcal{F}, P) \rightarrow S^T$$

$$\omega \mapsto X(\omega) = \{X_t(\omega), t \in T\}$$

Vergleich mit mehrdimensionalen ZV

Stochastischer Prozess $T = \mathbb{R}_+, \mathbb{N}_0, \dots$ $S \subset \mathbb{R}$ $X(\omega)$ Funktion bzw. Folge = Punkt im Funktionenraum S^T	p -dim ZV, $p = 1$: reelle ZV $T = \{1, 2, \dots, p\}$ $S = \mathbb{R}, \subset \mathbb{R}$ $X(\omega)$ Punkt im \mathbb{R}^p , $X(\omega) \in \mathbb{R}^p$
--	--

Also: Stochastischer Prozess X ist „ZV“ mit S^T als Wertebereich.

Problem: Lässt sich auf Funktionenraum eine σ -Algebra \mathcal{A} definieren, so dass

$$X : (\Omega, \mathcal{F}) \rightarrow (S^T, \mathcal{A}) \text{ messbar ist?}$$

Antwort: Ja, wobei es oft zweckmäßig ist, den Funktionenraum einzuschränken.

Beispiel: Wiener Prozess; hat stetige Pfade

$$W : (\Omega, \mathcal{F}, P) \rightarrow (C(T), \mathcal{B}_C)$$

$$C(T) = \text{Raum aller stetigen Funktionen, } \mathcal{B}_C = \text{Borel-}\sigma\text{-Algebra}$$

Dann lässt sich P auf (Ω, \mathcal{F}) als Bild-Wahrscheinlichkeits-Maß P_X auf (S^T, \mathcal{A}) verpflanzen:

$$W : (\Omega, \mathcal{F}, P) \rightarrow (S^T, \mathcal{A}, P_X).$$

\Rightarrow Auffassung des stochastischen Prozesses X als W-Maß P_X auf Funktionenraum.

Details: Billingsley, Gänsler/Stute, 7.3

2.2 Existenzsatz von Kolmogorov

In 2.1.1 wurde zur Definition eines stochastischen Prozesses X als Familie von ZV $\{X_t, t \in T\}$, $X_t : (\Omega, \mathcal{F}, P) \rightarrow (S, \mathcal{S})$ die *Existenz eines gemeinsamen W-Raumes* (Ω, \mathcal{F}, P) vorausgesetzt.

Aus der Definition können dann insbesondere die endlich-dimensionalen Verteilungen abgeleitet werden.

Bei den Beispielen (Irrfahrt, Poisson-Prozess, Wiener-Prozess, ...) wurde aber $(\Omega, \mathcal{F}, P) \rightarrow (S, \mathcal{S})$ nicht explizit angegeben.

Diskrete Irrfahrt:

$$X_n = Z_1 + \dots + Z_n, \quad n \in \mathbb{N}$$

Ereignis $\omega = (z_1, \dots, z_n, \dots)$ mit $z_n \in \{-1, 0, 1\}$

$$x_n(\omega) = z_1 + \dots + z_n$$

$\Omega = (-1, 0, 1)^{\mathbb{N}}$ Folgenraum, $\mathcal{F} = \mathcal{P}(\Omega)$

$P = P_1 \times \dots \times P_n \times \dots$ Produkt-Maß

$$\text{mit } P_n := \begin{cases} p & \text{für } z_n = 1 \\ q & \text{für } z_n = -1 \\ r & \text{für } z_n = 0 \end{cases}$$

Poisson-Prozess:

$$\omega = (\omega_1, \omega_2, \dots, \omega_n, \dots), \quad \omega_n \in \mathbb{R}_+$$

$\omega_n =$ Ergebnis von $T_n \sim Ex(\lambda)$

Ω Menge aller solcher Folgen

$\mathcal{F} = \mathcal{B}_+ \times \mathcal{B}_+ \times \dots \times \mathcal{B}_+ \times \dots$ (Produkt- σ -Algebra)

$P = P_1 \times P_2 \times \dots \times P_n \times \dots$ Produkt-Maß zu Exp.-Verteilungen P_1, \dots, P_n

In beiden Beispielen lassen sich zu jedem ω die Pfade $X(\omega)$, $N(\omega)$ der ZVen X_n bzw. $N(t)$ als *messbare*

Abbildungen definieren.

In den meisten Fällen ist jedoch eine solche explizite Angabe von (Ω, \mathcal{F}, P) nicht möglich, z.B.: Aktienkurse, Folge von Marktwahlen, DNA-Strang.

Der Existenzsatz von Kolmogorov zeigt, dass es reicht, wenn man endlich-dimensionale Verteilungen in *konsistenter* Weise vorgibt.

Satz 2.1 Existenzsatz von Kolmogorov

Sei $\{F_{t_1, \dots, t_n}\}$ (bzw. $\{P_{t_1, \dots, t_n}\}$) ein konsistentes System von endlich-dimensionalen Verteilungsfunktionen (bzw. Verteilungen).

Dann existiert ein W-Raum (Ω, \mathcal{F}, P) und ein stochastischer Prozess

$$X = \{\Omega, \mathcal{F}, P, (X_t, t \in T)\}$$

mit F_{t_1, \dots, t_n} als System von endlich-dimensionalen Verteilungsfunktionen, d.h.

$$F_{t_1, \dots, t_n}(x_1, \dots, x_n) = P(X_{t_1} \leq x_1, \dots, X_{t_n} \leq x_n)$$

Bemerkung:

- (a) Der stochastische Prozess X ist durch den Existenzsatz *nicht eindeutig* bestimmt. Es lässt sich immer ein Prozess \tilde{X} konstruieren, der zwar andere Pfade besitzt als X , aber die gleichen endlich-dimensionalen Verteilungsfunktionen! (Vgl. auch Abschnitt 2.3).

Beispiel: X mit stetigen Pfaden.

X und \tilde{X} haben verschiedene Pfade, aber die gleichen endlich-dimensionalen Verteilungsfunktionen!

- (b) Spezialfall: Mehrdimensionale ZV, $T = \{1, \dots, p\}$.

Existenzsatz sichert zu vorgegebener gemeinsamer Verteilungsfunktion $F(x_1, \dots, x_p)$ die Existenz eines W-Raumes $\{\Omega, \mathcal{F}, P\}$ mit

$$F(x_1, \dots, x_p) = P(X_1 \leq x_1, \dots, X_p \leq x_p).$$

Dabei kann man sogar $(\Omega, \mathcal{F}) = (\mathbb{R}^p, \mathcal{B}^p)$ wählen, d.h. die Realisierungen x_1, \dots, x_p werden mit ω identifiziert, $\omega \equiv (x_1, \dots, x_p) \in \mathbb{R}^p$.

Analog bei stochastischen Prozessen:

$$\omega \equiv \text{Pfad} \in S^T,$$

$$\Omega = \text{Menge aller Pfade.}$$

\Rightarrow Man braucht sich keinen zugrundeliegenden W-Raum wie bei Irrfahrt und Poisson-Prozess konstruieren.

(c) Beispiel: Wiener Prozess

Vorgabe von (W1), (W2), d.h. unabhängige und stationäre Zuwächse und W(3), d.h. $W(0) = 0$

\Rightarrow endlich-dimensionale Verteilungsfunktionen wie in Abschnitt 1.2.2

\Rightarrow Existenzsatz garantiert stochastischen Prozess \tilde{W} , aber nicht notwendig stetige Pfade.

Es lässt sich aber eine „Version“ W mit *stetigen Pfaden* konstruieren, vgl. Billingsley, Sect. 3.6.

(d) In den weiteren Kapiteln (analog wie bei Irrfahrt, Poisson-Prozess, Wiener Prozess):

Vorgabe von „Konstruktionsvorschriften“ bzw. „Axiomen“

\Rightarrow endlich-dimensionale Verteilungsfunktionen

\Rightarrow stochastischer Prozess inklusive W-Raum

2.3 Äquivalenz- und Stetigkeitsbegriffe

2.3.1 Äquivalente stochastische Prozesse

$X = \{X_t, t \in T\}$ und $Y = \{Y_t, t \in T\}$ seien zwei stochastische Prozesse auf gleichem W-Raum (Ω, \mathcal{F}, P) mit gleichem (S, \mathcal{S}) .

Definition 2.5 Verteilungsäquivalenz, schwache Äquivalenz

X und Y *verteilungsäquivalent* (*schwach äquivalent*) $:\Leftrightarrow$

die endlich-dimensionalen Verteilungen von X und Y sind gleich \Leftrightarrow

$\forall \{t_1, \dots, t_n\} \subset T, \{B_1, \dots, B_n\} \subset \mathcal{S}, n \in \mathbb{N}$

$$P(X_{t_1} \in B_1, \dots, X_{t_n} \in B_n) = P(Y_{t_1} \in B_1, \dots, Y_{t_n} \in B_n).$$

Definition 2.6 Äquivalenz

X und Y heißen *äquivalent* $:\Leftrightarrow$

$$P(X_t = Y_t) = 1 \quad \forall t \in T$$

Bemerkung: Man sagt: Y ist „Version“ von X .

Definition 2.7 Ununterscheidbar

X und Y heißen *ununterscheidbar* $:\Leftrightarrow$

X und Y haben mit Wahrscheinlichkeit 1 gleiche Pfade \Leftrightarrow

$$P\{X_t = Y_t \quad \forall t \in T\} = 1 \quad \Leftrightarrow \quad P(\{\omega : X(\omega) = Y(\omega)\}) = 1$$

Es gilt:

X, Y ununterscheidbar \Rightarrow äquivalent \Rightarrow verteilungsäquivalent

Falls T abzählbar: X, Y ununterscheidbar \Leftrightarrow äquivalent

Beweis:

Ununterscheidbar \Rightarrow Äquivalent: klar

Äquivalent \Rightarrow Verteilungsäquivalent:

$$\begin{aligned} P(X_{t_1} \in B_1, \dots, X_{t_n} \in B_n) &= P(X_{t_1} \in B_1, \dots, X_{t_n} \in B_n, X_{t_1} = Y_{t_1}, \dots, X_{t_n} = Y_{t_n}) \\ &= P(Y_{t_1} \in B_1, \dots, Y_{t_n} \in B_n, X_{t_1} = Y_{t_1}, \dots, X_{t_n} = Y_{t_n}) \\ &= P(Y_{t_1} \in B_1, \dots, Y_{t_n} \in B_n) \end{aligned}$$

T abzählbar: Äquivalent \Rightarrow Ununterscheidbar

$$\begin{aligned} P(X_t = Y_t) = 1 \quad \forall t \in \mathbb{N} &\Leftrightarrow P(X_t \neq Y_t) = 0 \quad \forall t \in \mathbb{N} \\ &\Rightarrow \sum_{t=1}^{\infty} P(X_t \neq Y_t) = 0 \\ P\left(\bigcup_{t=1}^{\infty} \{X_t \neq Y_t\}\right) &\leq \sum_{t=1}^{\infty} P(X_t \neq Y_t) = 0 \\ \Rightarrow P\left(\overline{\bigcup_{t=1}^{\infty} \{X_t \neq Y_t\}}\right) &\stackrel{\text{(de Morgan)}}{=} P\left(\bigcap_{t=1}^{\infty} \{X_t = Y_t\}\right) = 1 \end{aligned}$$

□

Gegenbeispiel für nicht abzählbares T :

$$\begin{aligned} X : X_t(\omega) &= 0 \quad \forall t \in T, \omega \in \Omega, \text{ d.h. alle Pfade } \equiv 0 \\ Y : Y_t(\omega) &= \begin{cases} 1 & , \quad t = \tau(\omega) \\ 0 & , \quad \text{sonst} \end{cases} \end{aligned}$$

Dabei ist $\tau > 0$ stetige ZV, z.B. $\tau \sim \text{Ex}(\lambda)$

X und Y verteilungsäquivalent.

Da $X_t(\omega) = 0$ für alle $(t, \omega) \Rightarrow$

$$\begin{aligned} P(\omega : Y_t(\omega) = X_t(\omega)) &= P(\omega : Y_t(\omega) = 0) \\ &= P(\omega : \tau(\omega) \neq t) = 1 \quad \forall t \in T \end{aligned}$$

$\Rightarrow X$ und Y äquivalent, aber unterscheidbar.

2.3.2 Stetigkeitsbegriffe

$T \in \mathbb{R}_+$, $S \subset \mathbb{R}$, $\mathcal{S} = \mathcal{B} \cap S$

Definition 2.8 Pfadstetig

X heißt (fast sicher) pfadstetig $:\Leftrightarrow$

$$\begin{aligned} P(\omega : \lim_{s \rightarrow t} X(s, \omega) = X(t, \omega) \quad \forall t \in \mathbb{R}_+) &= 1 \\ (= P(\lim_{s \rightarrow t} X(s) = X(t) \quad \forall t \in \mathbb{R}_+) &= 1) \end{aligned}$$

Analog: fast sicher rechtsstetig, fast sicher cadlag (rechtsstetig mit Grenzwerten von links)

Definition 2.9 Stetig

X heißt (fast sicher) stetig $:\Leftrightarrow$

$$P(\omega : \lim_{s \rightarrow t} X(s, \omega) = X(t, \omega)) = 1 \quad \forall t \in \mathbb{R}_+$$

Definition 2.10 Stochastisch Stetig

X heißt stochastisch stetig $:\Leftrightarrow$

$$\begin{aligned} \lim_{s \rightarrow t} P(\omega : |X(s, \omega) - X(t, \omega)| > \epsilon) &= 0 \quad \forall \epsilon > 0, t \in \mathbb{R}_+ \\ \Leftrightarrow \text{p-}\lim_{s \rightarrow t} X(s) &= X(t) \quad \forall t \in \mathbb{R}_+ \end{aligned}$$

„ $X(s)$ konvergiert nach Wahrscheinlichkeit gegen $X(t)$, wenn $s \rightarrow t, \forall t \in \mathbb{R}_+$ “.

Beispiel: Poisson-Prozess

- (a) N nicht pfadstetig
- (b) N fast sicher stetig
- (c) N stochastisch stetig

Beweis:

(c) N stochastisch stetig:

$$P(|N(t) - N(s)| > \epsilon) \leq P(|N(t) - N(s)| \geq 1) = 1 - e^{-\lambda|t-s|}$$

$$\lim_{s \rightarrow t} P(|N(t) - N(s)| > \epsilon) \leq \lim_{s \rightarrow t} (1 - e^{-\lambda|t-s|}) = 0$$

(b) N fast sicher stetig:

$$\{\omega : N(t, \omega) \text{ stetig in } t \geq 0\} = \{\omega : S_n(\omega) \neq t, \forall n \in \mathbb{N}\}$$

Da S_n Gamma-verteilte ZV:

$$P\{\omega : S_n(\omega) = t\} = 0 \quad \forall n \in \mathbb{N}$$

$$\Rightarrow P\left(\overline{\bigcap_{n=1}^{\infty} \{\omega : S_n(\omega) \neq t\}}\right) = P\left(\bigcup_{n=1}^{\infty} \{\omega : S_n(\omega) = t\}\right)$$

$$\leq \sum_{n=1}^{\infty} P\{\omega : S_n(\omega) = t\} = 0$$

$$\Rightarrow P\left(\bigcap_{n=1}^{\infty} \{\omega : S_n(\omega) \neq t\}\right) = 1$$

□

Satz 2.2

X, Y äquivalent und X, Y fast sicher pfad-rechtsstetig $\Rightarrow X, Y$ ununterscheidbar.

Beweis: z.B. Todorovic, prop. 1.3.1, S. 7

2.4 Stationäre und nichtstationäre stochastische Prozesse

Stationäre stochastische Prozesse sind stochastische Prozesse bei denen sich die endlich-dimensionale Verteilungsfunktion oder andere Charakteristika bei einer Zeitverschiebung (näherungsweise) nicht ändern.

Beispiel:

- „Rauschen“ in elektronischen Systemen
- Abweichungen in Regelsystemen
- physiologische Daten (?), z.B. EKG
- Zeitreihen nach Trend-/Saisonbereinigung
- Renditen in effizienten Märkten (?)

FKO S. 209, Kap. 7/8

Definition 2.11 Strenge Stationarität

Ein stochastischer Prozess X heißt streng stationär $:\Leftrightarrow$

$$\forall n, t_1, \dots, t_n, h \quad F_{t_1, \dots, t_n}(x_1, \dots, x_n) = F_{t_1+h, \dots, t_n+h}(x_1, \dots, x_n)$$

d.h. die endlich-dimensionalen Verteilungsfunktionen sind invariant gegenüber einer Zeitverschiebung h .

Folgerungen: $F_{t_1}(x_1) = F_{t_1+h}(x_1)$, d.h. die eindimensionalen Verteilungen sind zeitinvariant.

- $E(X_t) = \mu = \text{constant}$, $\text{Var}(X_t) = \sigma^2 = \text{constant}$.

•

$$\begin{aligned}
\text{Cov}(X_s, X_t) &= \int (x_1 - \mu)(x_2 - \mu) dF_{s,t}(x_1, x_2) \\
&= \int (x_1 - \mu)(x_2 - \mu) dF_{0,t-s}(x_1, x_2) \\
&= \text{Cov}(X_0, X_{t-s}) \\
&=: \gamma(t-s)
\end{aligned}$$

d.h. die Kovarianz zwischen X_s, X_t hängt nur von Zeitdifferenz $t-s$, nicht von den Werten t, s auf der Zeitachse selbst ab!

Definition 2.12 Autokovarianzfunktion

$\gamma(h) = \text{Cov}(X_{t+h}, X_t) = \text{Cov}(X_h, X_0)$ heißt *(Auto-)Kovarianzfunktion*.

Eigenschaften:

$$\begin{aligned}
\gamma(h) &= \gamma(-h) \\
|\gamma(h)| &\leq \gamma(0) = \sigma^2 = \text{Var}(X_t)
\end{aligned}$$

Definition 2.13 Schwache Stationarität

Ein stochastischer Prozess X heißt *schwach stationär* $:\Leftrightarrow$

$$E(X_t) = \mu, \quad \text{Cov}(X_t, X_s) = \gamma(t-s)$$

d.h. die Autokovarianz hängt nur von Zeitdifferenz ab.

Offensichtlich gilt:

$$\text{Strenge Stationarität} \Rightarrow \text{Schwache Stationarität}$$

Die Umkehrung gilt für Gauß-Prozesse, d.h. Prozesse mit multivariat normalverteilten endlich-dimensionalen Verteilungsfunktionen, da die 1. und 2. Momente die endlich-dimensionalen Verteilungsfunktionen eindeutig bestimmen. Also:

X Gauß-Prozess \Rightarrow Schwache Stationarität und starke Stationarität sind äquivalent.

Bemerkung: Es gibt auch Nicht-Gauß-Prozesse, bei denen schwache und starke Stationarität äquivalent sind.

Definition 2.14 Autokorrelationsfunktion (ACF)

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)} = \frac{\gamma(h)}{\sigma^2} \quad h \geq 0$$

heißt *Autokorrelationsfunktion*.

Beispiel:

(a) Poisson-Prozess:

Der Poisson-Prozess N ist nichtstationär, da $E(N(t)) = \text{Var}(N(t)) = \lambda t \neq \text{constant}$. Dagegen sind die Zuwächse stationär (und unabhängig).

(b) Wiener-Prozess:

Der Wiener-Prozess W ist ebenfalls nichtstationär, da zwar $E(W(t)) = 0$, aber $\text{Var}(W(t)) = \sigma^2 t$ gilt. Die Zuwächse sind stationär (und unabhängig).

(c) Zufällige trigonometrische Polynome:

A und B seien identisch verteilte, unkorrelierte ZVen mit Erwartungswert 0 und Varianz σ^2 . Für eine feste Frequenz ω sei für $t \in \mathbb{Z}$

$$X_t = A \cos(\omega t) + B \sin(\omega t)$$

Dann ist $X = \{X_t, t \in \mathbb{Z}\}$ stationär mit $E(X_t) = 0$ und

$$\begin{aligned} \gamma(h) &= E(X_t X_{t+h}) \\ &= E\{(A \cos(\omega t) + B \sin(\omega t))(A \cos(\omega(t+h)) + B \sin(\omega(t+h)))\} \\ &= E\{A^2 \cos(\omega t) \cos(\omega(t+h)) + B^2 \sin(\omega t) \sin(\omega(t+h))\} \\ &= \sigma^2 \cos(\omega h) \end{aligned}$$

(d) Moving Average-Prozess:

Sei $\epsilon = \{\epsilon_t, t \in \mathbb{Z}\}$ eine Folge von unkorrelierten ZVen mit $E(\epsilon_t) = 0$, $\text{Var}(\epsilon_t) = \sigma^2$. Ein solcher stationärer stochastischer Prozess heißt (diskretes) Weißes Rauschen. Gilt zusätzlich ϵ_t iid $N(0, \sigma^2)$, so heißt ϵ Gauß'sches Weißes Rauschen.

Definition 2.15 Moving Average-Prozess

Der stochastische Prozess $X = \{X_t, t \in \mathbb{Z}\}$ mit

$$X_t = \sum_{j=0}^q \theta_j \epsilon_{t-j}, \quad \theta_q \neq 0$$

heißt moving average-Prozess (Prozess der gleitenden Durchschnitte) der Ordnung q , i.Z. $X \sim MA(q)$

Aus der Definition folgt sofort

$$E(X_t) = \sum_{j=0}^q \theta_j E(\epsilon_{t-j}) = 0.$$

Die Kovarianzfunktion berechnet sich wegen

$$\begin{aligned} X_t &= \theta_0 \epsilon_t + \theta_1 \epsilon_{t-1} + \dots + \theta_{h-1} \epsilon_{t-h+1} + \theta_h \epsilon_{t-h} + \dots + \theta_q \epsilon_{t-q} \\ X_{t-h} &= \theta_0 \epsilon_{t-h} + \dots + \theta_{q-h} \epsilon_{t-q} + \dots + \theta_q \epsilon_{t-q-h} \end{aligned}$$

und

$$E(\epsilon_t \epsilon_s) = \delta_{t,s} \sigma^2$$

zu

$$\gamma(h) = E(X_t X_{t-h}) = \begin{cases} \sigma^2 (\theta_0 \theta_h + \theta_1 \theta_{h+1} + \dots + \theta_{q-h} \theta_q) & , \quad |h| \leq q \\ 0 & , \quad |h| > q. \end{cases}$$

Speziell ergibt sich

$$\text{Var}(X_t) = \sigma_x^2 = \gamma(0) = \sigma^2 \sum_{j=0}^q \theta_j^2.$$

Es gilt also:

Jeder $MA(q)$ -Prozess (mit $q < \infty$) ist stationär mit ACF

$$\rho(h) = \begin{cases} \frac{\sum_{j=0}^{q-h} \theta_j \theta_{j+h}}{\sum_{j=0}^q \theta_j^2} & , \quad 0 \leq h \leq q \\ 0 & , \quad h > q. \end{cases}$$

(e) Modellierung ökonomischer Zeitreihen durch stationäre Prozesse:

Die Annahme, dass eine ökonomische Zeitreihe, d.h. eine Reihe von ökonomischen Daten y_t , $t = 1, 2, \dots$ als Pfad eines stationären stochastischen Prozesses angesehen werden kann, ist oft nicht sehr realistisch. Stationäre Prozesse können aber als Baustein für realitätsgetreuere Modelle dienen. Unterstellt man etwa einen linearen Trend, so ist ein möglicher Ansatz

$$Y_t = \beta_0 + \beta_1 t + X_t,$$

wobei X selbst stationär ist. Allgemeiner geht man beim sogenannten Komponentensatz (Kap 8.1) davon aus, dass

$$Y_t = T_t + S_t + X_t$$

gilt, mit der Trendkomponente T_t und der Saisonkomponente S_t und X_t stationär, z.B. MA- oder AR-Prozess.

(f) Autoregressive Prozesse:

Gauß'sche Irrfahrt (Random walk):

$$\begin{aligned} X_0 &= 0 \\ X_t &= X_{t-1} + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma^2) \quad \text{iid} \end{aligned}$$

Nichtstationär, da zwar $E(X_t) = 0$ aber, $\text{Var}(X_t) = t * \sigma^2$.

Aber: $AR(1)$ -Prozess

$$X_t = \delta X_{t-1} + \epsilon_t \quad \text{mit} \quad \epsilon_t \sim N(0, \sigma^2)$$

mit $|\delta| < 1$ (zentrale Bedingung!) ist (für $t \rightarrow \infty$) stationär.

Man kann zeigen, dass für jeden Startwert X_0 , die Verteilung von X_t für $t \rightarrow \infty$ gegen eine $N(0, \frac{\sigma^2}{(1-\delta^2)})$ konvergiert. Alternativ kann man $N(0, \frac{\sigma^2}{(1-\delta^2)})$ für X_0 schon fordern, d.h. der Prozess wird bereits im „Equilibrium“/„Gleichgewicht“ gestartet und es gilt (nicht nur asymptotisch, sondern exakt):

- $E(X_t) = 0$
- $\text{Var}(X_t) = \frac{\sigma^2}{(1-\delta^2)}$
- $\text{Corr}(X_s, X_t) = \delta^{|s-t|}$, insbesondere $\text{Corr}(X_t, X_{t+1}) = \delta$

Bei Vorliegen eines $AR(1)$ -Prozess mit unbekanntem δ kann man also ρ durch die empirische Autokorrelation (zum Lag 1)

$$\frac{\frac{1}{T} \sum_{t=2}^T (X_t - \bar{X})(X_{t-1} - \bar{X})}{\widehat{\text{Var}}(X_t)}$$

schätzen.

$$\widehat{\text{Var}}(X_t) = \frac{1}{T} \sum_{t=0}^T (X_t - \bar{X})^2 = \text{empirische Varianz}$$

- Mehr dazu in **Zeitreihenanalysen**
- $AR(p)$ -Prozess etc.
- Schätzung der empirischen Korrelationsfunktion.

$$\hat{\rho}(L) = \frac{\frac{1}{T} \sum_{t=L+1}^T (X_t - \bar{X})(X_{t-L} - \bar{X})}{\widehat{\text{Var}}(X_t)}, \quad L = \text{„Lag } L\text{“}.$$

(g) Stationäre Gauß-Prozesse:

$$\{X(t), t \geq 0\} \text{ oder } \{X_t, t \in \mathbb{N}_0\}$$

$$E(X(t)) = \mu, \text{ Var}(X(t)) = \sigma^2$$

$$\rho(h) = \text{Corr}(X(t), X(t+h)) = \text{Corr}(X(0), X(h)) \text{ (vorgegebene Korrelationsfunktion)}$$

Stationärer Gauß-Prozess : \Leftrightarrow Für alle $n \geq 1, t_1, \dots, t_n$ ist $X_{(n)} = (X(t_1), \dots, X(t_n))'$ multivariat normalverteilt mit

$$E(X_{(n)}) = (\mu, \dots, \mu)'$$

$$\text{Cov}(X_{(n)}) = \begin{pmatrix} \sigma^2 & \sigma^2 \rho(t_2 - t_1) & \dots & \sigma^2 \rho(t_n - t_1) \\ \sigma^2 \rho(t_1 - t_2) & \sigma^2 & & \sigma^2 \rho(t_n - t_2) \\ \vdots & \ddots & & \vdots \\ \sigma^2 \rho(t_1 - t_{n-1}) & & \ddots & \sigma^2 \rho(t_n - t_{n-1}) \\ \sigma^2 \rho(t_1 - t_n) & \dots & \sigma^2 \rho(t_{n-1} - t_n) & \sigma^2 \end{pmatrix}$$

Kapitel 3

Markov-Ketten

Diskrete Zeit $T = \mathbb{N}_0$, diskreter Zustandsraum S .

In Beispielen / Anwendungen: S oft endlich. Ausnahme: z.B. einfache Irrfahrt ohne Schranken.

Inhalt: FKO, Kapitel 2 in Auszügen und Ergänzungen

3.1 Grundlegende Eigenschaften, Beispiele

Beispiel: Einfache Irrfahrt

$$X_{t+1} = X_t + Z_t, \quad Z_t \in \{-1, 0, +1\}$$

$$P(Z_t = 0 \mid X_t = i) = r_i, \quad P(Z_t = -1 \mid X_t = i) = q_i, \quad P(Z_t = +1 \mid X_t = i) = p_i, \quad p_i + r_i + q_i = 1$$

Offensichtlich hängt damit die Wahrscheinlichkeitsfunktion von X_{t+1} , gegeben der bisherige Pfad

$$\{X_t = i, X_{t-1} = i_{t-1}, \dots, X_0 = i_0\}$$

nur vom gegenwärtigen Zustand $X_t = i$ ab.

Definition 3.1 Markov-Kette (MK) 1.Ordnung

Der SP $X = \{X_t, t \in \mathbb{N}_0\}$, S diskret, heißt MK $:\Leftrightarrow \forall t \in \mathbb{N}_0; j, i, i_{t-1}, \dots, i_0 \in S$

1. $P(X_{t+1} = j \mid X_t = i, X_{t-1} = i_{t-1}, \dots, X_0 = i_0) = P(X_{t+1} = j \mid X_t = i) \Leftrightarrow$
2. $P(X_{t+1} = j \mid X_t, X_{t-1}, \dots, X_0) = P(X_{t+1} = j \mid X_t)$

Bemerkung:

- (a) $P(X_{t+1} = j \mid X_t)$ in 2. ist ZV, mit Werten $P(X_{t+1} = j \mid X_t = i)$, $i \in S$.
- (b) Verbale Interpretation der Markov-Eigenschaft 1. bzw. 2. : Bei gegebener Gegenwart hängt die Zukunft des SP nicht von der Vergangenheit ab. Anders ausgedrückt:
Bei gegebener Gegenwart sind Zukunft und Vergangenheit bedingt unabhängig.

Definition 3.2 Übergangswahrscheinlichkeiten

$$p_{ij}(t) = P(X_{t+1} = j \mid X_t = i)$$

heißt (einschrittige) *Übergangswahrscheinlichkeit* (ÜW) von i nach j (zum Zeitpunkt t).

Im allgemeinen hängt $p_{ij}(t)$ von t tatsächlich ab. Die klassische Theorie behandelt jedoch weitgehend homogene MK.

Definition 3.3 Homogene MK

Eine MK heißt *homogen* bzw. besitzt *stationäre* ÜW $:\Leftrightarrow$

$$p_{ij}(t) = p_{ij} \quad \forall t \in \mathbb{N}_0.$$

Vereinbarung: Im weiteren MK homogen, falls nichts anderes vermerkt.

Bemerkungen:

- Eine inhomogene MK ist i.A. kein stationärer SP.
- Behandlung nichtstationärer ÜW z.B. so:

$$\text{logit } p_{ij}(t) = \log \left(\frac{P(X_{t+1} = j \mid X_t = i)}{1 - P(X_{t+1} = j \mid X_t = i)} \right) = f_{ij}(t)$$

⇒ Vorlesung Angewandte Stochastische Prozesse, Computerintensive Verfahren, Generalisierte Regression.

Definition 3.4 Übergangsmatrix ÜM

Für homogene MK heißt

$$P = (p_{ij}), \quad i, j \in S$$

Übergangsmatrix (ÜM).

Bemerkung: Bei inhomogenen MK ist die ÜM $P(t) = (p_{ij}(t))$ *zeitabhängig*.

Für $S = \{0, 1, 2, \dots\}$ ist also:

$$P = \begin{pmatrix} p_{00} & p_{01} & \cdots & p_{0j} & \cdots \\ \vdots & \vdots & \ddots & \vdots & \\ p_{i0} & p_{i1} & \cdots & p_{ij} & \cdots \\ \vdots & \vdots & & \vdots & \ddots \end{pmatrix}$$

P ist endlich, falls S endlich. Eigenschaften:

$$p_{ij} \geq 0 \quad \text{und} \quad \sum_{j \in S} p_{ij} = 1 \quad (\text{Zeilensumme} = 1)$$

Satz 3.1 Folgerungen aus der Definition einer MK

- 1.
- $\forall t, s \geq 1; i_0, \dots, i_s$

$$\begin{aligned} P(X_{t+s} = i_s, \dots, X_{t+1} = i_1 \mid X_t = i_0) &= P(X_s = i_s, \dots, X_1 = i_1 \mid X_0 = i_0) \\ &= p_{i_0 i_1} p_{i_1 i_2} \cdots p_{i_{s-1} i_s}. \end{aligned}$$

Ist zusätzlich eine Anfangsverteilung

$$p_i(0) := P(X_0 = i), \quad i \in S$$

$$\text{gegeben} \Rightarrow P(X_s = i_s, \dots, X_1 = i_1, X_0 = i_0) = p_i(0) p_{i_0 i_1} \cdots p_{i_{s-1} i_s}$$

- 2.
- $\forall t, s \geq 1; i_0, \dots, i_{t+s}$

$$\begin{aligned} P(X_{t+s} = i_{t+s}, \dots, X_{t+1} = i_{t+1} \mid X_t = i_t, \dots, X_0 = i_0) &= \\ P(X_{t+s} = i_{t+s}, \dots, X_{t+1} = i_{t+1} \mid X_t = i_t) & \end{aligned}$$

(Erweiterung der Markov-Eigenschaft auf zukünftigen Verlauf X_{t+1}, \dots, X_{t+s} .)

3. Weitere Verallgemeinerung:
- $\forall 0 \leq t_0 < t_1 < \dots < t_k$

$$P(X_{t_k} = i_k \mid X_{t_{k-1}} = i_{k-1}, \dots, X_{t_0} = i_0) = P(X_{t_k} = i_k \mid X_{t_{k-1}} = i_{k-1})$$

Beweis:

FKO S. 15, 16, 17.

Beweis zu 2.: Für $s = 2$; für $s \geq 3$ analog durch Induktion.Mit $V_{t-1} = \{X_{t-1} = i_{t-1}, \dots, X_0 = i_0\}$ gilt

$$\begin{aligned} &P\{X_{t+2} = i_{t+2}, X_{t+1} = i_{t+1} \mid X_t = i_t, V_{t-1}\} \\ &= P\{X_{t+2} = i_{t+2} \mid X_{t+1} = i_{t+1}, X_t = i_t, V_{t-1}\} \cdot P\{X_{t+1} = i_{t+1} \mid X_t = i_t, V_{t-1}\} \\ &= P\{X_{t+2} = i_{t+2} \mid X_{t+1} = i_{t+1}\} \cdot P\{X_{t+1} = i_{t+1} \mid X_t = i_t\} \\ &= p_{i_t i_{t+1}} p_{i_{t+1} i_{t+2}} \\ &= P\{X_{t+2} = i_{t+2}, X_{t+1} = i_{t+1} \mid X_t = i_t\}. \end{aligned}$$

□

Bemerkung:

Gelegentlich wird 3. zur Definition von MK benützt. Folgt hier aus der einfacheren Definition 3.1.

Bemerkung:

Damit lassen sich alle endlich-dimensionalen Verteilungen der MK berechnen,

 \Rightarrow Satz von Kolmogorov anwendbar,

⇒ Bei Vorgabe einer ÜM P und einer Startverteilung $p(0)$ existiert zugehöriger SP mit W-Raum. Zugleich lässt sich die Likelihood berechnen, vgl. Statistische Inferenz.

Mehrschrittige ÜW, Chapman-Kolmogorov-Gleichung

Definition 3.5 Mehrschrittige ÜW

$$p_{ij}^{(t)} := P(X_{t+s} = j \mid X_s = i) = P(X_t = j \mid X_0 = i), \quad t \in \mathbb{N}_0$$

heißt t -schrittige ÜW von i nach j .

Satz 3.2 Chapman-Kolmogorov-Gleichungen

1.

$$p_{ij}^{(t+s)} = \sum_{k \in S} p_{ik}^{(s)} p_{kj}^{(t)}$$

2. Weiter gilt

$$(p_{ij}^{(t)}) = P^t = \underbrace{P \cdot \dots \cdot P}_{t\text{-mal}}$$

3. Damit 1. in Matrixform:

$$P^{t+s} = P^s P^t.$$

Beweis:

zu 1:

$$\begin{aligned}
 p_{ij}^{(t+s)} &= P\{X_{t+s} = j \mid X_0 = i\} \\
 &= \sum_{k \in S} P\{X_{t+s} = j, X_s = k \mid X_0 = i\} \\
 &= \sum_{k \in S} P\{X_{t+s} = j \mid X_s = k, X_0 = i\} \cdot P\{X_s = k \mid X_0 = i\} \\
 &= \sum_{k \in S} P\{X_{t+s} = j \mid X_s = k\} \cdot P\{X_s = k \mid X_0 = i\} \\
 &= \sum_{k \in S} p_{ik}^{(s)} p_{kj}^{(t)}
 \end{aligned}$$

zu 2: Zunächst für $t = 1, s = 1 \Rightarrow$

$$\begin{aligned}
 p_{ij}^{(2)} &= \sum p_{ik} p_{kj} \\
 p_{ij}^{(2)} &= (P^2)_{ij}
 \end{aligned}$$

dann Induktion: $s = 1, t = 2, \dots$

□

Definition 3.6 Zustandswahrscheinlichkeit

$$p_i(t) := P(X_t = i), \quad i \in S,$$

heißen *Zustandswahrscheinlichkeiten*.

$$p(t) = (p_i(t), i \in S)$$

(Zeilenvektor) heißt *Zustandsverteilung*. Es gilt

$$p_j(t) = \sum_{i \in S} p_i(0) p_{ij}^{(t)} \quad \text{bzw.} \quad p(t) = p(0) P^t$$

Beweis:

Satz von der totalen Wahrscheinlichkeit

$$P(X_t = j) = \sum_{i \in S} P(X_t = j \mid X_0 = i) P(X_0 = i).$$



Darstellung von MK durch gerichtete Graphen

Beispiele:

(a) Irrfahrtmodelle FKO, S. 22

(b) Markov-Ketten auf einem DNA-Strang

DNA-Stränge sind gerichtete ($5' \rightarrow 3'$) Ketten X_t , $t = 1, 2, \dots$, mit

$$X_t \in \{A(\text{Adenin}), G(\text{Guanin}), C(\text{Cytosin}), T(\text{Thymin})\},$$

(angeordnet in Form einer Doppel-Helix).

Die Folge $\{X_t\}$ ist **nicht** unabhängig. Als (erste) Approximation wird eine homogene Markov-Kette mit Übergangsmatrix

$$P = \begin{array}{c} \\ A \\ C \\ G \\ T \end{array} \begin{array}{cccc} A & C & G & T \\ \left(\begin{array}{cccc} .32 & .18 & .23 & .27 \\ .37 & .23 & .05 & .35 \\ .30 & .21 & .25 & .24 \\ .23 & .19 & .25 & .33 \end{array} \right) \end{array}.$$

angenommen (vgl. Lange, K., 1997, Math. and Stat. Methods for Genetic Analysis, Springer, N.Y.)

Ein komplexes MK-Modell bezieht Restriktionsorte (restriction sites) auf der DNA mit ein. Restriktionsenzyme erkennen spezifische Sequenzen auf dem DNA-Strang und zerschneiden die DNA an dieser Stelle, z.B. Enzym AluI erkennt *AGCT*

↪ Neue MK mit Zuständen $\{A, C, G, T, AG, AGC, AGCT\}$

AG = Paar *A, G*; (*G* folgt auf *A*)

AGC analog

AGCT Restriktionsort.

Sobald *AGCT* erreicht wird, unterbricht die nächste Base das Muster. Die zugehörige MK hat ÜM

$$P = \begin{matrix} & & A & C & G & T & AG & AGC & AGCT \\ \begin{matrix} A \\ C \\ G \\ T \\ AG \\ AGC \\ AGCT \end{matrix} & \left(\begin{matrix} .32 & .18 & 0 & .27 & .23 & 0 & 0 \\ .37 & .23 & .05 & .35 & 0 & 0 & 0 \\ .30 & .21 & .25 & .24 & 0 & 0 & 0 \\ .23 & .19 & .25 & .33 & 0 & 0 & 0 \\ .30 & .0 & .25 & .24 & 0 & .21 & 0 \\ .37 & .23 & .05 & 0 & 0 & 0 & .35 \\ .23 & .19 & .25 & .33 & 0 & 0 & 0 \end{matrix} \right) \end{matrix}$$

Für Sequenzanalysen eignen sich MK nur bedingt. Sie sind jedoch wesentlicher Baustein in „Hidden Markov Modellen (3.6)“.

(c) Diskreter Erneuerungsprozess (FKO, S. 24)

Ein Gerät werde zu den diskreten Zeitpunkten $t = 0, 1, 2, \dots$ überprüft. Falls ein Fehler gefunden wird, wird es durch ein neues ersetzt, andernfalls bleibt es in Betrieb usw. Die Lebensdauer des k -ten Gerätes wird so zu einer Zufallsvariablen Z_k . Wir nehmen nun an, dass für diese Lebensdauern Z_k unabhängig von k gilt

$$P\{Z \geq i + 1 \mid Z \geq i\} = p_i, \quad i \geq 0. \tag{3.1}$$

Sei nun X_t das Alter des zu Beginn der t -ten Periode arbeitenden Gerätes. Mit Wahrscheinlichkeit p_i geht dann also gemäß 3.1 das Alter i in das Alter $i + 1$ über, mit Wahrscheinlichkeit $q_i = 1 - p_i$ muss das Gerät ersetzt werden, in der nächsten Periode gilt $X_{t+1} = 0$. Die Folge X_t bildet offensichtlich wieder eine MK mit der ÜM

$$\begin{pmatrix} q_0 & p_0 & & 0 \\ q_1 & 0 & p_1 & \\ q_2 & 0 & 0 & p_2 \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

Anschaulich ergibt sich der Übergangsgraph

(d) Wechsel von Arbeitslosigkeit in Arbeit und zurück.

$S = \{0, 1\}$, 0 = arbeitslos, 1 = in Arbeit; t = Dauer von Arbeitslosigkeit in Monaten.

$p_{01}(t)$, $t = 1, 2, \dots$, Wahrscheinlichkeit für Übergang in Zustand Arbeit.

I.a. inhomogen (und i.a. abhängig von persönlichen Merkmalen des Arbeitslosen)

Fraglich: überhaupt MK, d.h. gilt die Markoveigenschaft?

Analog: $p_{10}(t)$ ÜW von Arbeit nach Arbeitslosigkeit.

(e) Einfache Markenwahlmodelle (FKO, S. 25)

Bei einigen Beispielen ist die Markov-Eigenschaft zweifelhaft.

Naheliegende Erweiterung:

Definition 3.7 MK höherer Ordnung

Der SP $X = \{X_t, t \in \mathbb{N}_0\}$, S diskret, heißt MK p -ter Ordnung $:\Leftrightarrow$

$\forall i_0, i_1, \dots, i_{t+1}, t \geq p + 1$

$$\begin{aligned} P(X_{t+1} = i_{t+1} \mid X_t = i_t, \dots, X_{t-p+1} = i_{t-p+1}, \dots, X_0 = i_0) \\ = P(X_{t+1} = i_{t+1} \mid X_t = i_t, \dots, X_{t-p+1} = i_{t-p+1}) \end{aligned}$$

Bemerkung:

Modellierung z.B. durch konditionale Logit-Modelle

$$\log \left(\frac{P(X_{t+1} = i_{t+1} \mid X_t, \dots, X_{t-p+1})}{1 - P(\dots \mid \dots)} \right) = f_{i_{t+1}}(X_t, \dots, X_{t-p+1})$$

(vgl. Abschnitt 3.7.2)

3.2 Klassifizierung von Zuständen und Rückkehrverhalten

(Nur für homogene MK!)

Bei (wechselseitiger) Erreichbarkeit von Zuständen und asymptotischem ($t \rightarrow \infty$) Verhalten der MK zeigen sich unterschiedliche Typen von Zuständen.

Beispiel:

Sei X eine homogene MK mit der ÜM

$$P = \begin{pmatrix} 0.2 & 0.8 & 0 \\ 0.6 & 0.4 & 0 \\ 0.2 & 0.3 & 0.5 \end{pmatrix}.$$

Man berechnet

$$P^2 = \begin{pmatrix} 0.52 & 0.48 & 0 \\ 0.36 & 0.64 & 0 \\ 0.32 & 0.43 & 0.25 \end{pmatrix},$$

$$P^4 = \begin{pmatrix} 0.44332 & 0.5568 & 0 \\ 0.4176 & 0.5824 & 0 \\ 0.4012 & 0.5363 & 0.0625 \end{pmatrix},$$

$$P^{10} = \begin{pmatrix} 0.4286 & 0.5714 & 0 \\ 0.4285 & 0.5715 & 0 \\ 0.4281 & 0.5709 & 0.0010 \end{pmatrix}.$$

Die $p_{ij}^{(t)}$ mit $j = 1, 2$ streben anscheinend gegen feste, positive Werte, während $p_{33}^{(t)}$ gegen Null geht.

Definition 3.8 Erreichbarkeit

1. Der Zustand j heißt von i aus *erreichbar*, i.Z. $i \rightarrow j$: \Leftrightarrow

$$\exists t \in \mathbb{N}_0 \text{ mit } p_{ij}^{(t)} > 0$$

\Leftrightarrow Es muss ein Weg im MK-Graphen von i nach j führen.

direkte Kante $i \rightarrow j$ i.a. nicht notwendig!

2. Die Zustände i und j heißen *gegenseitig erreichbar*, i.Z. $i \leftrightarrow j$: \Leftrightarrow

$$i \rightarrow j \quad \text{und} \quad j \rightarrow i$$

Man kommt in endlich vielen Schritten mit positiver Wahrscheinlichkeit von i nach j und zurück.

Satz 3.3 Äquivalenzklassen

$i \leftrightarrow j$ ist eine Äquivalenzrelation, d.h. die Menge aller Zustände lässt sich zerlegen in Äquivalenzklassen wechselseitig erreichbarer Zustände.

Es existiert kein Weg von l zurück zu einem Zustand aus der Klasse.

Beweis:

Zu zeigen ist:

1. $i \leftrightarrow i$ (Reflexivität), gilt nach Definition.
2. $i \leftrightarrow j \Leftrightarrow j \leftrightarrow i$, (Symmetrie) gilt nach Definition.

3. $i \leftrightarrow j$ und $j \leftrightarrow k \Rightarrow i \leftrightarrow k$ (Transitivität): Wegen $i \leftrightarrow j$ und $j \leftrightarrow k$ existiert $t, s \in \mathbb{N}_0$ mit $p_{ij}^{(t)} > 0$, $p_{jk}^{(s)} > 0$. Nach der Chapman-Kolmogorov-Gleichung folgt

$$p_{ik}^{(t+s)} = \sum_{l \in S} p_{il}^{(t)} p_{lk}^{(s)} \geq p_{ij}^{(t)} p_{jk}^{(s)} > 0,$$

und damit $i \rightarrow k$. Analog zeigt man $k \rightarrow i$.

□

Definition 3.9 Klassifizierung gemäß Erreichbarkeit

1. $C \subset S$ heißt *abgeschlossen* $:\Leftrightarrow$ kein Zustand in $S \setminus C$ ist von C aus erreichbar.
 C heißt *offen* $:\Leftrightarrow C$ nicht abgeschlossen
2. Ein Zustand i heißt *absorbierend*: $\{i\}$ abgeschlossen $\Leftrightarrow p_{ii} = 1$
3. Ein abgeschlossenes C heißt *irreduzibel* $:\Leftrightarrow$ keine echte, nichtleere Teilmenge von C ist abgeschlossen

4. Eine MK heißt *irreduzibel* $:\Leftrightarrow S$ irreduzibel \Leftrightarrow alle Zustände gegenseitig erreichbar

Zusammenfassung:

S zerfällt in offene und abgeschlossene, irreduzible Teilmengen wechselseitig erreichbarer Zustände. Bei MK mit endlich vielen Zuständen existiert mindestens eine abgeschlossene irreduzible Klasse C .

Eine offene Klasse kann existieren, muss aber nicht.

Klassifizierung nach Rückkehrverhalten

Die MK sei (o.B.d.A.) für $t = 0$ in i . Es sei T_{ii} die zufällige Zeit bis zur ersten Rückkehr nach i und $E(T_{ii}|X_0 = i) = \mu_{ii}$ die erwartete Rückkehrzeit.

Definition 3.10 Rekurrenz/Transienz/Periode

Sei $i \in S$

1. i heißt *rekurrent* $:\Leftrightarrow P(T_{ii} < \infty | X_0 = i) = 1$
mit Wahrscheinlichkeit 1 Rückkehr in endlicher Zeit (d.h. Rückkehrwahrscheinlichkeit = 1.)
2. i heißt *transient* $:\Leftrightarrow P(T_{ii} < \infty | X_0 = i) < 1$,
mit positiver Wahrscheinlichkeit keine Rückkehr.
3. Bei rekurrenten Zuständen kann man unterscheiden
 i heißt *positiv-rekurrent* $:\Leftrightarrow \mu_{ii} < \infty$
 i *null-rekurrent* $:\Leftrightarrow \mu_{ii} = \infty$
4. i *periodisch* mit Periode d $:\Leftrightarrow d \geq 2$ größte natürliche Zahl mit

$$\sum_{n=1}^{\infty} P(T_{ii} = nd | X_0 = i) + P(T_{ii} = \infty | X_0 = i) = 1$$

d.h. Rückkehr nur zu Vielfachen von d möglich (Periodengitter).

i *aperiodisch* für $d = 1$ (bzw. $d = \infty$)

5. i *ergodisch* $:\Leftrightarrow i$ positiv-rekurrent und aperiodisch (wichtiger Spezialfall)

Bemerkung:

Bei MK mit *endlich vielen Zuständen* existieren nur positiv-rekurrente Zustände; keine Fallunterscheidung nötig. Weiterhin gilt für jede Äquivalenzklasse C :

C abgeschlossen \Leftrightarrow alle Zustände aus C rekurrent,

C offen \Leftrightarrow alle Zustände aus C transient.

Beispiel:

(a) Irrfahrtmodell mit $r = 0$: jeder Zustand i ist rekurrent für $p = \frac{1}{2}$ und transient für $p \neq \frac{1}{2}$.

(b) Erneuerungsprozess: Die Zustände des diskreten Erneuerungsprozesses sind in Abhängigkeit von den bedingten Überlebens- und Ausfallwahrscheinlichkeiten p_i, q_i einmal transient, einmal positiv-rekurrent oder null-rekurrent. Etwa sind sämtliche Zustände der MK transient für

$$p_i = 1 - \left(\frac{1}{2}\right)^i, \quad q_i = \left(\frac{1}{2}\right)^i, \quad i \geq 1, \quad p_0 = 1, \quad q_0 = 0,$$

positiv-rekurrent für

$$p_i = q_i = \frac{1}{2}, \quad i \geq 0,$$

null-rekurrent für

$$p_i = \frac{i}{i+1}, \quad q_i = \frac{1}{i+1}, \quad i \geq 1, \quad p_0 = 1, \quad q_0 = 0.$$

(c) Eine MK sei durch folgenden Erreichbarkeitsgraph gegeben:

Das Periodengitter für den Zustand 1 ist $\{4, 6, 8, \dots\}$, also $d = 2$. Man sieht, dass nicht alle Vielfachen $n \cdot 2$ zum Periodengitter gehören, in diesem Fall erst für $n \geq 2$.

Satz 3.4 Resultate

1. Rekurrenz, Transienz, Periodizität und Ergodizität sind Klasseneigenschaften, d.h. alle Zustände einer Klasse sind rekurrent, transient, periodisch oder ergodisch.

Bew.: Für endliche MK anschaulich klar; formaler Beweis in FKO S. 39/40.

2. i rekurrent und $i \rightarrow j \Rightarrow i \leftrightarrow j$ und auch j rekurrent.

Falls von j kein Weg zurück nach i existiert: keine Rückkehr nach i möglich, i wäre nicht rekurrent.

3. i rekurrent \Rightarrow es existiert eine irreduzible Klasse $C(i)$ von rekurrenten Zuständen mit $i \in C(i)$.

Satz 3.5 Kanonische Repräsentation

S lässt sich zerlegen in irreduzible Teilmengen C_1, C_2, \dots und eine Restmenge T transienter Zustände. Seien nach evtl. Umnummerierung P_1, P_2, \dots, Q die dazugehörigen Teilmatrizen der ÜM P , so gilt

Satz 3.6

1. C irreduzibel, endliche Klasse \Rightarrow alle Zustände positiv-rekurrent
2. C endliche Klasse \Rightarrow
 - C rekurrent $\Leftrightarrow C$ abgeschlossen
 - C transient $\Leftrightarrow C$ offen
 - also: Falls MK endlich: offen=transient und abgeschlossen=rekurrent
3. MK endlich \Rightarrow es existiert mindestens eine rekurrente Klasse.

Für MK mit *endlich vielen* Zuständen geht man demnach so vor: Man sucht zuerst (etwa mit Hilfe des Erreichbarkeitsgraphen) die irreduziblen Äquivalenzklassen. Diese sind alle positiv rekurrent, der Rest ist transient. Die Periode wird für jede Klasse festgestellt.

In vielen Anwendungen liegt eine ergodische MK vor.

Beispiel:

(a) Markov-Kette für DNA-Sequenzen mit $S = \{A, C, G, T\}$ ist ergodisch.

(b) Absorbierende Markovketten

Absorbierende MK bestehen aus einer endlichen Menge von absorbierenden Zuständen und einer Menge von transienten Zuständen. In der kanonischen Representation hat also P die Gestalt

$$P = \begin{pmatrix} I & 0 \\ L & Q \end{pmatrix}, \text{ mit } I = \begin{pmatrix} 1 & & \\ & \ddots & \\ & & 1 \end{pmatrix}.$$

tot, an. In der kanonischen Darstellung ist die ÜM

$$\begin{array}{ccccc}
 & S_0 & S_1 & S_2 & S_3 & S_4 \\
 \left(\begin{array}{ccccc}
 1 & 0 & & & \\
 0 & 1 & & 0 & \\
 p_{20} & p_{21} & p_{22} & 0 & 0 \\
 p_{30} & p_{31} & 0 & p_{33} & 0 \\
 p_{40} & 0 & p_{42} & p_{43} & p_{44}
 \end{array} \right) .
 \end{array}$$

Die Zustände S_0, S_1 sind absorbierend. Die Zustände S_2, S_3, S_4 sind unter der Annahme $p_{ii} < 1, i = 2, 3, 4$ transient. Sämtliche Zustände sind aperiodisch.

3.3 Das Grenzverhalten von homogenen MK

Fragestellung:

$$\begin{array}{ll}
 P^t = (p_{ij}^{(t)}) \rightarrow ? & \text{für } t \rightarrow \infty \\
 p(t) = (P(X_t = i), i \in S) \rightarrow ? & \text{für } t \rightarrow \infty
 \end{array}$$

Hier nur für *aperiodische* MK. Für periodische MK siehe FKO, Abschnitt 2.3.

Satz 3.7 Grenzwertsatz für die t -schrittigen ÜW $p_{ij}^{(t)}$

Sei j *transient* oder *null-rekurrent*. Dann ist

$$\forall i \in S \quad \lim_{t \rightarrow \infty} p_{ij}^{(t)} = 0. \tag{3.2}$$

Sei j *positiv-rekurrent* und *aperiodisch*. Dann gilt:

$$\lim_{t \rightarrow \infty} p_{jj}^{(t)} = \frac{1}{\mu_{jj}} \tag{3.3}$$

$$\lim_{t \rightarrow \infty} p_{ij}^{(t)} = \lim_{t \rightarrow \infty} p_{jj}^{(t)} \quad \forall i \text{ mit } i \leftrightarrow j \tag{3.4}$$

$$\lim_{t \rightarrow \infty} p_{ij}^{(t)} = f_{ij} \frac{1}{\mu_{jj}} =: l_{ij}^{(\infty)} \quad \forall i \in S \tag{3.5}$$

Dabei ist f_{ij} die Wahrscheinlichkeit in endlicher Zeit von i nach j zu gelangen. Für rekurrentes i aus der gleichen Klasse wie j ist $f_{ij} = 1$, d.h. wir erhalten (3.3). Für $i \in C$, C abgeschlossen, $j \notin C$, ist $f_{ij} = 0$. Für transientes i ist $f_{ij} = P(j \text{ wird von } i \text{ aus (irgendwann) erreicht})$.

In der kanonischen Repräsentation ergibt sich also folgende Blockgestalt:

$$P^\infty := \lim_{t \rightarrow \infty} P^t = \begin{pmatrix} P_1^\infty & & & \\ & P_2^\infty & & 0 \\ & 0 & & \\ & & L^\infty & 0 \end{pmatrix}$$

Die Elemente von P_i^∞ berechnen sich aus (3.3), (3.4), L^∞ aus (3.5).

Beweis:

Formale Beweise schwierig. Hier nur Plausibilitätserklärungen.

(3.2) Für transientes j plausibel. Für null-rekurrentes, aperiodisches j ergibt sich (3.2) aus (3.3), indem man dort $\mu_{jj} = \infty$ setzt.

(3.3) Da j rekurrent und aperiodisch ist, gilt $p_{jj}^{(t)} > 0$ für alle genügend großen t . Die mittlere Wiederkehrzeit zwischen zwei j -Besuchen beträgt μ_{jj} . Die mittlere „Trefferquote“ in einem Zeitintervall der Länge 1 ist also $\frac{1}{\mu_{jj}}$. Mit

$$A_t = \begin{cases} 1, & X_t = j \\ 0, & X_t \neq j \end{cases}$$

muss also gelten

$$E(A_t | X_0 = j) \approx \frac{1}{\mu_{jj}}.$$

Wegen $E(A_t | X_0 = j) = P(X_t = j | X_0 = j) = p_{jj}^{(t)}$ gilt daher für genügend große t

$$p_{jj}^{(t)} \approx \frac{1}{\mu_{jj}}.$$

(3.4), (3.5) Zuerst muss j von i aus erreicht werden. Dies geschieht mit Wahrscheinlichkeit f_{ij} ; dann weiter wie oben.

□

Bemerkung

Die Plausibilitätserklärung zu (3.3) liefert uns auch eine weitere Interpretation von $p_{ij}^{(\infty)} = \lim_{t \rightarrow \infty} p_{ij}^{(t)}$ für aperiodische, rekurrente $i \leftrightarrow j$: $p_{ij}^{(\infty)} = \frac{1}{\mu_{jj}}$ ist die relative Häufigkeit für $t \rightarrow \infty$ der Besuche in j , bezogen auf die Gesamtzahl aller Übergänge.

Satz 3.7 ist bislang nur von beschränktem Wert, da wir noch keine einfache Methode angegeben haben, wie die erwartete Wiederkehrzeit μ_{jj} zu berechnen ist. Abhilfe schafft

Satz 3.8 Grenzwertsatz für ergodische MK

Für eine irreduzible, aperiodische MK gilt:

Falls alle Zustände *positiv-rekurrent* sind, dann besitzt das lineare Gleichungssystem

$$\left. \begin{array}{l} \pi_j = \sum_{i \in S} \pi_i p_{ij}, \quad j \in S \\ \sum_{j \in S} \pi_j = 1 \end{array} \right\} \text{ bzw. mit dem Zeilenvektor } \pi = (\pi_j, j \in S) \left\{ \begin{array}{l} \pi = \pi P \\ \pi \mathbf{1} = 1, \end{array} \right. \quad (3.6)$$

eine eindeutige, strikt positive ($\pi_j > 0, \forall j \in S$) Lösung und es gilt

$$\pi_j = \lim_{t \rightarrow \infty} p_{ij}^{(t)} = \frac{1}{\mu_{jj}}, \quad (3.7)$$

sowie

$$\lim_{t \rightarrow \infty} p(t) = \lim_{t \rightarrow \infty} p(0)P^t = p(0)P^\infty = \pi \quad (3.8)$$

für jede beliebige Anfangsverteilung $p(0)$.

Umgekehrt gilt auch: Falls (3.6) eindeutig lösbar und strikt positiv lösbar ist, so sind alle Zustände positiv rekurrent.

Da endliche, irreduzible MK nur positiv-rekurrente Zustände enthalten, gilt noch die

Folgerung:

Für eine irreduzible, aperiodische MK mit endlich vielen Zuständen hat (3.6) eine eindeutige, strikt positive Lösung.

Beweis:

Anstelle eines exakten Beweises, für den auf die einschlägige Literatur verwiesen sei, bringen wir eine *Plausibilitätserklärung*.

Nach der Bemerkung stellt $\lim_{t \rightarrow \infty} p_{ij}^{(t)} = \frac{1}{\mu_{jj}}$ die Wahrscheinlichkeit der Besuche in j „auf lange Sicht“

dar, (3.6) sind dann „Bilanzgleichungen“ für jeden Zustand j .

Die Wahrscheinlichkeit $\pi_j = \lim_{t \rightarrow \infty} p_{ij}^{(t)}$ dafür, dass sich die MK in j befindet, setzt sich additiv zusammen aus den Wahrscheinlichkeiten π_i , $i \in S$, mit denen sich die MK in anderen Zuständen $i \in S$ befindet, gewichtet mit den Wahrscheinlichkeiten p_{ij} in einem Schritt von i nach j zu gelangen.

(3.8) folgt direkt aus (3.7). □

Zur Berechnung von π löst man sich zuerst $\tilde{\pi} = \tilde{\pi}P$ ohne Normierungsbedingung und normiert anschließend durch $\pi = \frac{\tilde{\pi}}{\tilde{\pi}\mathbb{1}}$. Falls S endlich ist, eliminiert man von den linear abhängigen Gleichungen zuerst die komplizierteste.

Definition 3.11 Stationäre Verteilung

Sei P die Übergangsmatrix einer Markovkette. Der Zeilenvektor π heißt stationäre Verteilung dieser Markovkette : \Leftrightarrow

$$\pi = \pi P \quad \text{und} \quad \pi \mathbb{1} = 1.$$

Beispiel: Stationäre Verteilung einer MK mit 2 Zuständen $\{0, 1\}$

$$P = \begin{bmatrix} 1 - p_0 & p_0 \\ p_1 & 1 - p_1 \end{bmatrix}$$

$$(\pi_0, \pi_1) \begin{bmatrix} 1 - p_0 & p_0 \\ p_1 & 1 - p_1 \end{bmatrix} \stackrel{!}{=} (\pi_0, \pi_1)$$

$$\Leftrightarrow \begin{cases} \pi_0(1 - p_0) + \pi_1 p_1 = \pi_0 & \Leftrightarrow \boxed{\pi_0 p_0 = \pi_1 p_1} \\ \pi_0 p_0 + \pi_1(1 - p_1) = \pi_1 & \Leftrightarrow \pi_0 p_0 = \pi_1 p_1 \end{cases}$$

Lösungsweg 1: $\tilde{\pi}_0, \tilde{\pi}_1$ unnormiert; keine Nebenbedingung

$$\tilde{\pi}_0 = p_1 \Rightarrow \tilde{\pi}_1 = p_0$$

$$\Leftrightarrow \boxed{\pi_0 = \frac{p_1}{p_0 + p_1}, \quad \pi_1 = \frac{p_0}{p_0 + p_1}}$$

Lösungsweg 2: $\tilde{\pi}_0 + \tilde{\pi}_1 = 1$, Nebenbedingung

$$\pi_1 = 1 - \pi_0 \Leftrightarrow \pi_0 + \pi_1 = 1$$

$$\pi_0(1 - p_0) + (1 - \pi_0)p_1 = \pi_0$$

$$-\pi_0 p_0 + p_1 - \pi_0 p_1 = 0$$

$$\pi_0(p_0 + p_1) = p_1,$$

$$\boxed{\pi_0 = \frac{p_1}{p_0 + p_1}, \quad \pi_1 = \frac{p_0}{p_0 + p_1}}$$

3.4 Instationäre und inhomogene MK

In Anwendungen liegen oft Zeitreihen oder Longitudinaldaten zu Y Ziel- (oder Response)variable, diskret bzw. kategorial vor

- $z = (z_1, \dots, z_p)'$ Vektor von Kovariablen
- $T = \mathbb{N}_0$, Ergebnisraum S_Y von Y diskret.
- Zeitreihendaten: $(Y_t, Z_t) \quad t = 1, \dots, T$, d.h. ein Pfad eines SP Y und eines Kovariablen-SP werden von 1 bis T beobachtet.

- Longitudinaldaten: Für $n = 1, \dots, N$ Individuen (Objekte) werden jeweils Pfade

$$\{(Y_{nt}, z_{nt}), t = 1, \dots, T\}$$

beobachtet.

Situationen:

- N groß, T klein
- N klein, T groß
- (N=1: Zeitreihe)
- N, T mittel bis groß

Datenstruktur der Pfade passt zu MK für Y.

Aber:

- Zeitliche Stationarität fraglich \Rightarrow MK mit nichtstationärer ÜW $p_{ij}(t)$
- Homogenität der Population fraglich,

$$\Rightarrow P(Y_{n,t+1} = j \mid Y_{n,t} = i) = p_{ij,n}(t)$$

von n abhängig!

Bei homogener Population würde man

$$p_{ij,n}(t) \equiv p_{ij}(t), \quad n = 1, \dots, N$$

annehmen.

- Auch Markov-Eigenschaft, zumindest 1. Ordnung fraglich.

Fragestellung: Wie modelliert man nichtstationäre inhomogene MK?

3.4.1 Instationäre und inhomogene binäre MK 1. Ordnung

Sei zunächst $\{Y_t, t \in \mathbb{N}_0\}$ eine binäre MK 1. Ordnung. Dann werden geeignete Modelle für die ÜW

$$p_{ij}(t) = P(Y_{t+1} = j \mid Y_t = i), \quad i, j \in \{0, 1\}$$

benötigt.

Liegen bei Longitudinaldaten $\{Y_{nt}, t \in \mathbb{N}_0\}$ zusätzlich Kovariablen $\{z_{nt}, t \in \mathbb{N}_0\}$ zu den Objekten $n = 1, \dots, N$ vor, so werden entsprechend Modelle für

$$p_{ij,n}(t) = P(Y_{n,t+1} = j \mid Y_{n,t} = i, z_{nt})$$

benötigt. Die Kovariablen können zeitabhängig oder auch zeitunabhängig sein, wie z.B. das Geschlecht oder die Art der Therapie für ein Individuum bei einer klinischen Studie.

Es reicht die ÜW für $j = 1$ zu modellieren, da $p_{i0,n}(t) + p_{i1,n}(t) = 1$ gilt. Basis für die Modellierung sind binäre Regressionsmodelle, insbesondere Logit- oder Probitmodelle (vgl. Lineare Modelle, Generalisierte Regression)

Separate Modellierung der ÜW

Hier werden $p_{01}(t)$ und $p_{11}(t)$ jeweils separat durch binäre Regressionsmodelle modelliert:

$$p_{01}(t) = h(w_t' \beta_0), \quad p_{11}(t) = h(z_t' \beta_1).$$

Dabei wird der Index n unterdrückt; h sind sog. Responsefunktionen $h : \mathbb{R} \rightarrow (0, 1)$. Für $h(\cdot) = \exp(\cdot) / \{1 + \exp(\cdot)\}$ bzw. $h(\cdot) = \Phi(\cdot)$ erhält man Logit- bzw. Probitmodelle. Die Design- bzw. Kovariablenvektoren w_t, z_t können auch identisch sein.

Im reinen Zeitreihenfall ($n=1$) wird durch $\eta_t = w_t' \beta_0$ bzw. $\eta_t = z_t' \beta_1$ z.B. ein zeitlicher Trend parametrisch spezifiziert, etwa mit $w_t' \beta_0 = \beta_{00} + \beta_{01}t$ ein linearer Trend, oder ein (stückweise) polynomialer Trend, usw. Für Longitudinaldaten enthalten w_t, z_t i.d.R. auch individuelle Kovariablen.

Logit-Modell:

$$p_{01}(t) = \frac{\exp(w_t' \beta_0)}{1 + \exp(w_t' \beta_0)} \Leftrightarrow \log \left(\frac{p_{01}(t)}{p_{00}(t)} \right) = w_t' \beta_0;$$

analog für $p_{11}(t)$.

Bemerkung:

Lässt man statt einer parametrischen Trendfunktion $w_t'\beta_0$ eine nichtparametrisch spezifizierte Funktion $f_0(t)$ (bzw. $f_1(t)$) zu, so führt dies zu non- und semiparametrischen Logit-Modellen (Fahrmeir, Tutz; ch. 5; Generalisierte Regression).

Konditionale (autoregressive) Modellierung

Alternativ modelliert man

$$P(Y_{t+1} = 1 \mid Y_t, w_t)$$

für $Y_t \in \{0, 1\}$ in einem simultanen Modellansatz durch ein autoregressives Modell, indem Y_t wie w_t als Regressor fungiert, z.B.

$$\log \left(\frac{P(Y_{t+1} = 1 \mid Y_t, w_t)}{P(Y_{t+1} = 0 \mid Y_t, w_t)} \right) = w_t'\beta + Y_t\alpha \quad (\text{additiv})$$

oder

$$\dots = w_t'\beta + Y_t w_t'\alpha \quad \text{mit Interaktion}$$

Einsetzen von $Y_t = 0$ bzw. 1 liefert dann den Zusammenhang zur separaten Modellierung.

3.4.2 Weitere inhomogene MK

MK höherer Ordnung

Eine homogene MK 2. Ordnung besitzt Übergangswahrscheinlichkeiten

$$P(Y_{t+1} = j \mid Y_t = i_1, Y_{t-1} = i_2) =: p_{i_1 i_2, j}(t).$$

Im binären Fall $Y_t \in \{0, 1\}$ genügt es wieder $p_{i_1 i_2 1}(t)$ zu spezifizieren.

In Analogie zu MK 1. Ordnung könnte man z.B. 4 separate logistische Ansätze machen, je einen für die Paare $(i_1, i_2) = (0, 0), (0, 1), (1, 0)$ oder $(1, 1)$. Mit einem Designvektor z_t ergibt dies

$$\log \left(\frac{p_{i_1 i_2 1}(t)}{p_{i_1 i_2 0}(t)} \right) = z_t'\beta_{i_1 i_2},$$

mit vier verschiedenen Parametervektoren $\beta_{00}, \beta_{01}, \beta_{10}$, und β_{11} . Alternativ kann wieder zu einer simultanen Logit-Autoregression

$$\log \left(\frac{P(Y_{t+1} = 1 \mid Y_t, Y_{t-1}, z_t)}{P(Y_{t+1} = 0 \mid Y_t, Y_{t-1}, z_t)} \right) = z_t'\beta + \alpha_1 Y_t + \alpha_2 Y_{t-1} \quad (\text{additiv})$$

oder

$$\dots = z_t'\beta + Y_t z_t'\alpha_1 + Y_{t-1} z_t'\alpha_2 + Y_t Y_{t-1} z_t'\alpha_3$$

übergangen werden. Das zweite Modell besitzt genauso viele Parameter wie die 4 separaten Modelle. Einsetzen der verschiedenen 0/1-Kombinationen für Y_t, Y_{t-1} liefert

$$\beta_{00} = \beta, \beta_{01} = \beta + \alpha_1, \beta_{10} = \beta + \alpha_2, \beta_{11} = \beta + \alpha_1 + \alpha_2 + \alpha_3.$$

Durch Nullsetzen von Komponenten in $\alpha_1, \alpha_2, \alpha_3$ erhält man parametersparsamere Submodelle. Analog kann man MK höherer Ordnung in autoregressiver Form modellieren, vgl. z.B. Fahrmeir/Tutz, ch. 6.

Ebenfalls in Analogie zum binären Fall lassen sich inhomogene MK durch multinominale autoregressive Regressionsmodelle behandeln, vgl. Fahrmeir/Tutz, ch. 3 und ch. 6.

3.5 Statistische Inferenz für MK

3.5.1 Likelihood-Inferenz für SP

Statistische Inferenz für SP beruht auf den gleichen Prinzipien wie die Inferenz für unabhängige und evtl. identische Zufallsvariablen, durch die die Daten generiert werden. Durch die größere Modellkomplexität und die inhärente Abhängigkeit der Variablen werden jedoch Durchführung und Theorie i.d.R. etwas schwieriger.

Im weiteren gehen wir auf parametrische Likelihood-Inferenz für MK ein. Generell können folgende Situationen vorliegen:

Beobachtung eines Pfades

Die Daten (x_0, x_1, \dots, x_T) werden aufgefasst als Realisierung

$$(x_0 = X_0(\omega), x_1 = X_1(\omega), \dots, x_t = X_t(\omega), \dots, x_T = X_T(\omega)) \text{ eines } SP\{X_t\}.$$

Bei MK handelt es sich also um eine beobachtete diskrete oder kategoriale Zeitreihe, die als bis T vorliegende Realisierung der MK aufgefasst wird.

Sei θ ein Vektor von unbekanntem Parametern. Dann ist die Likelihoodfunktion

$$L(\theta | X) := f(x_0, \dots, x_T | \theta)$$

die gemeinsame Dichte von X_0, \dots, X_T , ausgewertet für x_0, \dots, x_T , und als Funktion von θ betrachtet.

Eine generelle Möglichkeit die Likelihoodfunktion zu berechnen ist die Faktorisierung in das Produkt

bedingter Dichten:

$$\begin{aligned}
 L(\theta | X) &= f(x_T, x_{T-1}, \dots, x_0 | \theta) \\
 &= f_T(x_T | x_{T-1}, \dots, x_0; \theta) \cdot f(x_{T-1}, \dots, x_0 | \theta) \\
 &= \dots \\
 &= \prod_{t=1}^T f_t(x_t | x_{t-1}, \dots, x_0; \theta) f_0(x_0 | \theta).
 \end{aligned}$$

Für MK ist diese Faktorisierung die Methode der Wahl und vereinfacht sich wegen der Markov-Eigenschaft

$$f_t(x_t | x_{t-1}, \dots, x_0; \theta) = f_t(x_t | x_{t-1}; \theta)$$

zu

$$L(\theta | X) = \prod_{t=1}^T f_t(x_t | x_{t-1}; \theta) f_0(x_0 | \theta).$$

Bei homogenen MK kann weiter der Index t bei f_t entfallen.

Durchgeführt wird das ML-Prinzip: Wähle $\hat{\theta}_{ML}$ als Maximierer von $L(\theta | X)$ üblicherweise durch Übergang zur log-Likelihood

$$l(\theta | X) := \ln L(\theta | X) = \sum_{t=1}^T l_t(x_t | x_{t-1}; \theta) + l_0(x_0 | \theta)$$

mit den log-Likelihood Beiträgen $l_t(x_t | \cdot) = \ln f_t(x_t | \cdot)$.

Bemerkung:

Falls die $\{X_t\}$ unabhängig sind, ergibt sich

$$L(\theta | X) = f(x_0, \dots, x_T | \theta) = \prod_{t=1}^T f_t(x_t | \theta) f_0(x_0 | \theta)$$

wie üblich als Produkt der Randdichten.

Beobachtung mehrerer unabhängiger Pfade eines SP

Die Daten (x_{n0}, \dots, x_{nT}) , $n = 1, \dots, N$ werden aufgefasst als unabhängige Realisierungen

$$(x_{n0} = X_0(\omega_n), \dots, x_{nT} = X_T(\omega_n)), n = 1, \dots, N$$

des $SP\{X_t\}$. Dann ergibt sich die gemeinsame Dichte und die Likelihood als Produkt

$$L(\theta | X) = \prod_{n=1}^N f(x_{nT}, \dots, x_{n0} | \theta) = \prod_{n=1}^N \prod_{t=1}^T f_t(x_{nt} | x_{n,t-1}; \theta) f_0(x_{n0} | \theta)$$

der beteiligten Likelihoodfunktionen für die Realisierungen $n = 1, \dots, N$.

Longitudinaldaten mit Kovariablen

In diesem Fall liegen Beobachtungen zu einer Zielvariablen Y und einem Vektor z (oder x) von Kovariablen vor.

Die Daten werden aufgefasst als unabhängige Realisierungen von SP Y_n , $n = 1, \dots, N$, gegeben die Kovariablen, also:

$$((y_{nt}, z_{nt}), t = 0, \dots, T), \quad n = 1, \dots, N.$$

Die Likelihood für Objekt n ist gegeben durch

$$L_n(\theta | Y_n) = \prod_{t=1}^T f(y_{nt} | y_{n,t-1}, z_{nt}; \theta) f_0(y_{n0} | \theta).$$

Wegen der Unabhängigkeit der Beobachtungen zwischen Objekten ergibt sich die gesamte Likelihood zu

$$L(\theta | Y) = \prod_{n=1}^N L_n(\theta | Y_n).$$

3.5.2 Inferenz bei homogenen Markov-Ketten

Sei X eine homogene MK mit endlichem Zustandsraum $S = \{1, \dots, m\}$.

Unbekannter Parameter θ :

$$\begin{aligned} P = (p_{ij}) & \quad \text{Übergangsmatrix} \\ p(0) & \quad \text{Anfangsverteilung} \end{aligned}$$

Plausibel: ML-Schätzer für p_{ij} sind die entsprechenden Übergangshäufigkeiten, die an einem oder mehreren Pfaden von X beobachtet werden. Zunächst:

1 Pfad beobachtet

$$X_0(\omega) = i_0, X_1(\omega) = i_1, \dots, X_T(\omega) = i_T$$

Likelihoodfunktion

$$\begin{aligned} L(p(0), P) &= P\{X_0 = i_0, \dots, X_T = i_T\} \\ &= p_{i_0}(0) p_{i_0 i_1} \cdot \dots \cdot p_{i_{T-1} i_T}. \end{aligned}$$

Problem: Es gibt nur eine Beobachtung zur Schätzung von $p(0)$; $\Rightarrow \hat{p}_{i_0}(0) = 1, \hat{p}_i(0) = 0$ sonst.

Deshalb: Betrachte nur die *bedingte* Likelihood

$$L(P) = \prod_{i,j} p_{ij}^{n_{ij}},$$

mit n_{ij} = Anzahl der beobachteten Übergänge von i nach j .

Logarithmieren ergibt

$$\log L(P) = \sum_{i,j} n_{ij} \log(p_{ij}).$$

ML-Schätzwerte \hat{p}_{ij} erhält man durch Maximieren von $\log L(P)$ unter der Nebenbedingung, dass die Zeilensummen =1 sind. Nach der Lagrange-Methode ist dies äquivalent zur Maximierung von

$$l^* = \sum_{i,j} n_{ij} \log p_{ij} - \sum_i \lambda_i \left(\sum_{j \in S} p_{ij} - 1 \right)$$

ohne Nebenbedingungen. Nullsetzen der partiellen Ableitungen liefert

$$\frac{\partial l^*}{\partial p_{ij}} = \frac{n_{ij}}{p_{ij}} - \lambda_i = 0,$$

also

$$n_{ij} = \lambda_i p_{ij}.$$

Summation über j liefert wegen $\sum_j p_{ij} = 1$

$$\lambda_i = \sum_{j \in S} n_{ij} = n_i,$$

und daraus schließlich

$$\hat{p}_{ij} = \frac{n_{ij}}{n_i}.$$

Eigenschaften des ML-Schätzers

Bei ergodischen MK gilt für $T \rightarrow \infty$

$$\hat{p}_{ij} \stackrel{a}{\sim} N \left(p_{ij}, \frac{\hat{p}_{ij}(1 - \hat{p}_{ij})}{n_i} \right)$$

Mehrere Pfade beobachtet

Im *homogenen* Fall lassen sich die ÜW wie bei einem Pfad schätzen.

Prinzipiell möglich ist nun auch die Schätzung von ÜW $p_{ij}(t)$ im *inhomogenen* Fall:

$$\hat{p}_{ij}(t) = \frac{n_{ij}(t)}{n_i(t)}$$

mit

$$\begin{aligned} n_{ij}(t) & \text{ Anzahl der Übergänge von } i \text{ nach } j \text{ mit } X_t = i \text{ und } X_{t+1} = j \\ n_i(t) & \text{ Anzahl der Beobachtungen mit } X_t = i \text{ und } X_{t+1} \text{ beliebig} \end{aligned}$$

Problem: $p_{ij}(t)$ wird unrestringiert, d.h. ohne weitere Struktur-/Modellannahmen geschätzt.

⇒ hohe Varianz von $\hat{p}_{ij}(t)$, falls $n_i(t)$ klein.

Alternative: Parametrische oder nonparametrische Modellierung von $p_{ij}(t)$, z.B. durch Logitmodelle (vgl. 3.4), und ML-Schätzung wie in 3.5.1.

Likelihood-Quotienten-Tests für (homogene) MK

Zum Test von Hypothesen H_0 versus H_1 , z.B.

$$\begin{aligned} H_0 : \{X_t\} \text{ ist i.i.d.,} & & H_1 : \{X_t\} \text{ ist MK 1. Ordnung} \\ H_0 : \{X_t\} \text{ ist MK 1. Ordnung,} & & H_1 : \{X_t\} \text{ ist MK 2. Ordnung} \end{aligned}$$

etc., kann der Likelihood-Quotienten-Test verwendet werden:

$$\begin{aligned} l(\hat{\theta}_1) & \text{ maximierte Loglikelihood im } H_1\text{-Modell} \\ l(\hat{\theta}_0) & \text{ maximierte Loglikelihood im } H_0\text{-Modell} \end{aligned}$$

Dann gilt unter H_0 :

$$2(l(\hat{\theta}_1) - l(\hat{\theta}_0)) \stackrel{a}{\sim} \chi^2(r), \quad T \rightarrow \infty$$

mit r =Differenz der Anzahl der geschätzten Parameter in H_0 -und H_1 -Modell.

3.5.3 Fallstudie: Niederschlag bei den Snoqualmie-Wasserfällen, Cascade Mountains

Aus: Guttorp „Stochastic Modelling of Scientific Data“.

Daten: 36 Jahre $j = 1948 - 1983$; jeweils Januar $t = 1, \dots, 31$.

$$X_{tj} = \begin{cases} 1 & \text{Regen am Tag } t \text{ in Jahr } j \\ 0 & \text{kein Regen am Tag } t \text{ im Jahr } j. \end{cases}$$

Annahme: Unabhängige Pfade $j = 1, \dots, 36$ desselben SP X werden beobachtet.

Modell 1: $\{X_{tj}\}$ sind i.i.d.

rel. Häufigkeiten für $X_t = 0$: $\hat{p} = \frac{325}{791+325} = 0.291$

rel. Häufigkeiten für $X_t = 1$: $1 - \hat{p} = 0.709$

Likelihood: $L(p) \propto p^{325}(1 - p)^{791}$

Modell 2: X_t ist homogene MK 1. Ordnung mit (jahres-unabhängiger) ÜM P

Kontingenztafel:

	$X_t = 0$	$X_t = 1$	
$X_{t-1} = 0$	186	123	309
$X_{t-1} = 1$	128	643	771
	314	766	1080

Beachte: Hier gehen nur $30 \cdot 36 = 1080$ Beobachtungen ein.

Geschätzte Übergangsmatrix (z.B. $\hat{p}_{01} = \frac{123}{309} = 0.398$):

$$\hat{P} = \begin{pmatrix} 0.602 & 0.398 \\ 0.166 & 0.834 \end{pmatrix}$$

Likelihood:

$$L(p(0), P) = \left(\prod_{j=1}^{36} P(X_{0j} = x_{0j}) \right) p_{00}^{186} p_{01}^{123} p_{10}^{128} p_{11}^{643}$$

LQ-Test zu

$$H_0 : \{X_t\} \text{ i.i.d} \quad \textit{versus} \quad H_1 : \{X_t\} \text{ MK 1. Ordnung}$$

Beachte: „gleiche“ Daten werden benötigt;

deshalb müssen bei Modell 1 die Startwerte ($j = 1, \dots, 36$) weggelassen werden.

$$\Rightarrow \hat{p} = \frac{314}{766+314} = 0.291$$

LQ-Statistik ($\overset{a}{\sim} \chi^2(1)$):

$$\begin{aligned} 2\{\log(L(\hat{P})) - \log(\hat{p})\} &= 2\{[186 \cdot \log(0.602) + 123 \cdot \log(0.398) + 128 \cdot \log(0.166) + 643 \cdot \log(0.834)] \\ &\quad - [314 \cdot \log(0.291) + 766 \cdot \log(0.709)]\} \\ &= 193.49 \end{aligned}$$

\Rightarrow **i.i.d. Modell 1 ablehnen**

Modell 3: Homogene MK 2. Ordnung

Kontingenztafel:

		$X_t = 0$	$X_t = 1$	
$X_{t-2} = 0$	$X_{t-1} = 0$	109	67	176
$X_{t-2} = 0$	$X_{t-1} = 1$	25	94	119
$X_{t-2} = 1$	$X_{t-1} = 0$	70	52	122
$X_{t-2} = 1$	$X_{t-1} = 1$	100	527	627
		304	740	1044

\Rightarrow geschätzte ÜM für MK 2. Ordnung

$$\hat{P}_2 = \hat{p}_{ij,k} = \begin{pmatrix} 0.619 & 0.381 \\ 0.210 & 0.790 \\ 0.574 & 0.426 \\ 0.159 & 0.841 \end{pmatrix}$$

LQ-Test zu

$$H_0 : \{X_t\} \text{ MK 1. Ordnung} \quad \textit{versus} \quad H_1 : \{X_t\} \text{ MK 2. Ordnung}$$

$$\begin{aligned}\log(L(\hat{P}_2)) &= 109 \cdot \log(0.619) + \dots + 527 \cdot \log(0.841) \\ &= -536.48\end{aligned}$$

$$\log(L(\hat{P})) = -537.67$$

$\Rightarrow 2\{\log(L(\hat{P}_2)) - \log(L(\hat{P}))\} = 2.37$ bei 2 Freiheitsgraden

$\Rightarrow p$ -Wert 0.15, $\Rightarrow H_0$ wird nicht abgelehnt.

3.6 Hidden Markov Modelle

Kritik am Markov-Modell für Niederschlag:

- Das Markov-Modell impliziert eine (geometrische) Verteilung für die Dauer von trockenen/ nassen Perioden, die sich nicht mit der Realität deckt \rightarrow kürzere Perioden werden vom Modell vorhergesagt als beobachtet wurde.

- Schwierig zu Verallgemeinern, um *mehrere* Wetterstationen zu modellieren.

Daher: **Hidden Markov Modelle**

Latenter Zustand X_t (Klima) folgt einer Markovkette. Beobachtungen $Y_t \mid X_t$ sind bedingt unabhängig, gegeben den Zustand X_t .

HMM bilden eine flexible Modellklasse, insbesondere hat Y_t marginal i.A. nicht die Markov-Eigenschaft.

Beispiel: Regenfall-Daten

X_t : MK mit Übergangswahrscheinlichkeit $P = \begin{pmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{pmatrix}$

zwei Zustände:

$$X_t = \begin{cases} 0 & \text{gutes Wetter} \\ 1 & \text{schlechtes Wetter} \end{cases}$$

$$Y_t = \begin{cases} 1 & \text{Regen} \\ 0 & \text{kein Regen} \end{cases}$$

mit

$$P(Y_t = 1 \mid X_t = 1) = q_{11}$$

$$P(Y_t = 1 \mid X_t = 0) = q_{01}$$

$$P(Y_t = 0 \mid X_t = 1) = q_{10}$$

$$P(Y_t = 0 \mid X_t = 0) = q_{00}$$

Hidden Markov-Modelle: Das gelegentlich unehrliche Casino

Es wird mit 2 Würfeln gespielt:

Würfel 1 fair; Würfel 2 unfair, mit $P(Y = 6) = 0.5$, $P(Y = 1) = \dots = P(Y = 5) = 0.1$

Casino wechselt von fairem zu unfairem Würfel mit Wahrscheinlichkeit 0.05, und von unfairem zu fairem Würfel mit Wahrscheinlichkeit 0.1. Der Wechsel zwischen den Würfeln ist dann MK mit ÜM

$$P = \begin{matrix} & f & u \\ \begin{matrix} f \\ u \end{matrix} & \begin{pmatrix} 0.95 & 0,05 \\ 0.1 & 0.9 \end{pmatrix} \end{matrix}$$

ÜW $P(X_t = u \mid X_{t-1} = f) = 0,05$

$P(X_t = f \mid X_{t-1} = u) = 0.1$

usw.

Insgesamt ein **HMM**: (t indiziert Spiel t)

1 : 1/6
2 : 1/6
3 : 1/6
4 : 1/6
5 : 1/6
6 : 1/6
Fair

1 : 1/10
2 : 1/10
3 : 1/10
4 : 1/10
5 : 1/10
6 : 1/2
Unfair

$$P(Y_t = j \mid X_t = f) \quad j = 1, \dots, 6$$

$$P(Y_t = j \mid X_t = u)$$

Beispiel: DNA-Sequenzen mit CpG Inseln

Im menschlichen Genom wird bei Auftreten eines Paares CG (innerhalb **eines** Stranges) (Notation: CpG) in der DNA-Sequenz des Nucleotid C mit (relativ) hoher Wahrscheinlichkeit zu T mutiert (durch chemische Prozesse).

⇒ CpG Paare treten seltener auf, als wenn man C; G als unabhängig ansieht.

Es gibt aber Abschnitte im Genom, wo diese chemische Mutation unterdrückt wird, z.B. am Beginn von Genen: „CpG-Insel“

DNA in Abschnitte unterteilt:

$$Y_t \in \{A, C, G, T\}, \quad X_t \in \{\text{Insel, nicht Insel}\}$$

X HMM

Daten: $\{Y_t, \dots\}$ beobachteter Pfad

Gesucht: $\{X_t, \dots\}$ latenter Pfad, enthält Inseln.

Inferenzproblem in HMM bei Daten $\{Y_t, t = 1, \dots, T\}$:

- Schätze ÜM P der latenten MK X
- Schätze bedingte Wahrscheinlichkeiten $P(Y_t = j \mid X_t = k)$
- Schätze Pfad $\{X_t \mid t = 1, \dots, T\}$ der verborgenen MK („Viterbi-Algorithmus“).

3.7 Allgemeine Markov-Ketten und MCMC-Methoden

Hauptziel dieses Abschnitts ist eine kurze Einführung in Markov Chain Monte Carlo (MCMC)-Methoden. Diese erlauben das Ziehen von Zufallszahlen aus komplizierten, analytisch und numerisch

unzugänglichen Verteilungen. MCMC-Verfahren werden insbesondere in der modernen, computerintensiven Bayesinferenz eingesetzt.

Vorteile im Vergleich zu herkömmlichen Verfahren zur Ziehung von Zufallszahlen sind:

- Die interessierende Dichte muss nur bis auf eine Normierungskonstante bekannt sein.
- Es können auch Zufallszahlen aus extrem hochdimensionalen Dichten, wie sie etwa bei komplexer Bayes-Modellierung auftreten, gezogen werden.

Anstelle aus der Dichte selbst zu ziehen, wird dabei eine Markovkette generiert, die diese Dichte als stationäre Verteilung besitzt. Da der Zustandsraum der MK dabei dem (oft stetigen) Parameterraum des statistischen Modells entspricht, ist es nötig einige Grundlagen, insbesondere Grenzwertsätze und dafür notwendige Begriffe, auf MK mit stetigem bzw. allgemeinem Zustandsraum S zu erweitern.

3.7.1 Allgemeine Markovketten

Im Folgenden ist S ein allgemeiner Zustandsraum. Insbesondere kann S ein stetiger Teilraum von \mathbb{R}^p sein.

Definition 3.12 Markovkette

Eine MK ist ein SP $\{X_t, t \in \mathbb{N}_0\}$ mit Zustandsraum S , der die Markov-Eigenschaft

$$P(X_{t+1} \in A \mid X_t = x, X_{t-1} \in A_{t-1}, \dots, X_0 \in A_0) = P(X_{t+1} \in A \mid X_t = x) \quad (3.9)$$

für beliebige $x \in S$ und $A, A_{t-1}, \dots, A_0 \in \mathcal{S}$ erfüllt.

Die MK heißt homogen, falls die Wahrscheinlichkeiten in (3.9) nicht von t abhängen.

$$P(x, A) := P(X_{t+1} \in A \mid X_t = x) = P(X_1 \in A \mid X_0 = x)$$

heißt Übergangskern der homogenen MK.

Bemerkung:

(a) Für MK mit diskretem Zustandsraum ist $P(x, A)$ durch die ÜW bzw. ÜM

$$p_{ij} = P(X_{t+1} = j \mid X_t = i)$$

bestimmt.

(b) Für $S = \mathbb{R}$ ist $P(x, A)$ durch eine Übergangsdichte $p(x, y)$ mit

$$P(x, A) = \int_A p(x, y) dy$$

bestimmt.

Beispiele für Markov-Ketten sind Irrfahrtmodelle und autoregressive Prozesse:

- Einfache Irrfahrt (Random Walk 1. Ordnung):

$$X_{t+1} = X_t + \epsilon_t, \quad \epsilon_t \text{ iid } N(0, \sigma^2), \quad X_0 = 0$$

Hier gilt

$$P(x, A) = P(X_{t+1} \in A \mid X_t = x) = \int_A \phi(y \mid x, \sigma^2) dy$$

mit

$$\phi(y \mid x, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2} \frac{(y-x)^2}{\sigma^2}\right) = p(x, y).$$

- $AR(1)$ -Prozess:

$$X_{t+1} = aX_t + \epsilon_t, \quad \epsilon_t \text{ wie oben;}$$

$$p(x, y) = \phi(y \mid ax, \sigma^2).$$

- Analog lassen sich mehrdimensionale Irrfahrten bzw. AR -Prozesse definieren; $S = \mathbb{R}^p$:

$$X_{t+1} = AX_t + \epsilon_t,$$

A $p \times p$ -Matrix, ϵ_t iid $N_p(0, \Sigma)$.

- State-space-Modelle: Hier wird X_t nur indirekt beobachtet mit

$$Y_t = Z_t X_t + U_t, \quad X_{t+1} = AX_t + \epsilon_t,$$

Z_t bekannte Matrix, U_t iid $N_p(0, \Sigma_u)$. Die MK $\{X_t\}$ ist latent oder verborgen (hidden).

Der t -schrittigen ÜM $(p_{ij}^{(t)}) = P^t$ entspricht die t -te Iteration des Übergangskerns

Es gilt

$$P^t(x, A) = P(X_t \in A \mid X_0 = x)$$

$$P^t = PP^{(t-1)}, \quad P^1 = P.$$

=Wahrscheinlichkeit, dass die Kette nach t Schritten einen Zustand in A erreicht, ausgehend vom Startwert x .

Definition 3.13 Invariante Verteilung

Eine Verteilung π^* auf (S, \mathcal{S}) mit Dichte π (bezüglich des dominierenden Maßes μ) heißt invariante Verteilung für den Übergangskern $P(x, A) : \Leftrightarrow$

Für alle $A \in \mathcal{S}$ gilt

$$\pi^*(A) = \int P(x, A)\pi(x)dx. \quad (3.10)$$

Bemerkung:

Für MK mit diskretem S entspricht (3.10) der Gleichung $\pi = \pi P$, mit der ÜM P . Für $S = \mathbb{R}$ und mit der zu $P(x, A)$ gehörenden Übergangsdichte $p(x, y)$ gilt (unter Regularitätsannahmen)

$$\pi(y) = \int p(x, y)\pi(x)dx.$$

Definition 3.14 Irreduzible Markovkette

Eine MK mit invarianter Verteilung π^* heißt irreduzibel, wenn sie positive Wahrscheinlichkeiten besitzt von einem beliebigen Startpunkt x_0 aus jede Menge A zu erreichen, für die $\pi^*(A) > 0$ gilt.

Definition 3.15 (A-)periodische MK

Eine MK heißt periodisch, wenn sie Teile des Zustandsraums nur in einer bestimmten Reihenfolge erreichen kann; andernfalls heißt die MK aperiodisch.

Es gilt

Satz 3.9 Grenzwertsatz

Sei $\{X_t, t \in \mathbb{N}_0\}$ eine irreduzible, aperiodische MK mit Übergangskern P und invarianter Verteilung π^* . Dann ist π^* eindeutig bestimmt und es gilt

$$\|P^t(x, \cdot) - \pi^*\| \rightarrow 0 \quad \text{für } t \rightarrow \infty,$$

wobei

$$\|P^t(x, \cdot) - \pi^*\| = 2 \sup_{A \in \mathcal{S}} |P^t(x, A) - \pi^*(A)|.$$

3.7.2 Metropolis-Hastings-Algorithmus und MCMC

Bisher war die ÜM P bzw. der Übergangskern P gegeben. Daraus kann die (eindeutige) invariante stationäre Verteilung π bzw. π^* bestimmt werden. Jetzt betrachtet man das umgekehrte Problem: Finde zu gegebener Verteilung π eine MK mit Übergangskern P und π als invarianter Verteilung.

Offensichtlich ist P nicht eindeutig: Für

(Matrizen)

erhält man jeweils $\pi = (3/5, 2/5)$ als invariante Verteilung. Also: Viel Freiraum in der Wahl von P .

Anwendung in Bayesianischer Inferenz

Bezeichne $D = (d_1, \dots, d_n)$ Daten, x unbekannte Parameter in der Likelihood

$$L(D | x) = f(D | x),$$

f die gemeinsame Dichte der Beobachtungen $D = (d_1, \dots, d_n)$, und $p(x)$ die priori-Verteilung der Parameter x . Dann ist nach dem Satz von Bayes

$$\pi(x) := p(x | D) = \frac{L(D | x)p(x)}{\int L(D | x)p(x)dx} = \frac{L(D | x)p(x)}{L(D)}$$

die posteriori Dichte der unbekannt Parameter. Problem: $\pi(x)$ ist wegen der Integration im Nenner oft nicht analytisch zugänglich. Dagegen ist der Zähler bekannt, d.h.

$$\pi(x) \propto L(D | x)p(x),$$

bis auf die Normierungskonstante $L(D)$ zugänglich. Mit MCMC können damit auch für hochdimensionale x aus $\pi(x)$ Zufallszahlen $x^{(t)}$, $t = 1, 2, \dots, N$ gezogen werden. Momente der posteriori-Verteilung und andere interessierende Größen werden dann durch entsprechende Analoga der empirischen Verteilung approximiert, z.B.

$$E_{\pi}(g) = \int g(x)\pi(dx) \approx \frac{1}{N} \sum_{t=1}^N g(x^{(t)}),$$

für geeignete Funktionen g , oder

$$\pi(A) \approx \frac{1}{N} \sum_{t=1}^N I_{\{x^{(t)} \in A\}}.$$

3.8 Der Metropolis-Hastings-Algorithmus

Wir konstruieren im Folgenden eine Markovkette mit Übergangskern P , so dass P^n gegen π^* konvergiert. Dann definieren wir den Übergangskern $P(x, A)$ wie folgt:

Sei $q(x, y)$ eine sogenannte Vorschlagsdichte, mit der ein neuer Zustand Y der Markovkette vorgeschlagen wird.

Sei weiterhin

$$\alpha(x, y) = \min \left\{ \frac{\pi(y)q(y, x)}{\pi(x)q(x, y)}, 1 \right\}$$

die Akzeptanzwahrscheinlichkeit mit der Y als neuer Zustand akzeptiert wird.

Definiere

$$p(x, y) = \begin{cases} q(x, y)\alpha(x, y) & x \neq y \\ 0 & \text{sonst} \end{cases}$$

und

$$r(x) = 1 - \int p(x, y)dy$$

= Wahrscheinlichkeit, dass der alte Zustand beibehalten wird.

Damit definieren wir den Übergangskern

$$P(x, A) = \int_A p(x, y)dy + r(x)\delta_x(A)$$

mit

$$\delta_x(A) = 1 \Leftrightarrow x \in A$$

Kann man zeigen, dass eine Markovkette mit oben definiertem Übergangskern π^* als invariante Verteilung besitzt, so konvergieren Iterationen des Kerns gegen π^* , falls die Kette zusätzlich irreduzibel und aperiodisch ist.

Damit erhalten wir den Metropolis-Hastings-Algorithmus:

- i) Wähle Startwerte x_0 und die Länge N der Kette. Setze $t=1$
- ii) Ziehe eine Zufallszahl $Y = y$ aus $q(x_{t-1}, Y)$ und akzeptiere diese als neuen Zustand x_t mit Wahrscheinlichkeit $\alpha(x_{t-1}, y)$, andernfalls setze $x_t = x_{t-1}$
- iii) Falls $t = N$ beende den Algorithmus, ansonsten setze $t = t + 1$ und fahre fort mit ii).

Beweis der Invarianzeigenschaft:

Wir zeigen dass für

$$P(x, A) = \int_A p(x, y)dy + r(x)\delta_x(A)$$

die Invarianzeigenschaft

$$\pi^*(A) = \int P(x, A)\pi(x)dx$$

gilt.

Zunächst gilt

$$\begin{aligned} \pi(x)p(x, y) &= \pi(x)q(x, y)\alpha(x, y) \\ &= \pi(x)q(x, y) \min \left\{ \frac{\pi(y)q(y, x)}{\pi(x)q(x, y)}, 1 \right\} \\ &= \min \{ \pi(y)q(y, x), \pi(x)q(x, y) \} \\ &= \pi(y)q(y, x) \min \left\{ 1, \frac{\pi(x)q(x, y)}{\pi(y)q(y, x)} \right\} \\ &= \pi(y)p(y, x) \end{aligned}$$

Die Beziehung $\pi(x)p(x, y) = \pi(y)p(y, x)$ heißt Umkehrbedingung. Sie ist hinreichend dafür, dass π bzw. π^* invariant ist. Weiter folgt

$$\begin{aligned}
 \int P(x, A)\pi(x)dx &= \int \left(\int_A p(x, y)dy \right) \pi(x)dx + \int r(x)\delta_x(A)\pi(x)dx \\
 &= \int_A \left(\int p(x, y)\pi(x)dx \right) dy + \int_A r(x)\pi(x)dx \\
 &= \int_A \left(\int p(y, x)\pi(y)dx \right) dy + \int_A r(x)\pi(x)dx \\
 &= \int_A (1 - r(y))\pi(y)dy + \int_A r(x)\pi(x)dx \\
 &= \int_A \pi(y)dy - \int_A r(y)\pi(y)dy + \int_A r(x)\pi(x)dx \\
 &= \int_A \pi(y)dy = \pi^*(A)
 \end{aligned}$$

Bemerkungen:

- Nach einer gewissen Konvergenzphase (sog. „burn in“) können die gezogenen Zufallszahlen X_0, X_1, \dots, X_N als Realisierungen aus π^* angesehen werden. Eigenschaften von π^* kann man nun mit Hilfe der Zufallszahlen schätzen, z.B. den Erwartungswert durch \bar{X} , wobei üblicherweise der burn in unberücksichtigt bleibt.
- Falls $q(x, y)$ symmetrisch ist, d.h.

$$q(x, y) = q(y, x)$$

vereinfacht sich die Akzeptanzwahrscheinlichkeit zu

$$\alpha = \min \left\{ \frac{\pi(y)}{\pi(x)}, 1 \right\}$$

In diesem Fall spricht man vom sog. Metropolis-Algorithmus.

- Die Vorschlagsdichte $q(x, y)$ sollte so gewählt werden, dass man leicht Zufallszahlen aus ihr ziehen kann und die Akzeptanzraten nicht zu klein werden. (z.B. multivariate Normalverteilung).

Beispiele für die Konstruktion von Vorschlagsdichten

- Unabhängige Ketten

$$q(x, y) = q(y)$$

- Random Walk Kette

Ziehe eine Zufallszahl Z aus einer Dichte $f(z)$ und setze $Y = X_{n-1} + Z$ d.h.

$$q(X_{n-1}, Y) = f(Y - X_{n-1})$$

- Ist $x = (x_1, \dots, x_p)$ multivariat mit Skalaren oder Teilvektoren x_1, \dots, x_p , so kann man den MH-Algorithmus auch komponentenweise anwenden, d.h. erst x_1 bei festgehaltenen x_2, \dots, x_p aufdatieren, dann x_2, \dots , dann x_p bei festgehaltenen x_1, \dots, x_{p-1} . Dabei vereinfacht sich $\pi(y)/\pi(x^{(t)})$ zu

$$\frac{\pi(y_i | x_{-i}^{(t)})}{\pi(x_i^{(t)} | x_{-i}^{(t)})}$$

mit $\pi(y_i | x_{-i})$ als bedingter Dichte gegeben alle Komponenten x_{-i} von x ohne die i -te. Sind diese bedingten Dichten bekannt, d.h. kann man direkt aus ihnen ziehen, so wählt man sie als Vorschlagsdichte $q(y, x) = q(y)$. Dies ergibt $\alpha(x, y) = 1$, d.h. man erhält den sogenannten Gibbs sampler.

- Damit Iterationen des Übergangskerns tatsächlich gegen π^* konvergieren, ist neben der geeigneten Invarianzeigenschaft zusätzlich Irreduzibilität und Aperiodizität erforderlich. In der Praxis sind diese Eigenschaften fast immer erfüllt, können aber theoretisch schwer nachgewiesen werden. In der Praxis wird die Konvergenz daher durch Analyse des MCMC Outputs überprüft, z.B.
 - grafische Darstellung der Samplingpfade
 - Berechnung der Autokorrelationsfunktion der gezogenen Zufallszahlen.

Kapitel 4

Diskrete Markov-Prozesse

Bei Markov-Ketten ist $T = \mathbb{N}_0$, d.h. die Zeit wird als diskret betrachtet. Bei Markov-Prozessen läuft die Zeit kontinuierlich mit $T = \mathbb{R}_+$. Ist S diskret, wie in diesem Kapitel, sprechen wir von diskreten Markov-Prozessen.

Inhalt:

- Allgemeine Eigenschaften
- Poisson-Prozess, Geburts- und Todesprozesse
- Abriss der allgemeinen Theorie für reguläre Prozesse, insbesondere mit endlichem S .

4.1 Definition und elementare Eigenschaften

Beispiele

- Klinische Studien, Krankenverläufe
- Markenwahlprozess
- genetische Evolution
(Liò, P. & Goldmann, N.(1998) Models of Molecular Evolution and Phylogeny. Genome Research 8, 1233–1244).

Fragestellung: Gesucht sind stochastische Modelle für die molekulare Evolution, d.h. die durch Mutation verursachte Veränderung der Nukleotide A, T, C, G an bestimmten Positionen der DNA (oder analog für 20 Basen in Aminosäuren).

Vereinfachende Annahme: Die Evolution an verschiedenen Positionen verläuft unabhängig. Dann

genügt es, Modelle separat für die Positionen zu entwickeln.

Weitere Annahme: Homogenität und Stationarität. Verschiedene Modelle unterscheiden sich durch Annahme über die Raten λ_{ij} für Übergänge $i \rightarrow j$.

- Poisson-Prozess

Diskrete Markov-Prozesse sind einfache stochastische Modelle für zeitstetige stochastische Vorgänge. In Anwendungen können die Voraussetzungen zu restriktiv sein: \rightarrow Kapitel 5 und 7.

Definition 4.1 Diskreter Markov-Prozess

Ein SP $X = \{X(t), t \geq 0\}$ mit abzählbarem Zustandsraum heißt (*diskreter*) *Markov-Prozess* (MP) $:\Leftrightarrow \forall n \geq 0 \quad \forall t \geq s > s_n > \dots > s_0 \geq 0$ gilt

$$P\{X(t) = j \mid X(s) = i, X(s_n) = i_n, \dots, X(s_0) = i_0\} = P\{X(t) = j \mid X(s) = i\} \quad (4.1)$$

für beliebige $j, i, i_n, \dots, i_0 \in S$.

$$p_{ij}(s, t) := P\{X(t) = j \mid X(s) = i\}$$

heißt *Übergangswahrscheinlichkeit(-sfunktion)* (ÜW),

$$P(s, t) = (p_{ij}(s, t))$$

Übergangsmatrix (ÜM).

Der MP heißt *homogen* (oder besitzt *stationäre* ÜW) $:\Leftrightarrow$

$$P\{X(t+s) = j \mid X(s) = i\} = P\{X(t) = j \mid X(0) = i\} = p_{ij}(0, t) =: p_{ij}(t),$$

d.h. für die ÜW ist nur die Zeitdifferenz maßgeblich.

Vereinbarung:

Wie bei den MK wollen wir uns im folgenden meist auf homogene MP beschränken.

Bemerkung:

$p_{ij}(t)$ entspricht $p_{ij}^{(t)}$ (t -schrittige ÜW) bei MK.

Satz 4.1 Chapman-Kolmogorov-Gleichung

Für die ÜW gilt

$$\begin{aligned}
p_{ij}(t) &\geq 0, & \sum_{j \in S} p_{ij}(t) &= 1 \\
p_{ij}(s+t) &= \sum_{k \in S} p_{ik}(s)p_{kj}(t)
\end{aligned} \tag{4.2}$$

bzw.:

$$P(s+t) = P(s)P(t)$$

Beweis:

$$\begin{aligned}
P\{X(s+t) = j \mid X(0) = i\} &= \sum_k P\{X(s+t) = j, X(s) = k \mid X(0) = i\} \\
&= \sum_k P\{X(s) = k \mid X(0) = i\} \cdot P\{X(s+t) = j \mid X(s) = k, X(0) = i\} \\
&= \sum_k P\{X(s) = k \mid X(0) = i\} \cdot P\{X(s+t) = j \mid X(s) = k\}
\end{aligned}$$

Für homogene MP erhält man daraus (4.2). □**Satz 4.2 Gemeinsame Verteilung** $\forall n \in \mathbb{N} \quad \forall 0 \leq t_0 < t_1 < \dots < t_n, \quad i_0, i_1, \dots, i_n \in S$ gilt

$$\begin{aligned}
P\{X(t_n) = i_n, \dots, X(t_1) = i_1 \mid X(t_0) = i_0\} &= p_{i_0 i_1}(t_1 - t_0) \cdot \dots \cdot p_{i_{n-1} i_n}(t_n - t_{n-1}) \\
P\{X(t_n) = i_n, \dots, X(t_1) = i_1, X(t_0) = i_0\} &= \\
&= P\{X(t_0) = i_0\} \cdot p_{i_0 i_1}(t_1 - t_0) \cdot \dots \cdot p_{i_{n-1} i_n}(t_n - t_{n-1}).
\end{aligned}$$

Beweis:Völlig analog zum zeitdiskreten Fall bei MK. □Das bedeutet, dass die gemeinsame Verteilung von $X(t_0), \dots, X(t_n)$ nach Vorgabe einer Anfangsver-

teilung bestimmt werden kann. Über den Satz von Kolmogorov lässt sich auch entsprechend wie für MK zeigen, dass durch Vorgabe einer Anfangsverteilung und von ÜW ein MP, der die vorgegebenen ÜW besitzt, konstruiert werden kann.

Für die statistische Inferenz ist Satz 4.2 wichtig, da er die Likelihoodfunktion bei gegebenen Daten i_1, \dots, i_n zu den Zeitpunkten t_1, \dots, t_n liefert.

In Anwendungen werden statt der ÜW oft sogenannte (*Übergangs-*)*Intensitäten* bzw. *Raten* zugrundegelegt oder aus Daten geschätzt. Für die Definition setzen wir voraus

$$p_{ij}(0) = \lim_{h \downarrow 0} p_{ij}(h) = \delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}.$$

d.h. dass ein Übergang von i nach j ($i \neq j$) eine positive Zeitspanne beansprucht. Es lässt sich dann die Stetigkeit der $p_{ij}(t)$ für $t \geq 0$ und die Existenz der folgenden Grenzwerte bzw. Ableitungen zeigen.

Definition 4.2 (Übergangs-)Intensität, Rate

Für $i, j \in S$, $i \neq j$, heißt

$$\begin{aligned} \lambda_{ij} &= \lim_{h \downarrow 0} \frac{p_{ij}(h)}{h} = \lim_{h \downarrow 0} \frac{P(X(t+h) = j \mid X(t) = i)}{h} \\ &= \lim_{h \downarrow 0} \frac{P(X(t+h) = j \mid X(t) = i, \text{Vorgeschichte})}{h} = p'_{ij}(0) \end{aligned}$$

Übergangsintensität bzw. *Rate* von i nach j .

Für $i = j$ definieren wir

$$\lambda_{ii} = \lim_{h \downarrow 0} \frac{p_{ii}(h) - p_{ii}(0)}{h} = \lim_{h \downarrow 0} \frac{p_{ii}(h) - 1}{h}.$$

Zusammenfassend gilt für $i, j \in S$

$$\lambda_{ij} = p'_{ij}(0) = \lim_{h \downarrow 0} \frac{p_{ij}(h) - p_{ij}(0)}{h}.$$

Dabei gilt $0 \leq \lambda_{ij} < \infty$ für $i \neq j$, aber $\lambda_{ii} \leq 0$.

$\Lambda = (\lambda_{ij})$ heißt *Intensitätsmatrix*.

In $o(h)$ -Schreibweise gilt also für $i \neq j$

$$p_{ij}(h) = P(X(t+h) = j \mid X(t) = i) = \lambda_{ij}h + o(h),$$

und

$$p_{ii}(h) = 1 + \lambda_{ii}h + o(h).$$

Beispiel: Poisson-Prozess

Für den PP $\{N(t), t \geq 0\}$ gilt

$$\begin{aligned} P(N(t+h) - N(t) \geq 2) &= o(h) \\ P(N(t+h) - N(t) = 1) &= \lambda h + o(h) \end{aligned}$$

$$\begin{aligned} \Rightarrow \quad p_{i,i+1}(h) &= \lambda h + o(h), \text{ also } \lambda_{i,i+1} = \lambda \\ p_{ij}(h) &= \begin{cases} 0, & 0 \leq j \leq i-1 \\ o(h), & j \geq i+2 \end{cases} \end{aligned}$$

also $\lambda_{ij} = 0$ für $j \leq i-1$ und $j \geq i+2$, $p_{ii}(h) = 1 - \lambda h + o(h)$, also $\lambda_{ii} = -\lambda$.

Damit erhält man die Intensitätsmatrix

also eine Bandmatrix mit Zeilensumme = 0.

Die Definition 4.2 und das Beispiel zeigen, wie aus ÜW die Intensitäten erhalten werden können. Die schwierigere Fragestellung, wie aus Λ die ÜM $P(t)$ zu berechnen sind, wird in den folgenden Abschnitten behandelt.

4.2 Geburts- und Todesprozesse

4.2.1 Geburtsprozesse

Pfad: Treppenfunktion wie beim Poisson-Prozess

Aber: Intensität $\lambda \rightarrow$ zustandsabhängige Intensität λ_i

Definition 4.3 Geburtsprozess

Ein SP $X = \{X(t), t \geq 0\}$ mit Zustandsraum $S \subset \mathbb{N}_0$ heißt *Geburtsprozess* $:\Leftrightarrow X$ ist ein homogener, diskreter MP mit

$$\begin{aligned} p_{i,i+1}(h) &= \lambda_i h + o(h), \\ p_{i,i}(h) &= 1 - \lambda_i h + o(h), \quad \lambda_i \geq 0, \quad i \geq i_0 \\ p_{i,j}(h) &= \begin{cases} 0 & , \quad 0 \leq j \leq i-1 \\ o(h) & , \quad j \geq i+2 \end{cases} \end{aligned}$$

Damit sind analog zum Poisson-Prozess die Intensitäten bestimmt durch

$$\lambda_{i,i+1} = \lambda_i, \quad \lambda_{ii} = -\lambda_i$$

und $\lambda_{ij} = 0$ sonst. Damit hat auch Λ die gleiche Bandstruktur; es ist nur λ durch λ_i zu ersetzen.

Aus der Definition 4.3, d.h. aus den Intensitäten, lassen sich mit Hilfe der Chapman-Kolmogorov-Gleichungen ÜW (mit $i_0 = 0$)

$$p_{0n}(t) = P(X(t) = n \mid X(0) = 0)$$

herleiten:

$$\begin{aligned} p_{0n}(t+h) &= \sum_{k=0}^{\infty} p_{0k}(t)p_{kn}(h) \\ &= p_{0n}(t)(1 - \lambda_n h + o(h)) + p_{0,n-1}(t)(\lambda_{n-1} h + o(h)) + \sum_{k=0}^{n-2} p_{0k}(t)p_{kn}(h) \\ &= p_{0n}(t)(1 - \lambda_n h) + p_{0,n-1}(t)\lambda_{n-1} h + o(h) \quad \text{für } n \geq 1, \end{aligned}$$

bzw.

$$p_{00}(t+h) = p_{00}(t)(1 - \lambda_0 h) + o(h) \quad \text{für } n = 0.$$

Grenzübergang $h \downarrow 0$ ergibt

$$\begin{aligned} p'_{00}(t) &= -\lambda_0 p_{00}(t) \\ p'_{0n}(t) &= -\lambda_n p_{0n}(t) + \lambda_{n-1} p_{0,n-1}(t), \quad n \geq 1. \end{aligned}$$

Unter Berücksichtigung der Anfangsbedingungen

$$p_{00}(0) = 1, \quad p_{0n}(0) = 0 \quad \text{für } n > 0,$$

lässt sich das DGL-System lösen mit

$$p_{0n}(t) = \sum_{i=0}^n A_n^{(i)} e^{-\lambda_i t}, \quad n \geq 0,$$

falls $\lambda_i \neq \lambda_j$ für $i \neq j$ gilt. Dabei ist

$$A_n^{(i)} = \frac{\lambda_0 \lambda_1 \dots \lambda_{n-1}}{(\lambda_0 - \lambda_i) \dots (\lambda_{i-1} - \lambda_i) (\lambda_{i+1} - \lambda_i) \dots (\lambda_n - \lambda_i)}.$$

Man sieht, dass die Herleitung der ÜW aus den Intensitäten bereits in diesem einfachen Spezialfall eines MP aufwendig ist. Im allgemeinen existiert keine analytische Lösung, vgl. Abschnitt 4.3.

Eine Wahrscheinlichkeitsverteilung, d.h.

$$\sum_{n=0}^{\infty} p_{0n}(t) = 1 \quad \forall t \geq 0$$

ergibt sich genau dann (Feller), wenn

$$\sum_{i=0}^{\infty} \frac{1}{\lambda_i} = \infty.$$

Das bedeutet, dass der Geburtsprozess nicht zu rasch anwachsen, nicht „explodieren“ darf. Beispielsweise ist für $\lambda_i = 2^i$ die Summe endlich, und $X(t)$ wächst mit positiver Wahrscheinlichkeit in endlicher Zeit über alle Grenzen.

Führen wir für dieses Verhalten den zusätzlichen Zustand ∞ ein, so lässt sich dieses Phänomen auch mit dem folgenden Satz plausibel machen.

Satz 4.3 Exponentialverteilte Verweildauern

Die Verweildauern T_i , $i \in \mathbb{N}$, sind unabhängig und exponentialverteilt mit Parameter λ_i , d.h.

$$T_i \sim \text{Ex}(\lambda_i).$$

Beweis: *Spezialfall eines allgemeinen Satzes in Abschnitt 4.3.*

Es gilt also $E(T_i) = \frac{1}{\lambda_i}$. Daher ist $\sum_{i=0}^{\infty} \frac{1}{\lambda_i}$ die erwartete Zeit bevor $\{X(t) = \infty\}$ eintritt. Falls die Summe divergiert scheint plausibel, dass dann $P\{X(t) = \infty\} = 0$ ist. Andernfalls ist $P\{X(t) = \infty\} = 1 - \sum_{n=0}^{\infty} p_{0n}(t)$.

Beispiel: Yule-Prozess (linearer Wachstumsprozess)

Eine Population bestehe aus einem Anfangsbestand von $X(0) = i_0$ Individuen. Diese vermehren sich unabhängig voneinander, wobei jedes Individuum im Zeitintervall h mit Wahrscheinlichkeit $\lambda h + o(h)$ ein weiteres Mitglied der Population erzeugt, die Wahrscheinlichkeit für die Erzeugung mehrerer

Individuen sei $o(h)$. Der binomische Lehrsatz liefert

$$p_{i,i+1}(h) = \binom{i}{1} (\lambda h + o(h))(1 - \lambda h + o(h))^{i-1} = i\lambda h + o(h) \quad \text{und}$$

$$p_{i,j}(h) = \binom{i}{j-i} (\lambda h + o(h))^{j-i} (1 - \lambda h + o(h))^{i-(j-i)} = o(h) \quad \text{für } j - i \geq 2.$$

Es handelt sich also um einen Geburtsprozess mit der Rate $\lambda_i = i\lambda$.

Man kann zeigen, dass $X(t) - i_0$ eine negative Binomialverteilung mit den Parametern i_0 und $\exp(-\lambda t)$ besitzt. Also gilt

$$E(X(t) - i_0) = \frac{i_0(1 - \exp(-\lambda t))}{\exp(-\lambda t)}, \quad \text{Var}(X(t) - i_0) = \frac{i_0(1 - \exp(-\lambda t))}{\exp(-2\lambda t)}$$

und damit ergibt sich ein „exponentielles Wachstum“:

$$E(X(t)) = i_0 \exp(\lambda t), \quad \text{Var}(X(t)) = i_0 \exp(\lambda t)(\exp(\lambda t) - 1).$$

4.2.2 Geburts- und Todesprozesse

Pfad: Treppenfunktion mit Sprunghöhe ± 1

Die Definition erfolgt wieder über die Intensitäten nach oben oder unten zu springen.

Definition 4.4 Geburts- und Todesprozess

Ein SP $X = \{X(t), t \geq 0\}$ mit Zustandsraum $S \subset \mathbb{N}_0$ heißt Geburts- und Todesprozess $:\Leftrightarrow X$ ist ein homogener, diskreter MP mit

$$\begin{aligned} p_{i,i+1}(h) &= \lambda_i h + o(h), & i \geq 0 \\ p_{i,i-1}(h) &= \mu_i h + o(h), & i \geq 1 \\ p_{ii}(h) &= 1 - (\lambda_i + \mu_i)h + o(h), & i \geq 0 \\ p_{ij}(h) &= o(h), & |i - j| \geq 2, \quad i, j \geq 0 \\ (p_{ij}(0) &= 0, j \neq i, p_{ii}(0) = 1) \end{aligned}$$

mit

$$\begin{aligned} \mu_0 &= 0, \\ \lambda_0 &\underset{(\text{=})}{>} 0 \quad (> 0 \text{ „reflektierend“}, = 0 \text{ „absorbierend“}), \\ \mu_i, \lambda_i &\geq 0 \quad \text{für } i \geq 1. \end{aligned}$$

Für die Intensitäten gilt also

$$\begin{aligned} \lambda_{i,i+1} &= \lambda_i, \\ \lambda_{i,i-1} &= \mu_i, \\ \lambda_{ij} &= 0 \text{ für } |i - j| \geq 2, \\ \lambda_{ii} &= -(\lambda_i + \mu_i) \end{aligned}$$

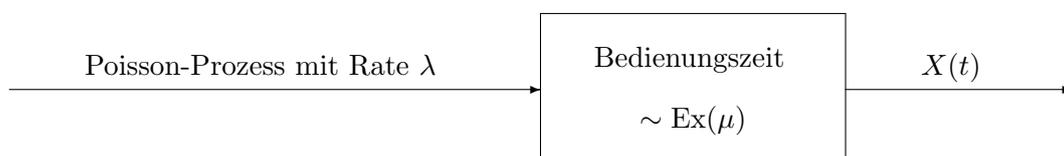
$$\Lambda = \begin{pmatrix} & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \end{pmatrix}$$

ist tridiagonal.

Die explizite Berechnung der ÜM $P(t)$ aus Λ ist i.a. nicht möglich, vgl. FKO S. 96 ff.

Spezielle GT-Prozesse spielen eine Rolle in der Warteschlangen- und Bedienungstheorie (vgl. FKO S. 101 ff).

Das Wartesystem M/M/1/ ∞



$X(t)$: Anzahl der im System befindlichen Kunden

$X(t)$ ist spezieller GT-Prozess mit $\lambda_i = \lambda$ und $\mu_i = \mu$

Für $\lambda < \mu$ existiert Grenzverteilung p_j :

$$p_j = \lim_{t \rightarrow \infty} p_{ij}(t) = (1 - r)r^j, \quad r = \frac{\lambda}{\mu}$$

Also:

$$\begin{aligned} \lim_{t \rightarrow \infty} X(t) = X(\infty) &\sim \text{geometrisch verteilt} \\ \rightarrow \text{Erwartungswert} &= \frac{r}{1 - r}, \quad \text{Varianz} = \frac{r}{(1 - r)^2} \end{aligned}$$

Variante

- begrenzte Kapazität N des „Warteraums“

$$\rightarrow \lambda_i = \begin{cases} \lambda & , 0 \leq i < N \\ 0 & , i \geq N \end{cases}$$

- von der Länge der angetroffenen Warteschlange verursachte Entmutigung, z.B.

$$\lambda_i = \frac{\lambda}{(i + 1)}, \quad i = 0, 1, \dots$$

Das Wartesystem M/M/s/ ∞

Nun: s Schalter verfügbar; jeweils Rate μ

$X(t)$ GT-Prozess mit

$$\lambda_i = \lambda \quad \text{und} \quad \mu_i = \begin{cases} \mu i & , i = 1, \dots, s \\ \mu s & , i = s + 1, \dots \end{cases}$$

Grenzverteilung existiert für $\frac{\lambda}{s\mu} < 1$.

Bevölkerungsentwicklung

Verallgemeinerung des Yule-Prozesses, zum Zeitpunkt t : i Individuen

→ Geburtsrate $\lambda_i = i\lambda$,

→ Sterberate $\mu_i = i\mu$

Für $X(0) = 1$ kann man verschiedene Resultate zeigen, z.B.

$$E(X(t)) = \exp((\lambda - \mu)t) \quad (\text{für } \lambda \neq \mu)$$

oder

$$\text{Wahrscheinlichkeit des Aussterbens} = \begin{cases} 1 & , \lambda \leq \mu \\ \frac{\mu}{\lambda} & , \lambda > \mu \end{cases}$$

Abschließend gehen wir kurz ein auf

Struktur und Simulation der Pfade von Geburts- und Todesprozessen

Es ist $(\lambda_i + \mu_i)h + o(h)$ die Wahrscheinlichkeit dafür, dass im Zeitraum $(t, t+h]$ der Zustand i verlassen wird. Gegenüber dem Poisson-Prozess tritt also $\lambda_i + \mu_i$ an die Stelle von λ und es lässt sich zeigen, dass die Verweildauer im Zustand i exponentialverteilt mit Erwartungswert $\frac{1}{(\lambda_i + \mu_i)}$ ist. Während des Zeitraums h findet eine Zustandsänderung $i \rightarrow i+1$ mit der Wahrscheinlichkeit $\lambda_i h + o(h)$, eine Änderung $i \rightarrow i-1$ mit der Wahrscheinlichkeit $\mu_i h + o(h)$ statt. Deshalb ist es plausibel, dass $\frac{\lambda_i}{(\lambda_i + \mu_i)}$ bzw. $\frac{\mu_i}{(\lambda_i + \mu_i)}$ die bedingte Wahrscheinlichkeit (unter der Bedingung, dass eine Änderung stattfindet) für einen Übergang $i \rightarrow i+1$ bzw. $i \rightarrow i-1$ ist.

Damit kann man sich die Entstehung der Pfade von Geburts- und Todesprozessen folgendermaßen veranschaulichen: Der momentane Zustand sei i . Es wird eine Realisation einer $(\lambda_i + \mu_i)$ -exponentialverteilten Zufallsvariablen erzeugt. Diese Realisierung ist die Verweildauer in i . Anschließend wird ein Bernoulli-Experiment durchgeführt mit $p_i = \frac{\lambda_i}{\lambda_i + \mu_i}$, das entscheidet, ob (mit Wahrscheinlichkeit p_i) nach $i+1$ oder nach $i-1$ (mit Wahrscheinlichkeit $1-p_i$) gesprungen wird, usw.

4.3 Diskrete MP: Skizze der allgemeinen Theorie

Die Definition 4.1 lässt noch Pfade mit sehr irregulären Eigenschaften zu, z.B. Geburtsprozesse mit Explosion. Deshalb stellen wir Zusatzforderungen:

- $p_{ij}(0) = \lim_{h \downarrow 0} p_{ij}(h) = \begin{cases} 1 & , i = j \\ 0 & , i \neq j \end{cases}$
- Pfade (m. Wkeit 1) rechtsseitig stetig

Bezeichnungen:

- S_n Zeitpunkt der n -ten Zustandsänderung, $S_0 = 0$,
- $Y_n = X(S_n)$ zum Zeitpunkt S_n angenommener Zustand,

- $T_{n+1} = S_{n+1} - S_n$ Verweildauer in Y_n ,
- $S_{n+1} = S_n + T_{n+1}$.
- Falls „absorbierende“ Zustände Y_n möglich sind, d.h. $S_{n+1} = \infty$, definieren wir

$$Y_{n+1} = \begin{cases} Y_n & , S_{n+1} = \infty \\ X(S_{n+1}) & , S_{n+1} < \infty. \end{cases}$$

- $V(t)$ Dauer (vom Zeitpunkt t) bis zur nächsten Zustandsänderung, „Vorwärtsrekurrenzzeit“.

Für die Vorwärtsrekurrenzzeit gilt in Verallgemeinerung des entsprechenden Ergebnisses für den Poisson-Prozess:

Satz 4.4 Exponentialverteilte Dauern

Die Verweildauern T_n , $n \in \mathbb{N}$ sind exponentialverteilt mit Parameter λ_i , sodass gilt $T_n | Y_{n-1} = i \sim \text{Exp}(\lambda_i)$. Für jeden Zeitpunkt t ist

$$P(V(t) \leq s | X(t) = i) = 1 - e^{-\lambda_i s} \quad , s \geq 0,$$

also ebenfalls exponentialverteilt mit Parameter λ_i .

Beweis:

(für $V(t)$)

Sei $f_i(s) := P(V(t) > s | X(t) = i)$. Wegen der Homogenität des MP ist $f_i(s)$ von t unabhängig und es gilt für alle $r, s \geq 0$

$$\begin{aligned} f_i(s+r) &= P(V(t) > s+r | X(t) = i) \\ &= P(V(t) > s, V(t+s) > r | X(t) = i) \\ &= P(V(t) > s | X(t) = i) \cdot P(V(t+s) > r | X(t) = i, V(t) > s) \\ &= P(V(t) > s | X(t) = i) \cdot P(V(t+s) > r | X(t+s) = i) \\ &= f_i(s) \cdot f_i(r). \end{aligned}$$

Wegen $0 \leq f_i(s) \leq 1$ muss $f_i(s) = e^{-\lambda_i s}$ mit $\lambda_i \geq 0$ sein. Für $\lambda_i = \infty$ ist $f_i(s) \equiv 0$. □

Definition 4.5 Zustände

Ein Zustand i heißt

$$\begin{aligned} \text{absorbierend} & :\Leftrightarrow \lambda_i = 0 \\ \text{stabil} & :\Leftrightarrow 0 < \lambda_i < \infty \\ \text{instabil} & :\Leftrightarrow \lambda_i = \infty \end{aligned}$$

Falls i absorbierend ist, gilt also für alle $s \geq 0$

$$P(V(t) > s \mid X(t) = i) = P(X(t+s) = i \mid X(t) = i) = 1,$$

der Zustand i wird also m. Wkeit 1 nicht mehr verlassen.

Falls i stabil ist, gilt

$$P(0 < V(t) < \infty \mid X(t) = i) = 1,$$

d.h. falls X zur Zeit t in i ist, verweilt X dort eine positive, aber *endliche* Zeit (m. Wkeit 1).

Falls i instabil ist, gilt

$$P(V(t) = 0 \mid X(t) = i) = 1,$$

d.h. der MP verlässt einen instabilen Zustand sofort wieder. Dass ein solches Verhalten zu irregulären Pfaden führt, erscheint plausibel. Man sieht leicht ein, dass insbesondere die Pfade nicht rechtsseitig stetig sein können. Da dies gegen unsere Vereinbarung verstößt, sind *instabile Zustände im weiteren ausgeschlossen*.

Durch den folgenden wichtigen Satz wird die Struktur von MP mit rechtsseitig stetigen Pfaden weitgehend aufgeklärt. Er zeigt, dass die nacheinander besuchten Zustände Y_n eine MK bilden und die Verweildauern exponentialverteilt sind.

Satz 4.5 Struktursatz

Sei X ein MP mit rechtsseitig stetigen Pfaden. Dann gilt

$$\forall n \in \mathbb{N}_0 \quad \forall j, i, \dots, i_0 \in S \quad \forall t, t_1, \dots, t_n \in \mathbb{R} \quad \forall 0 < s_1 \dots < s_n$$

$$1. P(Y_{n+1} = j, T_{n+1} > t \mid Y_n = i, \dots, Y_0 = i_0, S_n = s_n, \dots, S_0 = 0) = q_{ij} e^{-\lambda_i t}$$

$$\text{mit } q_{ij} \geq 0, \sum_{j \in S} q_{ij} = 1, q_{ii} = \begin{cases} 0 & , i \text{ stabil} \\ 1 & , i \text{ absorbierend} \end{cases}$$

$$2. \{Y_n, n \in \mathbb{N}\} \text{ ist eine MK mit der ÜM } Q = (q_{ij})$$

$$3. P(T_{n+1} > t \mid Y_n = i, Y_{n+1} = j) = P(T_{n+1} > t \mid Y_n = i) = e^{-\lambda_i t}$$

$$4. P(T_1 > t_1, \dots, T_n > t_n \mid Y_n = i, \dots, Y_0 = i_0) = e^{-\lambda_{i_0} t_1} \dots e^{-\lambda_{i_{n-1}} t_n}$$

Bemerkungen:

- (a) $\{Y_n, n \in \mathbb{N}\}$ heißt „eingebettete“ MK, Erreichbarkeits-/Rekurrenzeigenschaften der eingebetteten MK übertragen sich auf diskreten MP.
- (b) Aussage 3 besagt, dass λ_i nur vom eben besuchten Zustand, nicht aber vom nächsten abhängt.
- (c) Aussage 4 besagt, dass die Verweildauern exponentialverteilt und bedingt unabhängig sind, falls die Folge der besuchten Zustände gegeben ist.
- (d) Der Struktursatz liefert auch die gemeinsame Verteilung von $\{T_n, Y_n\}$ und damit die Grundlage für die Likelihood zur ML-Schätzung der Strukturparameter λ_i und q_{ij} (bzw. der Intensitäten λ_{ij})

$$\begin{aligned}
 P(T_1 > t_1, \dots, T_n > t_n, Y_0 = i_0, Y_1 = i_1, \dots, Y_n = i_n) \\
 &= P(T_1 > t_1, \dots, T_n > t_n \mid Y_0 = i_0, Y_1 = i_1, \dots, Y_n = i_n) \\
 &\quad \cdot P(Y_0 = i_0, Y_1 = i_1, \dots, Y_n = i_n) \\
 &= e^{-\lambda_0 t_1} \dots e^{-\lambda_{i_{n-1}} t_n} \cdot P(Y_0 = i_0) \cdot q_{i_0 i_1} \dots q_{i_{n-1} i_n}
 \end{aligned}$$

Beweis:

Die linke Seite von 1 ist gleich

$$P\{X(S_n + T_{n+1}) = j, T_{n+1} > t \mid X(S_n) = i, \dots, X(0) = i_0, S_n = s_n, \dots, S_0 = 0\}.$$

Wegen der (starken) Markov-Eigenschaft und der Homogenität ist dies gleich

$$\begin{aligned}
 P\{X(T_1) = j, T_1 > t \mid X(0) = i\} \\
 &= P\{T_1 > t \mid X(0) = i\} \cdot P\{X(T_1) = j \mid X(0) = i, T_1 > t\} \\
 &\stackrel{\text{Satz 4.4}}{=} e^{-\lambda_i t} \cdot P\{X(T_1) = j \mid X(0) = i, T_1 > t\} \\
 &= e^{-\lambda_i t} \cdot P\{X(t + V(t)) = j \mid X(t) = i\} \\
 &= e^{-\lambda_i t} \cdot P\{Y_1 = j \mid Y_0 = i\} \\
 &= e^{-\lambda_i t} q_{ij}.
 \end{aligned}$$

Daraus folgt 1. Setzt man in 1 $t = 0$, so ergibt sich 2; 3 ergibt sich ebenfalls aus 1; 4 aus 3 durch Induktion. □

Satz 4.5 liefert folgende

Konstruktion bzw. Simulation der Pfade eines MP

Man gibt einen Startwert $X(\omega, 0) = Y_0(\omega) = i_0$ vor. Befindet man sich im Zustand $Y_n(\omega) = i_n, n \in \mathbb{N}_0$, so würfelt man gemäß der Verteilung $q_{i_n j}, j \in S$, den nächsten Zustand $Y_{n+1}(\omega) = i_{n+1}$. Die Verweil-

dauer $T_{n+1}(\omega) = t_{n+1}$ in i_n bis zum Sprung nach i_{n+1} wird gemäß einer λ_{i_n} -Exponentialverteilung realisiert.

Zu klären bleibt noch, in welcher Weise die ÜW $p_{ij}(t)$ die Strukturparameter q_{ij} , λ_i bestimmen und umgekehrt. Dies soll im folgenden geschehen.

Dazu wollen wir noch sicherstellen, dass sich die Sprungstellen S_n nicht im Endlichen häufen; dann können auch keine Explosionen in endlicher Zeit stattfinden. Deshalb fordern wir

$$\sup_n S_n = +\infty \quad \text{m. Wkeit 1}$$

$X(t)$ lässt sich dann durch die Y_n , S_n bzw. T_n festlegen:

$$X(t) = Y_n \quad \text{für } t \in [S_n, S_{n+1}).$$

Zusammenfassend kommen wir so zur

Definition 4.6 Regulärer MP

Ein diskreter (homogener) MP heißt *regulär* $:\Leftrightarrow$

1. $p_{ij}(0) = \lim_{h \downarrow 0} p_{ij}(h) = \begin{cases} 1 & , i = j \\ 0 & , i \neq j \end{cases}$
2. Die Pfade sind (m. Wkeit 1) rechtsseitig stetig.
3. $\sup_n S_n = +\infty$ m. Wkeit 1

Vereinbarung: Im weiteren betrachten wir nur reguläre MP. Reguläre MP werden ausführlicher in *Cinlar*, Kap. 8 behandelt. Eine grundlegende Darstellung der Theorie nicht notwendig regulärer MP bietet *Chung*. Einige der folgenden Ergebnisse gelten auch für derartige allgemeinere Fälle.

Satz 4.6 Regularitätskriterien

Für einen MP X gelten 1 und 2 der Definition 4.6. Dann ist X regulär, falls eines der folgenden Kriterien zutrifft:

1. S endlich
2. $\lambda_i \leq c$ für alle $i \in S$
3. Alle Zustände der eingebetteten MK Y sind rekurrent.
4. Die MK Y verbleibt mit Wahrscheinlichkeit 0 in einer transienten Klasse.

Beweis: FKO S. 111

Der folgende Satz zeigt, wie sich aus vorgegebenen ÜW $p_{ij}(t)$ die Intensitäten λ_{ij} bestimmen lassen und wie sie mit den λ_i und q_{ij} zusammenhängen.

Satz 4.7 Intensitäten und Strukturparameter

X sei ein regulärer MP. Dann sind die ÜW $p_{ij}(t)$ stetig differenzierbar und es gilt

$$p'_{ij}(0) = \lim_{h \downarrow 0} \frac{p_{ij}(h) - p_{ij}(0)}{h} = \lambda_{ij} = \begin{cases} -\lambda_i & , i = j \\ \lambda_i q_{ij} & , i \neq j, \end{cases}$$

bzw. in $o(h)$ -Schreibweise

$$\begin{aligned} P(X(t+h) = j \mid X(t) = i) &= h \underbrace{\lambda_i q_{ij}}_{\lambda_{ij}} + o(h), \quad i \neq j \quad \left(\Rightarrow q_{ij} = \frac{\lambda_{ij}}{\lambda_i}, \text{ falls } \lambda_i > 0 \right) \\ P(X(t+h) = i \mid X(t) = i) &= 1 - \lambda_i h + o(h). \end{aligned}$$

In Matrixschreibweise (S endlich):

$$P(h) = I + \Lambda h + o(h).$$

bzw.

$$\lim_{h \rightarrow \infty} \frac{P(h) - I}{h} = P'(0) = \Lambda$$

Beweis:

Wir zeigen nur die Hauptaussage, jedoch nicht, dass die ÜW stetig differenzierbar sind. Da die Verweildauern exponentialverteilt sind, ist die Wahrscheinlichkeit für mehr als einen Übergang im Zeitraum h gleich $o(h)$. Deshalb gilt

$$\begin{aligned} p_{ii}(h) = P\{V(t) > h \mid X(t) = i\} + o(h) &\stackrel{\text{Satz 4.4}}{=} e^{-\lambda_i h} + o(h) \\ &= 1 - \lambda_i h + o(h) \end{aligned}$$

also

$$\frac{p_{ii}(h) - p_{ii}(0)}{h} = -\lambda_i + \frac{o(h)}{h}.$$

Grenzübergang $h \rightarrow 0$ liefert $p'_{ii}(0) = \lambda_{ii} = -\lambda_i$.

Für $i \neq j$ gilt entsprechend

$$\begin{aligned} p_{ij}(h) &= P\{Y_{n+1} = j, T_{n+1} \leq h \mid Y_n = i\} + o(h) \\ &\stackrel{\text{Satz 4.5, 1}}{=} q_{ij} - q_{ij} e^{-\lambda_i h} + o(h) \\ &= q_{ij} \lambda_i h + o(h), \end{aligned}$$

$h \rightarrow 0$ liefert $p'_{ij}(0) = \lambda_{ij} = \lambda_i q_{ij}$ für $i \neq j$. □

Da

$$q_{ii} = \begin{cases} 0 & , i \text{ stabil } (0 < \lambda_i < \infty) \\ 1 & , i \text{ absorbierend } (\lambda_i = 0), \end{cases}$$

ergibt sich für reguläre MP die

Folgerung

Die Zeilensummen der Intensitätsmatrix Λ sind gleich 0, d.h. für alle $i \in S$ gilt

$$\lambda_{ii} = -\lambda_i = - \sum_{j \in S, j \neq i} \lambda_{ij}$$

Beispiel: Molekulare Evolution

$$S = \{A, T, G, C\}$$

Beispiel: Poisson-/Geburtsprozess

In Anwendungen sind meist λ_{ij} bzw. λ_i und q_{ij} vorgegeben oder aus Daten geschätzt, und die ÜW $p_{ij}(t)$ sind daraus zu berechnen. Prinzipiell, jedoch selten analytisch, lässt sich diese Aufgabe mit Hilfe des folgenden Satzes lösen.

Satz 4.8 $\Lambda \rightarrow P(t)$

Sei X ein regulärer MP. Dann ist die ÜM $P(t) = (p_{ij}(t))$ durch die ÜM Q der eingebetteten MK Y und die Erwartungswerte $\frac{1}{\lambda_i}$ der Verweildauern bzw. die Intensitätsmatrix Λ eindeutig bestimmt. Die ÜW $p_{ij}(t)$ sind Lösung von

$$1. \quad p'_{ij}(t) = \sum_{k \in S} \lambda_{ik} p_{kj}(t) \quad \text{bzw.} \quad P'(t) = \Lambda P(t), \quad P(0) = I$$

(Rückwärtsgleichungen von Kolmogorov)

$$2. \quad p'_{ij}(t) = \sum_{k \in S} p_{ik}(t) \lambda_{kj} \quad \text{bzw.} \quad P'(t) = P(t) \Lambda, \quad P(0) = I$$

(Vorwärtsgleichungen von Kolmogorov)

3. Für endliches S ist die Lösung gegeben durch

$$P(t) = e^{t\Lambda} = I + t\Lambda + \frac{(t\Lambda)^2}{2!} + \frac{(t\Lambda)^3}{3!} + \dots$$

Die (numerische) Berechnung erfolgt mit Hilfe der Spektralzerlegung von Λ

$$\Lambda = U \operatorname{diag}(d_1, \dots, d_m) U^{-1},$$

mit den Eigenwerten d_1, \dots, d_m und der Matrix U der Eigenvektoren. Dann gilt

$$P(t) = U \operatorname{diag}(e^{d_1 t}, \dots, e^{d_m t}) U^{-1}.$$

Beweis:

Nur zu 1 bzw. 2; informell, in Matrixdarstellung (vgl. FKO S. 113). Die Beziehung zwischen $p'_{ij}(0)$ und λ_{ij} in Satz 4.7 lässt sich in Matrizenform durch

$$\lim_{h \downarrow 0} \frac{P(h) - I}{h} = \Lambda$$

darstellen. Dann folgt mit Hilfe der Chapman-Kolmogorov-Gleichungen (Satz 4.1)

$$\frac{P(t+h) - P(t)}{h} = \frac{P(t)P(h) - P(t)}{h} = \frac{P(t)(P(h) - I)}{h}$$

bzw.

$$\frac{P(t+h) - P(t)}{h} = \frac{(P(h) - I)P(t)}{h}.$$

Grenzübergang $h \rightarrow 0$ liefert

$$P'(t) = P(t)\Lambda \quad \text{bzw.} \quad P'(t) = \Lambda P(t)$$

mit $P(0) = I$.

Die Lösung in 3 ergibt sich aus der Theorie linearer Differentialgleichungssysteme. \square

Beispiel: MP mit 2 Zuständen

$$\Lambda = \begin{pmatrix} -1 & 1 \\ 2 & -2 \end{pmatrix}$$

Eigenwertzerlegung von Λ : EW sind 0 und -3 ; es ist

$$D = \text{diag}(d_1, d_2) = \begin{pmatrix} 0 & 0 \\ 0 & -3 \end{pmatrix}, \quad U = \begin{pmatrix} 1 & 1 \\ 1 & -2 \end{pmatrix}, \quad U^{-1} = \begin{pmatrix} \frac{2}{3} & \frac{1}{3} \\ \frac{1}{3} & -\frac{1}{3} \end{pmatrix}.$$

$$\begin{aligned} P(t) &= U \text{diag}(e^{d_1 t}, e^{d_2 t}) U^{-1} \\ &= U \begin{pmatrix} 1 & 0 \\ 0 & e^{-3t} \end{pmatrix} U^{-1} \\ &= \dots \\ &= \begin{pmatrix} \frac{2}{3} & \frac{1}{3} \\ \frac{2}{3} & \frac{1}{3} \end{pmatrix} + e^{-3t} \begin{pmatrix} \frac{1}{3} & -\frac{1}{3} \\ -\frac{2}{3} & \frac{2}{3} \end{pmatrix} \end{aligned}$$

Bemerkung: Es existiert der Grenzwert

$$\lim_{t \rightarrow \infty} P(t) = \begin{pmatrix} \frac{2}{3} & \frac{1}{3} \\ \frac{2}{3} & \frac{1}{3} \end{pmatrix}.$$

Beispiel: Allgemeiner MP mit 2 Zuständen

$$\Lambda = \begin{pmatrix} -\lambda & \lambda \\ \mu & -\mu \end{pmatrix}$$

Die Kolmogorov-Gleichung $P'(t) = P(t)\Lambda$ ergibt

$$\begin{pmatrix} p'_{00}(t) & p'_{01}(t) \\ p'_{10}(t) & p'_{11}(t) \end{pmatrix} = \begin{pmatrix} p_{00}(t) & p_{01}(t) \\ p_{10}(t) & p_{11}(t) \end{pmatrix} \cdot \begin{pmatrix} -\lambda & \lambda \\ \mu & -\mu \end{pmatrix}.$$

$$\Rightarrow p'_{00}(t) = -\lambda p_{00}(t) + \mu p_{01}(t), \quad p_{01}(t) = 1 - p_{00}(t),$$

$$\Rightarrow p'_{00}(t) = +\mu - (\lambda + \mu)p_{00}(t).$$

Lösung:

$$\begin{aligned} p_{00}(t) &= \frac{\mu}{\lambda + \mu} + \frac{\lambda}{\lambda + \mu} e^{-(\lambda + \mu)t} \\ p_{01}(t) &= 1 - p_{00}(t) \\ &= \frac{\lambda}{\lambda + \mu} - \frac{\lambda}{\lambda + \mu} e^{-(\lambda + \mu)t} \end{aligned}$$

Aus Symmetriegründen

$$\begin{aligned} p_{11}(t) &= \frac{\lambda}{\lambda + \mu} + \frac{\mu}{\lambda + \mu} e^{-(\lambda + \mu)t} \\ p_{10}(t) &= \frac{\mu}{\lambda + \mu} - \frac{\mu}{\lambda + \mu} e^{-(\lambda + \mu)t} \end{aligned}$$

Grenzwert:

$$\lim_{t \rightarrow \infty} P(t) = \begin{pmatrix} \frac{\mu}{\lambda + \mu} & \frac{\lambda}{\lambda + \mu} \\ \frac{\mu}{\lambda + \mu} & \frac{\lambda}{\lambda + \mu} \end{pmatrix}.$$

Grenzwertsätze

Die Klassifizierung der Zustände des MP X erfolgt durch die zugrundeliegende MK Y mit der ÜM Q . Mit den t-schrittigen ÜW $q_{ij}^{(t)}$ folgt zunächst für *transientes* j aus $\lim_{t \rightarrow \infty} q_{ij}^{(t)} = 0$ sofort

$$\lim_{t \rightarrow \infty} p_{ij}(t) = 0.$$

Deshalb beschränken wir uns wie bei MK auf irreduzible, rekurrente MP (d.h. die zugrundeliegende MK Y sei irreduzibel rekurrent).

Dann gilt

Satz 4.9 Grenzwertsatz

Der MP X sei regulär und irreduzibel positiv-rekurrent. Dann gilt:

1. $\lim_{t \rightarrow \infty} P(X(t) = j) = \pi_j$ existiert $\forall j \in S$,
und ist von der Anfangsverteilung unabhängig.

2. Die Grenzverteilung π lässt sich berechnen über

(a) $\pi_j = \frac{\eta_j}{\sum_i \eta_i}$ mit $\eta \Lambda = 0$ bzw. über

(b) $\pi_j = \frac{\nu_j / \lambda_j}{\sum_i \nu_i / \lambda_i}$ mit $\nu = \nu Q$.

$$3. \lim_{t \rightarrow \infty} P(t) = \begin{pmatrix} \pi \\ \vdots \\ \pi \\ \vdots \\ \pi \end{pmatrix}$$

4. $\pi = (\dots, \pi_j, \dots)$ ist strikt positiv.

5. π ist die einzige *stationäre* Verteilung von X , d.h. es gilt

$$\pi = \pi \cdot P(t) \quad \forall t \geq 0.$$

Bemerkungen:

1. $(\eta_j) = (\nu_j / \lambda_j)$ ist die *unnormierte* Grenzverteilung.
2. $(\nu_j) = \nu$ ist die Grenzverteilung der eingebetteten MK Y .

Plausibilitätsbeweis zu 2: (vgl. auch FKO S. 116)

Kolmogorov'sche Vorwärtsgleichung:

$$P'(t) = P(t) \cdot \Lambda$$

Annahme: $\lim_{t \rightarrow \infty} P(t)$ existiert

$$\lim_{t \rightarrow \infty} P(t) = \begin{pmatrix} \text{---} \pi \text{---} \\ \text{---} \pi \text{---} \end{pmatrix}$$

Dann folgt offensichtlich

$$\lim_{t \rightarrow \infty} P'(t) = 0$$

da

$$\frac{d}{dt} \pi_j = \frac{d}{dt} \lim_{t \rightarrow \infty} p_{ij}(t) = 0$$

(π_j hängt ja nicht von t ab, da Grenzverteilung)

Insgesamt folgt

$$0 = \begin{pmatrix} - & - & - & \pi & - & - & - \\ - & - & - & \pi & - & - & - \end{pmatrix} \cdot \Lambda,$$

in jeder Zeile steht also $2(a) \pi \cdot \Lambda = 0$.

Die Äquivalenz zu 2(b) gilt wegen $\nu = \nu \cdot Q \Leftrightarrow \nu(Q - I) = 0$.

Erweiterung mit $\frac{1}{\lambda_i} \cdot \lambda_i$:

$$\underbrace{\left(\frac{\nu_1}{\lambda_1}, \frac{\nu_2}{\lambda_2}, \dots \right)}_{\mu} \cdot \underbrace{\begin{pmatrix} -\lambda_1 & \lambda_1 q_{12} & \cdots \\ \lambda_2 q_{21} & -\lambda_2 & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix}}_{\Lambda} = 0$$

2(b) ist *plausibel*: Grenzverteilung ist proportional zu

$$\underbrace{\left(\frac{1}{\lambda_j} \right)}_{\text{durchschnittliche Aufenthaltsdauer in } j} \cdot \underbrace{\nu_j}_{\text{Wahrscheinlichkeit sich in } j \text{ aufzuhalten}}$$

Beispiel: 2-Zustands-MP

$$\Lambda = \begin{pmatrix} -\lambda_1 & \lambda_1 \\ \lambda_2 & -\lambda_2 \end{pmatrix}$$

$$-\eta_1 \lambda_1 + \eta_2 \lambda_2 = 0 \Leftrightarrow -\eta_1 \lambda_1 + (1 - \eta_1) \lambda_2 = 0$$

$$\eta_1 \lambda_1 - \eta_2 \lambda_2 = 0 \Leftrightarrow \lambda_2 - \eta_1 (\lambda_1 + \lambda_2) = 0$$

$$\Rightarrow \eta_1 = \frac{\lambda_2}{\lambda_1 + \lambda_2}, \quad \eta_2 = \frac{\lambda_1}{\lambda_1 + \lambda_2}$$

oder alternativ über eingebettete MK Y (2(b)):

Y hat ÜM

$$Q = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

mit stationärer Verteilung ($\nu = \nu Q$)

$$\nu = \left(\frac{1}{2}, \frac{1}{2} \right) \Rightarrow \text{Grenzverteilung } (\pi_1, \pi_2) \text{ ist proportional zu } \left(\frac{1}{\lambda_1}, \frac{1}{\lambda_2} \right)$$

Normieren liefert

$$\begin{aligned} (\pi_1, \pi_2) &= \pi = \frac{1}{\frac{1}{\lambda_1} + \frac{1}{\lambda_2}} \left(\frac{1}{\lambda_1}, \frac{1}{\lambda_2} \right) \\ &= \frac{\lambda_1 \lambda_2}{\lambda_1 + \lambda_2} \left(\frac{1}{\lambda_1}, \frac{1}{\lambda_2} \right) = \left(\frac{\lambda_2}{\lambda_1 + \lambda_2}, \frac{\lambda_1}{\lambda_1 + \lambda_2} \right) \end{aligned}$$

Beispiel: Peter Prinzip

(Taylor/Karlin, S. 366)

Betrieb: 3 berufliche Kategorien

 T Trainee, J Junior-Pos., S Senior-Pos.

$$\Lambda = \begin{array}{c} T \\ J \\ S \end{array} \begin{array}{ccc} T & J & S \\ \left(\begin{array}{ccc} -\lambda_T & \lambda_T & 0 \\ \lambda_{JT} & -\lambda_J & \lambda_{JS} \\ \lambda_S & 0 & -\lambda_S \end{array} \right) \end{array}.$$

J bleibt in J Pos. mit $\text{Ex}(\lambda_J) = \text{Ex}(\lambda_{JT} + \lambda_{JS})$ -verteilter Verweildauer; dann verlässt er die Position und wird durch Trainee ersetzt mit Wahrscheinlichkeit $q_{JT} = \frac{\lambda_{JT}}{\lambda_J}$, oder er bekommt eine S -Position mit Wahrscheinlichkeit $q_{JS} = \frac{\lambda_{JS}}{\lambda_J}$.

Alternative Beschreibung:

$$\begin{aligned} P(X(t+h) = J \mid X(t) = T) &= \lambda_T h + o(h) \\ P(X(t+h) = T \mid X(t) = J) &= \lambda_{JT} h + o(h) \\ P(X(t+h) = S \mid X(t) = J) &= \lambda_{JS} h + o(h) \\ P(X(t+h) = T \mid X(t) = S) &= \lambda_S h + o(h) \\ P(X(t+h) = i \mid X(t) = i) &= 1 - \lambda_i h + o(h) \quad \text{für } i = T, J, S. \end{aligned}$$

Grenzverteilung: $\pi = (\pi_T, \pi_J, \pi_S)$

$$\begin{aligned} \lambda_T \pi_T &= \lambda_{JT} \pi_J + \lambda_S \pi_S \\ \lambda_J \pi_J &= \lambda_T \pi_T \\ \lambda_S \pi_S &= \lambda_{JS} \pi_J \\ 1 &= \pi_T + \pi_J + \pi_S \end{aligned}$$

Lösung ist

$$\begin{aligned} \pi_T &= \frac{\lambda_S \lambda_J}{N}, \\ \pi_J &= \frac{\lambda_S \lambda_T}{N}, \\ \pi_S &= \frac{\lambda_T \lambda_{JS}}{N}. \end{aligned}$$

mit $N = \lambda_S \lambda_J + \lambda_S \lambda_T + \lambda_T \lambda_{JS}$.**4.4 Zur statistischen Analyse**

Zwei Situationen treten in Anwendungen auf.

1. Vollständige Kenntnis über einen oder mehrere Pfade

Bemerkung: Hier *keine* Links- oder Rechtszensierungen berücksichtigt.

2. Beobachtungen nur zu Zeitpunkten $t_1 < t_2 < \dots < t_n$, d.h. Pfade *nicht* vollständig beobachtet.

In beiden Situationen: Likelihood-basierte Inferenz für die Strukturparameter $\Lambda = \{\lambda_{ij}\}$ bzw. $Q = \{q_{ij}\}$, $\lambda = \{\lambda_i\}$, bzw. $\Lambda = \Lambda(\theta)$, $Q = Q(\theta)$, $\lambda = \lambda(\theta)$

Situation 1

Die Likelihood wird gemäß Struktursatz 4.5 bestimmt:

$$\begin{aligned}
 P(Y_n = i_n, T_n > t_n \mid Y_{n-1} = i_{n-1}, S_{n-1} = s_{n-1}, \dots, Y_0 = i_0, S_0 = 0) \\
 = q_{i_{n-1}i_n} \cdot \underbrace{e^{-\lambda_{i_{n-1}} t_n}}_{=, \text{Exponential-Survivorfkt. } 1-F(t_n)}
 \end{aligned}$$

Gemeinsame (bedingte) Dichte für Y_n diskret, T_n stetig (bezüglich Produktmaß von λ , mit ν Abzählmaß, λ Lebeguemaß):

$$q_{i_{n-1}i_n} \cdot \lambda_{i_{n-1}} e^{-\lambda_{i_{n-1}} t_n}$$

Damit:

$$\begin{aligned}
 \text{Likelihood } L(\lambda_{ij} \mid i_0, i_1, \dots, i_{n-1}, i_n; t_1, \dots, t_n) & \\
 &= \text{Produkt bedingter Dichten wie im zeitdiskreten Fall} \\
 &= p_{i_0}(0) \cdot \underbrace{q_{i_0 i_1} \cdot \lambda_{i_0}}_{\lambda_{i_0 i_1}} e^{-\lambda_{i_0} t_1} \cdot q_{i_1 i_2} \lambda_{i_1} e^{-\lambda_{i_1} t_2} \cdot \dots \cdot \\
 &\quad q_{i_j i_{j+1}} \lambda_{i_j} e^{-\lambda_{i_j} t_{j+1}} \cdot \dots \cdot q_{i_{n-1} i_n} \lambda_{i_{n-1}} e^{-\lambda_{i_{n-1}} t_n} \\
 &= p_{i_0}(0) \cdot \prod_{i \in S} \exp(-\lambda_i \gamma_i) \cdot \prod_{i, j \in S, i \neq j} (q_{ij} \lambda_i)^{n_{ij}}
 \end{aligned}$$

γ_i gesamte Verweildauer in i , n_{ij} Gesamtzahl der Übergänge $i \rightarrow j$.

Loglikelihood: ($p_{i_0}(0)$ kann ignoriert werden)

$$\begin{aligned}
 l(\lambda_{ij}) &= \sum_{i \in S} -\lambda_i \gamma_i + \sum_{i, j \in S, i \neq j} n_{ij} \log(\lambda_{ij}) \\
 &\stackrel{\lambda_i = \sum_{j \neq i} \lambda_{ij}}{=} \sum_{i, j; i \neq j} (-\lambda_{ij} \gamma_i + n_{ij} \log(\lambda_{ij})) \\
 \frac{\partial l(\lambda_{ij})}{\partial \lambda_{ij}} &= -\gamma_i + \frac{n_{ij}}{\lambda_{ij}} \stackrel{!}{=} 0 \\
 \Rightarrow \hat{\lambda}_{ij} &= \frac{n_{ij}}{\gamma_i} \\
 &= \frac{\text{Anzahl der Übergänge von } i \text{ nach } j}{\text{gesamte Verweildauer in } i}
 \end{aligned}$$

Es gilt: $\hat{\lambda}_{ij}$ ist konsistent für λ_{ij} mit asymptotischer Varianz $\frac{n_{ij}}{\gamma_i^2}$.

Korollar:

Sei $n_i = \sum_j n_{ij}$ die Anzahl der Sprünge weg von i . Dann ist $\hat{\lambda}_i = \frac{n_i}{\gamma_i}$ und $\hat{q}_{ij} = \frac{n_{ij}}{n_i}$.

Beweis:

Invarianz des ML-Schätzers bzgl. Reparametrisierung; hier $\lambda_i = \sum_{j \neq i} \lambda_{ij}$

$$\Rightarrow \hat{\lambda}_i = \sum_{j \neq i} \hat{\lambda}_{ij} = \sum_{j \neq i} \frac{n_{ij}}{\gamma_i} = \frac{n_i}{\gamma_i}.$$

Schließlich folgt für ÜW der eingebetteten MK Y :

$$\hat{q}_{ij} = \frac{\hat{\lambda}_{ij}}{\hat{\lambda}_i} = \frac{n_{ij}/\gamma_i}{n_i/\gamma_i} = \frac{n_{ij}}{n_i}$$

□

Beispiel: (aus Guttorp) Pavian Stamm im Amboseli NP, Kenya

Veränderung der Mitgliederanzahl durch Geburt (B), Einwanderung (I), Auswanderung (E), Tod (D).
Annahme eines Geburts-Tod-Prozesses. Tabelle 1-4.

Tabelle 1: Ereignisse im Pavian Stamm

nach so vielen Tagen	Anzahl	Ereignis
41	40	B
5	41	B
22	42	B
2	43	D
17	42	D
26	41	I
0	42	I
55	43	B
35	44	I
20	45	E
5	44	D
6	43	E
32	42	D
4	41	D
0	40	D
22	39	D
10	38	B
0	39	B
7	40	D
4	39	B
17	40	D
11	39	E
3	38	B
4	39	D
8	38	D
2	37	D
5	36	B
10	37	B

Tabelle 2

i, j	n_{ij}	$\hat{\lambda}_{ij}$	i, j	n_{ij}	$\hat{\lambda}_{ij}$
36, 37		$\frac{1}{5} = 0.2$	37, 36		$\frac{1}{12}$
37, 38		$\frac{1}{12}$	38, 37		$\frac{1}{21}$
38, 39		$\frac{2}{21}$	39, 38		$\frac{3}{41}$
39, 40		$\frac{2}{41}$	40, 39		$\frac{3}{65}$
40, 41		$\frac{1}{65}$	41, 40		$\frac{1}{35}$
41, 42		$\frac{2}{35}$	42, 41		$\frac{2}{71}$
42, 43		$\frac{2}{71}$	43, 42		$\frac{2}{63}$
43, 44		$\frac{1}{63}$	44, 43		$\frac{1}{40}$
44, 45		$\frac{1}{40}$	45, 44		$\frac{1}{20}$

Tabelle 3: γ_i

$\gamma_{36} =$	5	= 5
$\gamma_{37} =$	2+10	= 12
$\gamma_{38} =$	10+3+8	= 21
$\gamma_{39} =$	22+0+4+11+4	= 41
$\gamma_{40} =$	41+0+7+17	= 65
$\gamma_{41} =$	5+26+4	= 35
$\gamma_{42} =$	22+17+0+32	= 71
$\gamma_{43} =$	2+55+6	= 63
$\gamma_{44} =$	35+5	= 40
$\gamma_{45} =$	20	= 20

Tabelle 4: $\hat{\lambda}_i$

$\hat{\lambda}_{36} =$	$\frac{1}{5}$	= 0.2
$\hat{\lambda}_{37} =$	$\frac{2}{12}$	= 0.17
$\hat{\lambda}_{38} =$	$\frac{3}{21}$	= 0.14
$\hat{\lambda}_{39} =$	$\frac{5}{41}$	= 0.12
$\hat{\lambda}_{40} =$	$\frac{4}{65}$	= 0.06
$\hat{\lambda}_{41} =$	$\frac{3}{35}$	= 0.09
$\hat{\lambda}_{42} =$	$\frac{4}{71}$	= 0.06
$\hat{\lambda}_{43} =$	$\frac{3}{63}$	= 0.03
$\hat{\lambda}_{44} =$	$\frac{2}{40}$	= 0.025
$\hat{\lambda}_{45} =$	$\frac{1}{20}$	= 0.05

Situation 2

Daten $X(t_1) = i_1, \dots, X(t_n) = i_n$

\Rightarrow Likelihood

$$L(\lambda_{ij} \mid X(t_1) = i_1, \dots, X(t_n) = i_n) = p_{i_0}(0) p_{i_1 i_2}(t_2 - t_1) \cdot \dots \cdot p_{i_{n-1} i_n}(t_n - t_{n-1})$$

Die ÜW auf der rechten Seite müssen dabei aus den Kolmogorov-Gleichungen bzw. über die Spektralzerlegung von Λ bestimmt werden \Rightarrow numerisch aufwendig.

Also: Falls möglich, dieses Studiendesign vermeiden; gesamten Pfadverlauf retrospektiv erfragen. Dies ist allerdings in Beobachtungsstudien nicht immer möglich.

Kapitel 5

Erneuerungs- und Semi-Markov-Prozesse

Für den Poisson-Prozess und (reguläre) diskrete Markov-Prozesse impliziert die Markov-Eigenschaft, dass die Zwischenzeiten bzw. Verweildauern exponentialverteilt sind. Bei Erneuerungs- und Semi-Markov-Prozessen dürfen diese beliebig verteilt sein. Die Markov-Eigenschaft geht damit verloren bzw. gilt nur noch in eingeschränktem Sinn.

Ziel des Kapitels ist eine kurze Einführung in die Struktur dieser Prozesse. Details, insbesondere zur „klassischen“ Erneuerungstheorie, und Beweise finden sich in FKO, Kapitel 4 und 5.

5.1 Erneuerungsprozesse

Die Pfade sind Treppenfunktionen mit Sprunghöhe +1 wie beim Poisson-Prozess; statt Exponentialverteilungen für die Verweildauern T_n sind nun beliebige Verteilungen zugelassen:

T_n iid $\sim F(t)$, nichtnegativ. Interpretation und Anwendungsbeispiele analog zum Poisson-Prozess.

5.1.1 Definition und grundlegende Begriffe

Definition 5.1 Erneuerungsprozess

Sei $\{T_n, n \in \mathbb{N}\}$ eine Folge unabhängiger, nichtnegativer Zufallsvariablen mit gleicher Verteilungsfunktion F , für die $F(0) < 1$ gilt. Dann heißt die Folge $\{S_n, n \in \mathbb{N}_0\}$ mit

$$S_0 := 0, \quad S_{n+1} = S_n + T_{n+1}, \quad n = 0, 1, 2, \dots$$

Erneuerungsprozess (EP). Sei

$$N(t) = \max_{n \in \mathbb{N}_0} \{n : S_n \leq t\}$$

die Anzahl von Erneuerungen im Intervall $[0, t]$. Dann heißt $N = \{N(t), t \geq 0\}$ *Erneuerungszählprozess*, oft auch einfach Erneuerungsprozess.

Beispiele, Anwendungen

(a) Zuverlässigkeitstheorie

Ein bestimmter Bestandteil eines technischen Gerätes werde sofort nach seinem Ausfall durch ein neues Gerät ersetzt, usw. Die T_1, T_2, \dots sind dann die Lebensdauern des ersten, zweiten, ... Gerätes. $N(t)$ gibt die Anzahl von Erneuerungen an.

Denkt man statt an ein technisches Gerät an andere Systeme, so sind viele weitere Interpretationen möglich.

(b) Marktforschung

Ein bestimmtes Produkt wird nach seinem Verbrauch erneut gekauft, usw. Der Erneuerungsprozess heißt in diesem Zusammenhang Wiederholkaufprozess.

(c) Sachversicherung

S_1, S_2, \dots sind die Zeitpunkte, zu denen Schadensfälle auftreten. $N(t)$ ist die Anzahl der Schadensfälle.

(d) Warteschlangen, Lagerhaltung

Der Ankunftsprozess von Kunden kann in Verallgemeinerung zu den entsprechenden Beispielen

des letzten Kapitels einen Erneuerungsprozess bilden. Auch vom Bedarf in Lagerhaltungssystemen wird oft angenommen, dass er einen Erneuerungsprozess bildet; dabei ist T_n der Bedarf in der n -ten Periode.

(e) Klinische Studien, Survival-Analyse

In klinischen Studien sind T_1, T_2, \dots z.B. die Zeitdauern nach denen epileptische Anfälle, Rückfälle, etc. auftreten. Ein wichtiger Spezialfall ist die Survival-oder Überlebenszeit-Analyse: Hier wird für (homogene) Gruppen von Patienten jeweils die Überlebenszeit T_1 bis zum Ende (Tod) oder auch nur zensiert beobachtet, d.h. das Ereignis Tod ist bis zum Ende der Studie noch nicht eingetreten (vgl. Fahrmeir, Hamerle und Tutz, 1996, Kap.7).

Vor allem im Kontext des ersten Beispiels nennt man die T_n *Lebensdauern*, die S_n *Erneuerungszeitpunkte*. Dabei wird S_0 *nicht* zu den Erneuerungszeitpunkten gezählt (in der Literatur ist dies gelegentlich der Fall). Für die Lebensdauer ist auch $P\{T_n = \infty\} > 0$, also $F(\infty) := \lim_{t \rightarrow \infty} F(t) < 1$, zugelassen. Die Annahme, dass T_1 dieselbe Verteilung wie T_2, T_3, \dots besitzt, kann so gedeutet werden, dass zum Zeitpunkt $t = 0$ gerade eine Erneuerung stattgefunden hat (die aber vereinbarungsgemäß nicht gezählt wird). Besitzt T_1 eine von T_2, T_3, \dots verschiedene Verteilung, so spricht man vom *modifizierten* Erneuerungsprozess.

Die T_n können diskrete oder stetige ZV sein. Im diesem Kapitel werden nur stetige ZV betrachtet, d.h. die Lebensdauerverteilung F besitzt eine Dichte f .

Beispiele:

(a) Die Exponentialverteilung ergibt den Spezialfall des Poisson-Prozesses.

(b) Weibull-Verteilung

Dichte:

$$f(t) = \lambda \alpha (\lambda t)^{\alpha-1} \exp(-(\lambda t)^\alpha), \quad t \geq 0$$

wobei $\lambda > 0$ und $\alpha > 0$ Parameter der Verteilung sind.

Für $\alpha = 1$ ergibt sich die Exponentialverteilung als Spezialfall.

$$\begin{aligned} E(T) &= \frac{1}{\lambda} \Gamma\left(\frac{1+\alpha}{\alpha}\right) \\ \text{Var}(T) &= \frac{1}{\lambda^2} \left[\Gamma\left(\frac{\alpha+2}{\alpha}\right) - \left\{ \Gamma\left(\frac{\alpha+1}{\alpha}\right) \right\}^2 \right] \end{aligned}$$

(c) Gammaverteilung

Dichte

$$f(t) = \frac{\lambda^\alpha}{\Gamma(\alpha)} \cdot t^{\alpha-1} \exp(-\lambda t), \quad \alpha > 0, \lambda > 0, t > 0.$$

Für $\alpha = 1$ ergibt sich wiederum die Exponentialverteilung als Spezialfall.

$$\begin{aligned} E(T) &= \frac{\alpha}{\lambda} \\ \text{Var}(T) &= \frac{\alpha}{\lambda^2} \\ \text{Modus} &= \frac{(\alpha - 1)}{\lambda}, \quad \text{für } \alpha > 1 \end{aligned}$$

(d) Die log-Normalverteilung

Sie ergibt sich aus einer normalverteilten ZV

$X \sim N(\mu, \sigma^2)$ durch Exp-Transformation:

$$T = \exp(X), \quad X \sim N(\mu, \sigma^2)$$

Es gilt:

$$\begin{aligned} E(T) &= \exp\left(\mu + \frac{1}{2}\sigma^2\right) \\ \text{Var}(T) &= \exp(2\mu + \sigma^2) \cdot [\exp(\sigma^2) - 1] \\ f(t) &= \frac{1}{t} \frac{1}{\sigma} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} \frac{(\log(t) - \mu)^2}{\sigma^2}\right), \quad t > 0. \end{aligned}$$

Man definiert ferner die Survivorfunktion und insbesondere die Hazardrate:

Definition 5.2 Survivorfunktion und Hazardrate

Die *Survivorfunktion*

$$S(t) = 1 - F(t)$$

ist die unbedingte Wahrscheinlichkeit bis zum Zeitpunkt t zu überleben.

Die *Hazardrate* ist

$$\lambda(t) = \lim_{h \downarrow 0} \frac{P(t \leq T \leq t + h \mid T \geq t)}{h},$$

also die infinitesimale Rate für einen Ausfall zum Zeitpunkt t, gegeben Überleben bis zum Zeitpunkt t.

In $o(h)$ -Schreibweise: Die Wahrscheinlichkeit für einen Ausfall im Intervall $[t, t+h]$, gegeben Überleben bis zum Zeitpunkt t, ist

$$P(t \leq T \leq t + h \mid T \geq t) = \lambda(t)h + o(h).$$

Das Integral

$$\Lambda(t) = \int_0^t \lambda(u) du$$

wird als *kumulierte Hazardrate* bezeichnet.

Beispiel:

- (a) Exponentialverteilung: $\lambda(t) = \lambda$
 (b) Weibull-Verteilung: $\lambda(t) = \lambda\alpha(\lambda t)^{\alpha-1}$

Es gelten folgende

Zusammenhänge

$$\begin{aligned}\lambda(t) &= \frac{f(t)}{1 - F(t)} = \frac{f(t)}{S(t)} \\ S(t) &= \exp \left[- \int_0^t \lambda(u) du \right] = \exp(-\Lambda(t)) \\ f(t) &= \lambda(t) \cdot S(t) = \lambda(t) \exp \left[- \int_0^t \lambda(u) du \right]\end{aligned}$$

Beweis:

Die erste Beziehung erfolgt direkt aus der Definition der Hazardrate. Die zweite Gleichung folgt wegen

$$\int_0^t \lambda(u) du = \int_0^t \frac{f(u)}{1 - F(u)} du = [-\ln(1 - F(u))]_0^t = -\ln(1 - F(t)).$$

□

5.1.2 Zur Theorie der Erneuerungsprozesse

Die klassische Theorie der EP ist insbesondere interessiert an der Verteilung von S_n und $N(t)$, sowie der erwarteten Zahl $R(t) = E(N(t))$ von Erneuerungen bis t . Es gelten folgende Sätze (FKO S131-133):

Satz 5.1 Verteilung von S_n und $N(t)$

$$F_n(t) := P\{S_n \leq t\} = \int_{[0,t]} F_{n-1}(t-x) dF(x), \quad t \geq 0,$$

mit $F_1(t) := F(t)$; in der Schreibweise mit dem Faltungssymbol also

$$F_n = F * F_{n-1} = F^{n*} = \underbrace{F * \dots * F}_{n\text{-mal}}.$$

$$P\{N(t) = n\} = F_n(t) - F_{n+1}(t), \quad n = 0, 1, \dots; \quad F_0(t) \equiv 1.$$

Für die Dichte $f_n(t)$ von $F_n(t)$ gilt

$$f_n(t) = \int_0^t f_{n-1}(t-x) f(x) dx, \quad f_1(t) := f(t)$$

Es ist

$$E(N(t)) = R(t) = \sum_{n=1}^{\infty} F_n(t).$$

Analytisch ist die Faltung nur in Spezialfällen lösbar. Da die numerische Integration sehr aufwendig wird, sind folgende asymptotischen Ergebnisse (für $t \rightarrow \infty$) nützlich. Dabei werden *rekurrente* EP vorausgesetzt, d.h. es muss gelten

$$\lim_{t \rightarrow \infty} F(t) = 1.$$

Ansonsten treten mit positiver Wahrscheinlichkeit ab einem gewissen Erneuerungszeitpunkt keine weiteren Erneuerungen mehr auf. Im folgenden sei für rekurrente EP mit stetigen T_n

$$\mu = E(T_n) = \int_0^{\infty} x f(x) dx = \int_0^{\infty} (1 - F(x)) dx.$$

Dann gilt

Satz 5.2 Grenzwertsätze

Für einen rekurrenten EP mit stetigen T_n ,

$$\mu = E(T_n), \quad \sigma^2 = \text{Var}(T_n) < \infty \text{ gelten :}$$

1. $\lim_{t \rightarrow \infty} \frac{N(t)}{t} = \frac{1}{\mu}$ m. Wkeit 1
2. $\lim_{t \rightarrow \infty} \frac{R(t)}{t} = \frac{1}{\mu}$, *elementares Erneuerungstheorem*

$$\lim_{t \rightarrow \infty} (R(t) - \frac{t}{\mu}) = \frac{\sigma^2 - \mu^2}{2\mu^2},$$

$$3. \lim_{t \rightarrow \infty} P \left(\frac{N(t) - \frac{t}{\mu}}{\frac{\sigma}{\mu} \sqrt{\frac{t}{\mu}}} \leq x \right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \exp(-\frac{t^2}{2}) dt,$$

d.h. $N(t)$ ist für große t approximativ normalverteilt mit Erwartungswert $\frac{t}{\mu}$ und der Varianz $\frac{\sigma^2 t}{\mu^3}$.

Aussage 1. bzw. 2. enthalten das plausible Resultat, dass auf lange Sicht die (durchschnittliche) Anzahl von Erneuerungen pro Zeiteinheit gleich dem Kehrwert der erwarteten Dauer $\mu = E(T_n)$ ist.

5.2 Semi-Markov-Prozesse**5.2.1 Definition und grundlegende Eigenschaften**

Die Struktur eines typischen Pfades eines Semi-Markov-Prozesses gleicht der eines Markov-Prozesses, mit der Ausnahme, dass die Verweilzeiten nicht notwendig exponentialverteilt sein müssen.

- Y_n Zustand nach dem n -ten Übergang, $n = 0, 1, \dots$,
- T_{n+1} Verweildauer im Zustand Y_n
- S_n Zeitpunkt des n -ten Übergangs.

Die Folge $\{Y_n, n \in \mathbb{N}_0\}$ der Zustände soll eine Markov-Kette bilden, während die Zeitpunkte S_n , zu denen ein Übergang erfolgt, und damit die Verweildauer durch einen gesonderten Wahrscheinlichkeitsmechanismus beschrieben werden. Dies führt zur

Definition 5.3 Markov-EP, Semi-Markov-Prozess

$\{Y_n, n \in \mathbb{N}_0\}$ sei eine Folge von ZV mit abzählbarem Zustandsraum
 $\{S_n, n \in \mathbb{N}_0\}$ eine Folge nichtnegativer ZV mit $0 =: S_0 \leq S_1 \leq S_2 \leq \dots$
 Der SP $(Y, S) = \{(Y_n, S_n), n \in \mathbb{N}_0\}$ heißt *Markov-Erneuerungsprozess* (MEP) : \Leftrightarrow
 $\forall t \geq s \geq s_{n-1} \geq \dots \geq 0, \forall j, i_n, \dots, i_0 \in S$

$$(SM) \quad P\{Y_{n+1} = j, S_{n+1} - S_n \leq t \mid Y_n = i, Y_{n-1} = i_{n-1}, \dots, Y_0 = i_0, S_n = s, S_{n-1} = s_{n-1}, \dots\} \\ = P\{Y_{n+1} = j, S_{n+1} - S_n \leq t \mid Y_n = i\}.$$

Ist (Y, S) ein MEP, so heißt $X = \{X(t), t \geq 0\}$ der zu (Y, S) gehörige *Semi-Markov-Prozess* (SMP): \Leftrightarrow

$$X(t) = Y_n \quad \text{für } S_n \leq t < S_{n+1}, \quad n = 0, 1, 2, \dots$$

Interpretation: Die Markov-Eigenschaft gilt nur noch bei Bedingen auf $X(S_n) = i$, d.h. bedingt auf $t = S_n$; nicht jedoch für beliebige Zeitpunkte. Die Eigenschaft (SM) heißt deshalb Semi-Markov-Eigenschaft.

Definition 5.4 Homogener SMP

Gilt außerdem

$$P\{Y_{n+1} = j, S_{n+1} - S_n \leq t \mid Y_n = i\} = P\{Y_1 = j, S_1 \leq t \mid Y_0 = i\} =: Q_{ij}(t)$$

unabhängig von n , so heißt der MEP bzw. SMP *homogen*. Die bedingten Wahrscheinlichkeiten $Q_{ij}(t)$ heißen *Semi-Markov-Übergangswahrscheinlichkeiten* (SÜW), $Q(t) = (Q_{ij}(t))$ *Semi-Markov-Übergangsmatrix* (SÜM).

Vereinbarung: Im weiteren werden nur *homogene* SMP betrachtet und zusätzlich gefordert, dass für die Verteilung

$$Q_i(t) := \sum_{j \in S} Q_{ij}(t) = P\{T_{n+1} \leq t \mid Y_n = i\} = P\{S_1 \leq t \mid Y_0 = i\}$$

der Verweilzeit in i gilt:

$$Q_i(0) < 1 \text{ und } \lim_{t \rightarrow \infty} Q_i(t) = 1.$$

Also: Der SMP verbleibt eine positive Zeitspanne in i (dies entspricht der Forderung $F(0) < 1$ bei EP)

und verlässt m. Wkeit 1 nach endlicher Zeit i wieder (dies entspricht $F(\infty) = 1$ für einen rekurrenten EP).

Bemerkung: Alternativ können die Verweildauerverteilungen auch durch Hazardraten charakterisiert werden:

$$\lambda_i(t) = \frac{q_i(t)}{1 - Q_i(t)}, \quad q_i(t) \text{ Dichte.}$$

Satz 5.3 Struktur von SMP

$\{Y_n, n \in \mathbb{N}_0\}$ ist eine homogene MK mit Übergangsmatrix $Q = (q_{ij})$,

wobei

$$q_{ij} = \lim_{t \rightarrow \infty} Q_{ij}(t).$$

Die bedingte Verteilungsfunktion der Verweildauer T_{n+1} bedingt auf den Zustand $Y_n = i$ und $Y_{n+1} = j$ ist $G_{ij}(t) = P(T_{n+1} \leq t \mid Y_n = i, Y_{n+1} = j) = \frac{Q_{ij}(t)}{q_{ij}}$. (Für $q_{ij} = 0$ setze $G_{ij}(t) = 0$.)

Daraus kann man leicht einen *Algorithmus zur Simulation von SMP* herleiten:

1. Simuliere die MK Y gemäß ÜM (q_{ij})
2. Simuliere dann die Verweildauer T gemäß der Verteilung $G_{ij}(t)$.

Bemerkung: Alternativ kann ein SMP auch über die (abgangsspezifischen) Hazardraten

$$\begin{aligned} \lambda_{ij}(t) &= \lim_{h \downarrow 0} \frac{P(Y_{n+1} = j, t \leq T_{n+1} \leq t + h \mid Y_n = i, T_{n+1} \geq t)}{h} \\ &= \lim_{h \downarrow 0} \frac{P(Y_{n+1} = j, t \leq T_{n+1} \leq t + h \mid Y_n = i)}{h \cdot P(T_{n+1} \geq t \mid Y_n = i)} \\ &= \frac{q_{ij}(t)}{1 - Q_i(t)}, \quad \text{mit } q_{ij}(t) : \text{„Dichte“ zu } Q_{ij}(t), \text{ d.h.} \\ q_{ij}(t) &= \frac{\partial Q_{ij}(t)}{\partial t} \end{aligned}$$

charakterisiert werden.

Beispiel:

MEP mit

$$Q(t) = (Q_{ij}(t)) = \begin{pmatrix} 0.6(1 - \exp(-5t)) & 0.4(1 - \exp(-2t)) \\ 0.5 - 0.2 \exp(-3t) - 0.3 \exp(-5t) & 0.5 - 0.5 \exp(-2t) - t \exp(-2t) \end{pmatrix}$$

1. Verteilungsfunktion der Verweildauer in i

$$\begin{aligned} Q_1(t) &= 1 - 0.6 \exp(-5t) - 0.4 \exp(-2t) \\ Q_2(t) &= 1 - 0.2 \exp(-3t) - 0.5 \exp(-2t) \\ &\quad - 0.3 \exp(-5t) - t \exp(-2t) \end{aligned}$$

2. Übergangsmatrix der eingebetten MK

$$Q = \lim_{t \rightarrow \infty} Q(t) = \begin{pmatrix} 0.6 & 0.4 \\ 0.5 & 0.5 \end{pmatrix}$$

3. Bedingte Verteilungsfunktionen

$$\begin{aligned} G_{ij}(t) &= P(T_{n+1} \leq t \mid Y_n = i, Y_{n+1} = j) \\ &= \frac{Q_{ij}(t)}{q_{ij}} \end{aligned}$$

$$G_{11}(t) = 1 - \exp(-5t)$$

$$G_{12}(t) = 1 - \exp(-2t)$$

$$G_{21}(t) = 1 - \frac{2}{5} \exp(-3t) - \frac{3}{5} \exp(-5t)$$

$$G_{22}(t) = 1 - \exp(-2t) - 2t \exp(-2t)$$

4. Erwartete Verweildauern

$$\mu_i = \int_0^{\infty} [1 - Q_i(t)] dt$$

$$\begin{aligned} \mu_1 &= \int_0^{\infty} 0.6 \exp(-5t) dt + \int_0^{\infty} 0.4 \exp(-2t) dt \\ &= 0.6 \left[-\frac{1}{5} \exp(-5t) \right]_0^{\infty} + 0.4 \left[-\frac{1}{2} \exp(-2t) \right]_0^{\infty} \\ &= 0.6 \cdot \frac{1}{5} + 0.4 \cdot \frac{1}{2} \\ &= \frac{6}{50} + \frac{4}{20} \\ &= \frac{12}{100} + \frac{20}{100} = \frac{32}{100} = 0.32 \end{aligned}$$

$$\mu_2 = 0.626$$

5. Hazardrate

$$\lambda_{ij}(t) = \frac{q_{ij}(t)}{1 - Q_i(t)}$$

z.B.

$$\begin{aligned} q_{12}(t) &= \frac{\partial Q_{12}(t)}{\partial t} \\ &= 0.4 \cdot 2 \exp(-2t) \\ &= 0.8 \exp(-2t). \end{aligned}$$

$$Q_1(t) = 1 - 0.6 \cdot \exp(-5t) - 0.4 \cdot \exp(-2t).$$

$$\begin{aligned} \Rightarrow \lambda_{12}(t) &= \frac{0.8 \exp(-2t)}{0.6 \exp(-5t) + 0.4 \exp(-2t)} \\ &= \frac{2}{1 + 1.5 \exp(-3t)}. \end{aligned}$$

Bemerkung MK, MP, EP als Spezialfälle

(a) Ein Erneuerungsprozess ergibt sich mit $S = \{1\}$ und daher $Q_{11}(t) = F(t)$.

(b) Eine Markov-Kette erhält man durch die Wahl

$$Q_{ij}(t) = \begin{cases} 0 & \text{für } t < 1 \\ q_{ij} & \text{für } t \geq 1 \end{cases}$$

(da $q_{ij} = \lim_{t \rightarrow \infty} Q_{ij}(t) \Rightarrow T_1 = T_2 = \dots = 1$ deterministische Verweildauern.

(c) Mit $G_{ij}(t) = 1 - \exp(-\lambda_i t)$, $t \geq 0$

$$q_{ii} = \begin{cases} 0 & , i \text{ stabil} \\ 1 & , i \text{ absorbierend} \end{cases}$$

erhält man einen MP.

5.2.2 Grenzwertsätze

Rekurrenzeigenschaften und Irreduzibilität eines Semi-Markov-Prozesses leiten sich aus den entsprechenden Eigenschaften der Markovkette Y ab.

Der SMP heißt aperiodisch, wenn für alle i die Zwischenzeiten T_{ii} zwischen aufeinanderfolgenden Besuchen in i nicht gitterförmig sind, d.h. es gibt kein $d \geq 0$ mit $\sum_{n=0}^{\infty} P(T_{ii} = nd) = 1$. (Dies ist insbesondere bei stetigen Verweildauern erfüllt.)

Es gelten ähnliche Aussagen wie bei MP, wobei der Erwartungswert $\frac{1}{\lambda_i}$ von exponentialverteilten Verweildauern durch den Erwartungswert

$$\mu_i = \int_0^{\infty} tq_i(t)dt = \int_0^{\infty} (1 - Q_i(t))dt$$

ersetzt wird. Zum Beispiel gilt

Satz 5.4 Grenzwertsatz

Der SMP sei aperiodisch, irreduzibel und positiv-rekurrent. Dann gilt für die Zustandswahrscheinlichkeit $p_j(t) = P(X(t) = j)$

$$\pi_j := \lim_{t \rightarrow \infty} p_j(t) = \frac{\nu_j \mu_j}{\sum_{i \in S} \nu_i \mu_i}$$

(unabhängig von der Anfangsverteilung), wobei ν mit

$$\nu = \nu Q$$

ein zur ÜM Q von Y gehöriges strikt positives invariantes Maß (d.h. bis auf Normierung die stationäre Grenzverteilung von Y) ist.

Eine Plausibilitätserklärung verläuft wie für MP.

5.3 Bemerkung zur statistischen Inferenz

Im Vergleich zu MP sind statt Parametern λ_i der Exponentialverteilungen die entsprechenden Verweildauern parametrisch oder nichtparametrisch zu schätzen sowie die ÜM Q der eingebetteten MK. Die moderne Inferenz benutzt das Instrumentarium von Zählprozessen (vgl. Kap.7) um die (Übergangsspezifischen) Hazardraten $\lambda_{ij}(t)$ als Funktion von t zu schätzen.