Kapitel 3

Markov-Ketten

3.1 Grundlegende Eigenschaften, Beispiele

- Diskreter Parameterraum (Zeit) $T = \mathbb{N}_0$.
- Zustandsraum S meist diskret (3.1 - 3.6), manchmal auch stetig (3.7 - 3.8).
- In Beispielen / Anwendungen ist S oft endlich.
- Definition: Markov-Kette 1. Ordnung

Der stochastische Prozess $X = \{X_t, t \in \mathbb{N}_0\}$, S diskret, heißt Markov-Kette (1. Ordnung) $:\Leftrightarrow \forall t \in \mathbb{N}_0, j, i, i_{t-1}, \dots, i_0 \in S$

$$\begin{aligned} P(X_{t+1} = j | X_t = i, X_{t-1} = i_{t-1}, \dots, X_0 = i_0) &= P(X_{t+1} = j | X_t = i) \\ &\iff \\ P(X_{t+1} = j | X_t, X_{t-1}, \dots, X_0) &= P(X_{t+1} = j | X_t) \end{aligned}$$

- Beispiel: Einfache Irrfahrt

$$X_{t+1} = X_t + Z_t, \quad Z_t \in \{-1, 0, +1\}, \quad t = 1, 2, \dots$$

$$P(Z_t = 0 | X_t = i) = r_i,$$

$$P(Z_t = -1 | X_t = i) = q_i,$$

$$P(Z_t = +1 | X_t = i) = p_i,$$

$$p_i + r_i + q_i = 1$$

Die Wahrscheinlichkeitsfunktion von X_{t+1} gegeben den bisherigen Pfad

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

hängt nur vom gegenwärtigen Zustand $X_t = i$ ab.

- Bemerkungen:

- $P(X_{t+1} = j|X_t)$ ist eine Zufallsvariable mit Werten $P(X_{t+1} = j|X_t = i)$, $i \in S$.
- Verbale Interpretation der Markov-Eigenschaft:

Bei gegebener Gegenwart hängt die Zukunft der Markov-Kette nicht von der Vergangenheit ab.

Bei gegebener Gegenwart sind Zukunft und Vergangenheit bedingt unabhängig.

- Definition: Übergangswahrscheinlichkeiten

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

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

- Definition: Homogene Markov-Kette

Eine Markov-Kette heißt homogen bzw. besitzt stationäre Übergangswahrscheinlichkeiten : $\Leftrightarrow$

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

- Bemerkungen:

- Im Folgenden betrachten wir homogene Markov-Ketten, falls nichts anderes vermerkt.
- Eine inhomogene Markov-Kette ist i.A. kein stationärer stochastischer Prozess.
- Nichtstationäre Übergangswahrscheinlichkeiten können z.B. so modelliert werden:

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

$\implies$ Computerintensive Verfahren, Generalisierte Regression.

- Definition: Übergangsmatrix

Für homogene Markov-Ketten heißt

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

Übergangsmatrix. Für inhomogene Markov-Ketten ist die Übergangsmatrix $P(t) = (p_{ij}(t))$ zeitabhängig.

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

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

- P ist endlich, falls S endlich ist.

- Eigenschaften der Übergangsmatrix

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

- Folgerungen aus der Definition einer Markov-Kette:

1. $\forall t, s \geq 1, i_0, \dots, i_s \in S$ gilt

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

Ist zusätzlich eine Anfangsverteilung

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

gegeben, so gilt

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

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

$$P(X_{t+s} = i_{t+s}, \dots, X_{t+1} = i_{t+1} | X_t = i_t, \dots, X_0 = i_0) = \\ P(X_{t+s} = i_{t+s}, \dots, X_{t+1} = i_{t+1} | X_t = i_t)$$

$$3. \forall 0 \leq t_0 < t_1 < \dots < t_k$$

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

- Bemerkungen:

- Gelegentlich wird 3. zur Definition von Markov-Ketten benutzt.

- Aus 1.-3. lassen sich alle endlich-dimensionalen Verteilungen einer Markov-Kette berechnen, so dass der Satz von Kolmogorov anwendbar ist.

⇒ Bei Vorgabe einer Übergangsmatrix P und einer Startverteilung $p(0)$ existiert ein zugehöriger stochastischer Prozess mit Wahrscheinlichkeitsraum.

- Zugleich lässt sich die Likelihood berechnen, vgl. Statistische Inferenz in 3.5.

- Definition: Mehrschrittige Übergangswahrscheinlichkeiten

$$p_{ij}^{(t)} := P(X_{t+s} = j | X_s = i) = P(X_t = j | X_0 = i), \quad t \geq 1$$

heißt t -schrittige Übergangswahrscheinlichkeit von i nach j .

- Chapman–Kolmogorov–Gleichungen

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

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

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

- Definition: Zustandswahrscheinlichkeit

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

heißen Zustandswahrscheinlichkeiten. Der Zeilenvektor

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

heißt Zustandsverteilung. Es gilt

$$p_i(t) = \sum_{j \in S} p_j(0) p_{ji}^{(t)}$$

bzw.

$$p(t) = p(0)P^t.$$

- Darstellung von Markov-Ketten durch gerichtete Graphen.

- Markov-Ketten höherer Ordnung

Der stochastische Prozess $X = \{X_t, t \in N_0\}$, S diskret, heißt Markov-Kette p -ter Ordnung $:\Leftrightarrow \forall i_0, i_1, \dots, i_{t+1} \in S, \quad t \geq p + 1$

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

- Modellierung z.B. durch konditionale Logit-Modelle

$$\log \left(\frac{P(X_{t+1} = i_{t+1} | X_t, \dots, X_{t-r+1})}{1 - P(X_{t+1} = i_{t+1} | X_t, \dots, X_{t-r+1})} \right) = f(X_t, \dots, X_{t-r+1})$$

3.2 Klassifizierung von Zuständen und Rückkehrverhalten

- Definition: Erreichbarkeit

1. Der Zustand j heißt von i aus erreichbar ($i \rightarrow j$) : $\Leftrightarrow$

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

d.h. es muss ein Weg im Markov-Graphen von i nach j führen. Eine direkte Kante $i \rightarrow j$ ist aber i.A. nicht notwendig!

2. Die Zustände i und j heißen gegenseitig erreichbar ($i \leftrightarrow j$) : $\Leftrightarrow$

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

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

- Klassifizierung nach Erreichbarkeit:
 1. $C \subseteq S$ 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 $:\Leftrightarrow \{i\}$ abgeschlossen, d.h. $p_{ii} = 1$.
 3. Ein abgeschlossenes C heißt irreduzibel $:\Leftrightarrow$ keine echte, nichtleere Teilmenge von C ist abgeschlossen
 4. Eine Markov-Kette heißt irreduzibel $:\Leftrightarrow S$ irreduzibel $\Leftrightarrow$ alle Zustände sind gegenseitig erreichbar.
- Damit zerfällt S in offene und abgeschlossene, irreduzible Teilmengen wechselseitig erreichbarer Zustände. Bei einer Markov-Kette mit endlich vielen Zuständen existiert mindestens eine abgeschlossene irreduzible Klasse C . Eine offene Klasse kann existieren, muss aber nicht.

- Klassifizierung nach Rückkehrverhalten ($i \in S$):

Es sei T_{ii} die zufällige Zeit bis zur ersten Rückkehr nach i bei einem Start in i , und $E(T_{ii}|X_0 = i) = \mu_{ii}$ die erwartete Rückkehrzeit.

1. i heißt rekurrent $:\Leftrightarrow$

$$P(T_{ii} < \infty | X_0 = i) = 1$$

(mit Wahrscheinlichkeit 1 Rückkehr in endlicher Zeit).

2. i heißt transient $:\Leftrightarrow$

$$P(T_{ii} < \infty | X_0 = i) < 1$$

(mit positiver Wahrscheinlichkeit keine Rückkehr).

3. Bei rekurrenten Zuständen kann man weiter unterscheiden in

positiv-rekurrent $:\Leftrightarrow \mu_{ii} < \infty$

null-rekurrent $:\Leftrightarrow \mu_{ii} = \infty$

4. i heißt periodisch mit Periode $d : \Leftrightarrow d \geq 2$ größte natürliche Zahl mit

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

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

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

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

- Für eine Markov-Kette mit endlichem S gilt für Zustand i und Äquivalenzklasse C
 1. i positiv-rekurrent $\Leftrightarrow i$ rekurrent (keine Fallunterscheidung).
 2. C abgeschlossen $\Leftrightarrow$ alle Zustände aus C (positiv-)rekurrent.
 3. C offen $\Leftrightarrow$ alle Zustände aus C transient.

- Resultate:

1. Rekurrenz, Transienz, Periodizität und Ergodizität sind Klasseneigenschaften, d.h. alle Zustände einer Klasse sind rekurrent, transient, periodisch oder ergodisch.
2. i rekurrent und $i \rightarrow j \Rightarrow i \leftrightarrow j$ und auch j rekurrent.
3. i rekurrent $\Rightarrow$ es existiert eine irreduzible Klasse $C(i)$ von rekurrenten Zuständen mit $i \in C(i)$.
4. C irreduzible, endliche Klasse $\Rightarrow$ alle Zustände positiv rekurrent
5. C endliche Klasse $\Rightarrow$
 C rekurrent $\Leftrightarrow C$ abgeschlossen
 C transient $\Leftrightarrow C$ offen
Für endliches S gilt also: offen=transient und abgeschlossen=rekurrent
6. MK endlich $\Rightarrow$ es existiert mindestens eine rekurrente Klasse.

- 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 Übergangsmatrix P , so gilt

$$P = \begin{pmatrix} P_1 & 0 & \dots & \dots & 0 \\ 0 & P_2 & 0 & \dots & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & \dots & \dots & \ddots & 0 \\ L_1 & L_2 & \dots & \dots & Q \end{pmatrix}$$

- Klassifizierung für endliche Markov-Ketten:

- Bestimme (etwa mit Hilfe des Erreichbarkeitsgraphen) die irreduziblen Äquivalenzklassen. Diese sind alle positiv rekurrent, der Rest ist transient.
- Bestimme für jede Klasse die Periode.

3.3 Grenzverhalten homogener Markov-Ketten

- Fragestellung:

$$P^t = (p_{ij}^{(t)}) \rightarrow ? \quad \text{für } t \rightarrow \infty$$
$$p(t) = \{P(X_t = i)\} \rightarrow ? \quad \text{für } t \rightarrow \infty$$

Wir betrachten nur aperiodische Markov-Ketten.

- Grenzwertsatz für die t -schrittigen Übergangswahrscheinlichkeiten $p_{ij}^{(t)}$

Sei j transient oder null-rekurrent. Dann ist $\forall i \in S$

$$\lim_{t \rightarrow \infty} p_{ij}^{(t)} = 0.$$

Sei j positiv-rekurrent und aperiodisch. Dann gilt

$$\begin{aligned}\lim_{t \rightarrow \infty} p_{jj}^{(t)} &= \frac{1}{\mu_{jj}} = \frac{1}{E(T_{jj})} \\ \lim_{t \rightarrow \infty} p_{ij}^{(t)} &= \lim_{t \rightarrow \infty} p_{jj}^{(t)} \quad \forall i \text{ mit } i \leftrightarrow j \\ \lim_{t \rightarrow \infty} p_{ij}^{(t)} &= f_{ij} \frac{1}{\mu_{jj}} =: l_{ij}^{(\infty)} \quad \forall i \in S\end{aligned}$$

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 $f_{ij} = 1$.
- 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:

$$P^\infty = \begin{pmatrix} P_1^\infty & 0 & \cdots & \cdots & 0 \\ 0 & P_2^\infty & 0 & \cdots & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & \cdots & \cdots & \ddots & 0 \\ L_1^\infty & L_2^\infty & \cdots & \cdots & 0 \end{pmatrix}$$

- Interpretation von $p_{ij}^{(\infty)}$ für aperiodische, rekurrente $i \leftrightarrow j$: $p_{ij}^{(\infty)} = \frac{1}{\mu_{jj}}$ ist die relative Häufigkeit der Besuche in j , bezogen auf die Gesamtzahl aller Übergänge (für $t \rightarrow \infty$).
- Problem: Berechnung der Erwartungswerte $E(T_{jj}) = \mu_{jj}$.

- Grenzwertsatz für ergodische Markov-Ketten:

Für eine irreduzible, aperiodische MK gilt: Sind alle Zustände positiv rekurrent, dann besitzt das lineare Gleichungssystem

$$\begin{aligned} \pi_j &= \sum_{i \in S} \pi_i p_{ij}, & j \in S \\ \sum_{j \in S} \pi_j &= 1 \end{aligned} \quad \text{bzw.} \quad \begin{aligned} \pi &= \pi P \\ \pi \mathbf{1} &= 1, \end{aligned}$$

mit dem Zeilenvektor $\pi = (\pi_j; j \in S)$, eine eindeutige, strikt positive Lösung ($\pi_j > 0, \forall j \in S$) und es gilt

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

sowie

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

für jede beliebige Anfangsverteilung $p(0)$.

Umgekehrt gilt auch: Ist das obige Gleichungssystem eindeutig und strikt positiv lösbar, so sind alle Zustände positiv rekurrent.

- Folgerung: Für eine irreduzible, aperiodische Markov-Kette mit endlich vielen Zuständen hat das Gleichungssystem eine eindeutige, strikt positive Lösung.

3.4 Instationäre und inhomogene Markov-Ketten

- Häufige Datensituation in Anwendungen:
 - Y_t diskrete bzw. kategoriale Ziel- (oder Response)variable.
 - $z = (z_1, \dots, z_p)'$ Vektor von Kovariablen
 - $T = \mathbb{N}_0$
- Zeitreihendaten: $(Y_t, Z_t), t = 1, \dots, T$, d.h. ein Pfad eines stochastischen Prozesses Y und eines Kovariablen-Prozesses werden von 1 bis T beobachtet.
- Longitudinaldaten: $\{(Y_{nt}, z_{nt}), t = 1, \dots, T\}$, d.h. jeweils Pfade für $n = 1, \dots, N$ Individuen / Objekte beobachtet.
- Mögliche Situationen:
 - N groß, T klein,
 - N klein, T groß ($N=1$: Zeitreihe),
 - N, T mittel bis groß.

- Datenstruktur der Pfade passt zu einer Markov-Kette für Y , aber
 - Zeitliche Stationarität fraglich $\Rightarrow$ Markov-Kette mit nichtstationären Übergangswahrscheinlichkeiten $p_{ij}(t)$
 - Homogenität der Population fraglich,

$$\Rightarrow P(Y_{n,t-1} = j | Y_{n,t} = i) = p_{ij,n}(t).$$

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.

3.4.1 Instationäre und inhomogene binäre Markov-Kette 1. Ordnung

- Modell für (binäre) Zeitreihendaten $\{Y_t, t \in \mathbb{N}_0\}$:

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

- Modell für Longitudinaldaten $\{Y_{nt}, t \in \mathbb{N}_0\}$, $n = 1, \dots, N$ mit Kovariablen $\{z_{nt}, t \in \mathbb{N}_0\}$:

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

Die Kovariablen können dabei zeitabhängig oder auch zeitunabhängig sein.

- Es reicht die Übergangswahrscheinlichkeiten 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 Übergangswahrscheinlichkeiten:

$$p_{01,n}(t) = h(w'_{nt}\beta_0), \quad p_{11,n}(t) = h(z'_{nt}\beta_1).$$

h ist die sogenannte Responsefunktion $h : \mathbb{R} \rightarrow [0, 1]$. Für

$$h(x) = \exp(x)/(1 + \exp(x))$$

bzw.

$$h(x) = \phi(x)$$

erhält man das Logit- bzw. das Probitmodell. Die Kovariablenvektoren w_{nt}, z_{nt} 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, z.B. ein linearer Trend

$$w'_t\beta_0 = \beta_{00} + \beta_{01}t$$

oder ein (stückweise) polynomialer Trend. Für Longitudinaldaten enthalten w_{nt}, z_{nt} in der Regel auch individuelle Kovariablen.

- Lässt man statt einer parametrischen Trendfunktion $w_t'\beta_0$ eine nichtparametrisch spezifizierte Funktion $f_0(t)$ zu, so führt dies zu non- und semiparametrischen binären Regressionsmodellen (Generalisierte Regression, Computerintensive Verfahren).
- Konditionale Modellierung: Simultanes, autoregressives Modell in dem Y_t als Regressor fungiert:

$$P(Y_{t+1} = 1|Y_t, w_t).$$

- Beispiele:

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

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

- Das Modell mit Interaktion ist äquivalent zur separaten Modellierung (Einsetzen von $Y_t = 0$ bzw. 1 liefert die Zusammenhänge).

3.4.2 Weitere inhomogene Markov-Ketten

- Eine inhomogene Markov-Kette 2. Ordnung besitzt die Übergangswahrscheinlichkeiten

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

- Im binären Fall $Y_t \in \{0, 1\}$ genügt es wieder $p_{i_1 i_2, 1}(t)$ zu spezifizieren.
- Separate Modellierung: 4 separate binäre Regressionsmodelle für die Paare $(i_1, i_2) = (0, 0), (0, 1), (1, 0)$ und $(1, 1)$.

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

- Simultane (autoregressive) Logit-Modelle:

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

oder

$$\log \left(\frac{P(Y_{t+1} = 1 | Y_t, Y_{t-1}, z_t)}{P(Y_{t+1} = 0 | Y_t, Y_{t-1}, z_t)} \right) = z_t' \beta + Y_t z_t' \alpha_1 + Y_{t-1} z_t' \alpha_2 + Y_t Y_{t-1} z_t' \alpha_3$$

- Einsetzen der verschiedenen 0/1-Kombinationen für Y_t, Y_{t-1} liefert

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

Durch Nullsetzen von Komponenten in $\alpha_1, \alpha_2, \alpha_3$ erhält man parametersparsamere Submodelle.

- Analoges Vorgehen bei Markov-Ketten höherer Ordnung möglich.
- Allgemeinere inhomogene Markov-Ketten (mit mehr als zwei Zuständen) lassen sich durch kategoriale autoregressive Regressionsmodelle behandeln.

3.5 Statistische Inferenz

3.5.1 Likelihood-Inferenz für stochastische Prozesse

- Beobachtung eines Pfades: Die Daten $(x_0), x_1, \dots, x_T$ werden aufgefasst als Realisierungen

$$x_0 = X_0(\omega), x_1 = X_1(\omega), \dots, x_t = X_t(\omega), \dots, x_T = X_T(\omega)$$

eines stochastischen Prozess $\{X_t, t \in \mathbb{N}_0\}$.

- Bei Markov-Ketten handelt es sich also um eine diskrete oder kategoriale Zeitreihe, die als bis T vorliegende Realisierung der Markov-Kette 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 an $x_0, \dots, x_T$ und als Funktion von θ betrachtet.

- Faktorisierung der Likelihood in das Produkt bedingter Dichten:

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

- Bei Markov-Ketten ergibt sich

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

Bei homogenen Markov-Ketten kann der Index t bei f_t entfallen.

- Maximum-Likelihood-Prinzip: Wähle $\hat{\theta}_{ML}$ als Maximierer von $L(\theta|X)$.

Maximierung üblicherweise durchs 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 $l_t(x_t|.) = \ln f_t(x_t|.)$.

- Falls die X_t unabhängig sind, ergibt sich wie üblich

$$L(\theta|X) = f(x_0, \dots, x_T|\theta) = \prod_{t=0}^T f_t(x_t|\theta).$$

- Beobachtung mehrerer unabhängiger Pfade eines stochastischen Prozesses: Die Daten $x_{n0}, \dots, x_{nT}$, $n = 1, \dots, N$ werden aufgefasst als unabhängige Realisierungen

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

des Prozesses $\{X_t, t \in \mathbb{N}_0\}$. Die gemeinsame Dichte und die Likelihood ergeben sich als Produkt

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

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

- Longitudinaldaten mit Kovariablen: Beobachtungen zu einer Zielvariablen Y und einem Vektor z von Kovariablen.

Die Daten werden aufgefasst als unabhängige Realisierungen von stochastischen Prozessen Y_n , $n = 1, \dots, N$, gegeben die Kovariablen.

Die Likelihood für Beobachtung n ist gegeben durch

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

Wegen der Unabhängigkeit der verschiedenen Pfade 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 Markov-Kette mit endlichem Zustandsraum $S = \{1, \dots, m\}$.
- Unbekannte Parameter θ :

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

- Für einen beobachteten Pfad

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

ergibt sich die 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 $p(0)$.

$$\Rightarrow \hat{p}_{i_0}(0) = 1, \quad \hat{p}_i(0) = 0, i \neq i_0.$$

Betrachte deshalb nur die bedingte Likelihood

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

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

- Log-Likelihood:

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

- Maximieren von $\log L(P)$ unter der Nebenbedingung, dass die Zeilensummen =1 sind ergibt

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

- Bei ergodischen Markov-Ketten gilt für $T \rightarrow \infty$

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

- Im homogenen Fall lassen sich die Übergangswahrscheinlichkeiten bei mehreren beobachteten Pfaden wie bei einem Pfad schätzen.
- Prinzipiell lassen sich bei mehreren beobachteten Pfaden auch Übergangswahrscheinlichkeiten $p_{ij}(t)$ im inhomogenen Fall schätzen:

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

mit

$n_{ij}(t)$ Anzahl der Übergänge von i nach j mit $X_t = i$ und $X_{t+1} = j$

$n_i(t)$ Anzahl der Beobachtungen mit $X_t = i$

- Likelihood-Quotienten-Tests für homogene Markov-Ketten:

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

$$\begin{array}{ll} H_0 : \{X_t\} \text{ ist i.i.d.,} & H_1 : \{X_t\} \text{ ist Markov-Kette 1. Ordnung} \\ H_0 : \{X_t\} \text{ ist Markov-Kette 1. Ordnung,} & H_1 : \{X_t\} \text{ ist Markov-Kette 2. Ordnung} \end{array}$$

kann der Likelihood-Quotienten-Test verwendet werden:

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

Unter H_0 gilt:

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

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

3.6 Hidden Markov Modelle

- Daten prinzipiell wie bei Markov-Ketten: $X_1, \dots, X_T$ Realisationen eines stochastischen Prozesses $X = \{X_t, t \in \mathbb{N}\}$ mit diskretem Zustandsraum S .
- Die Annahme einer Markov-Kette wird ersetzt durch einen zweistufigen Mechanismus:
 - Den Daten unterliegt eine unbeobachtete (latente, hidden) Markov-Kette $C = \{C_t, t \in \mathbb{N}\}$ mit Zustandsraum $\{1, \dots, m\}$.
 - Gegeben $C_t = i$ wird X_t generiert aus einer diskreten Verteilung mit

$$\pi_{xi} = P(X_t = x | C_t = i).$$

Diese Verteilung kann parametrisch spezifiziert sein, z.B.

$$X_t | C_t = i \sim Po(\lambda_i)$$

oder (für endliches S) nonparametrisch modelliert werden.

- Zusätzliche Annahme: $X_t | C_t$ sind bedingt (gegeben C_t) unabhängig.

- Vorteil im Vergleich zu Markov-Ketten: Flexiblere Modellklasse, z.B. impliziert eine Markov-Kette eine (geometrische) Verteilung für die Rückkehrzeiten T_{ii} . In Anwendungen passt diese Verteilung häufig nicht zu den beobachteten Wartezeiten.
- $\{X_t\}$ hat im Allgemeinen marginal nicht die Markov-Eigenschaft!
- Unbekannte Modellparameter:
 - Startverteilung $\delta = p(0)$ der latenten Markov-Kette.
 - Übergangsmatrix $\Gamma = (\gamma_{ij}) = P$ der latenten Markov-Kette.
 - $\Pi = (\pi_{si})$ Matrix der Wahrscheinlichkeitsverteilungen $P(X_t = s | C_t = i)$ für nonparametrische Hidden Markov Modelle bzw. $\theta = (\theta_i)$ Vektor der Modellparameter der parametrischen Verteilungen.
- Die beobachteten Größen X_t sind unabhängig für
 - $\pi_{si} \equiv \pi_s$,
 - $m = 1$,
 - $\gamma_{ij} \equiv \gamma_j$.

- Likelihood in Hidden Markov-Modellen:

$$\begin{aligned}
 L_T &= P(X_0 = x_0, \dots, X_T = x_T) \\
 &= \sum_{i_0=1}^m \dots \sum_{i_T=1}^m (\pi_{x_0 i_0} \cdot \dots \cdot \pi_{x_T i_T}) (\delta_{i_0} \gamma_{i_0 i_1} \cdot \dots \cdot \gamma_{i_{T-1} i_T}).
 \end{aligned}$$

- Umschreiben ergibt einen numerisch günstigeren Ausdruck:

$$L_T = \delta \Lambda(x_0) \left(\prod_{t=1}^T B_t \right) \mathbf{1}',$$

mit

$$\begin{aligned}
 B_t &= \Gamma \Lambda(x_t), \\
 \Lambda(x) &= \text{diag}(\pi_{x_1}, \dots, \pi_{x_m}).
 \end{aligned}$$

- Wenige Parameter (insbesondere parametrische Modelle) $\Rightarrow$ direkte Maximierung über numerische Verfahren.
- Für nonparametrische Modelle mit vielen Parametern: Baum-Welsh-Algorithmus (EM-Algorithmus).
- Vorhersagewahrscheinlichkeiten:

$$P(X_{T+1} = x_{t+1} | X_T = x_T, \dots, X_0 = x_0) = \frac{L_{T+1}}{L_T}.$$

- Schätzung des (wahrscheinlichsten) Pfads der latenten Markov-Kette:

$$\max_{i_0, \dots, i_T} P(C_0 = i_0, \dots, C_T = i_T | X_0 = x_0, \dots, X_T = x_T)$$

bzw. äquivalent dazu

$$\max_{i_0, \dots, i_T} P(C_0 = i_0, \dots, C_T = i_T, X_0 = x_0, \dots, X_T = x_T).$$

Lösung: Viterbi-Algorithmus.

3.7 Allgemeine Markov-Ketten

- Definition: Markov-Kette (mit allgemeinem Zustandsraum S)

Eine Markov-Kette ist ein stochastischer Prozess $\{X_t, t \in \mathbb{N}_0\}$ mit Zustandsraum $(S, \mathcal{S})$, der die Markov-Eigenschaft

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

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

Die MK heißt homogen, falls die Übergangswahrscheinlichkeiten nicht von t abhängen.

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

heißt dann Übergangskern der homogenen MK.

- Bemerkungen:

(a) Für Markov-Ketten mit diskretem Zustandsraum ist $P(x, A)$ durch die Übergangsmatrix

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

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.

- Der t -schrittigen Übergangsmatrix $(p_{ij}^{(t)}) = P^t$ entspricht die t -te Iteration des Übergangskerns (Wahrscheinlichkeit, dass die Kette nach t Schritten einen Zustand in A erreicht, ausgehend vom Startwert x):

$$\begin{aligned} P^t(x, A) &= P(X_t \in A | X_0 = x) \\ &= \int P^{t-1}(y, A) P(x, dy) \end{aligned}$$

- Definition: Invariante Verteilung

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

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

- Definition: Irreduzible Markov-Ketten

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:

$\pi^*(A) > 0 \Rightarrow$ für jedes $x_0 \in S$ gibt es ein $t \geq 1$, so dass $P^t(x_0, A) > 0$.

- Definition: (A-)periodische Markov-Ketten

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

- Grenzwertsatz

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

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

wobei

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

3.8 MCMC und Metropolis-Hastings-Algorithmus

- Bayesianische Inferenz:

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

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

f die gemeinsame Dichte der Beobachtungen $D = (d_1, \dots, d_n)$ und $p(x)$ die Prior-Verteilung der Parameter x .

Dann ist nach dem Satz von Bayes

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

die Dichte der Posteriori-Verteilung der unbekannt Parameter.

Im Allgemeinen ist $\pi(x)$ wegen der Integration im Nenner nicht analytisch zugänglich. Dagegen ist der Zähler bekannt und berechenbar.

- Könnten Zufallszahlen $x^{(t)}$, $t = 1, 2, \dots, N$ aus der Posteriori-Verteilung gezogen werden, so könnten Momente der Posteriori-Verteilung und andere interessierende Größen durch entsprechende Analoga der empirischen Verteilung approximiert werden, z.B.

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

für geeignete Funktionen g , oder

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

Dabei kann x auch hochdimensional sein.

- Ziel von Markov Chain Monte Carlo (MCMC) Techniken ist die Konstruktion einer Markov-Kette, deren Übergangskern P gegen die gesuchte Posteriori-Verteilung π^* konvergiert (d.h. π^* soll die invariante Verteilung der Markov-Kette sein).

- Bisher war die Übergangsmatrix 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 Markov-Kette mit Übergangskern P und π^* als invarianter Verteilung. Dabei ist P nicht eindeutig!
- 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 $Y = y$ 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) und betrachte den Übergangskern

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

mit

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

Eine Markovkette mit dem so definiertem Übergangskern besitzt π^* als invariante Verteilung. Somit konvergieren die Iterationen des Kerns gegen π^* , falls die Kette zusätzlich irreduzibel und aperiodisch ist.

- Metropolis-Hastings-Algorithmus:

- i) Wähle Startwerte x_0 und die Länge N der Kette. Setze $t = 1$
- ii) Ziehe eine Zufallszahl 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, andernfalls setze $t = t + 1$ und fahre fort mit ii).

- Bemerkungen:

- Nach einer gewissen Konvergenzphase (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 die Burn In Phase unberücksichtigt bleibt.
- Falls $q(x, y)$ symmetrisch ist, d.h. $q(x, y) = q(y, x)$ vereinfacht sich die Akzeptanzwahrscheinlichkeit zu

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

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 Teilvektoren (oder Skalaren) $x_1, \dots, x_P$, so kann man den MH-Algorithmus auch komponentenweise anwenden.

- Damit Iterationen des Übergangskerns tatsächlich gegen π^* konvergieren, ist neben der Invarianzeigenschaft zusätzlich Irreduzibilität und Aperiodizität erforderlich. In der Praxis sind diese Eigenschaften fast immer erfüllt, können aber theoretisch schwer nachgewiesen werden. Die Konvergenz wird daher in der Regel durch Analyse des MCMC Outputs überprüft, z.B. durch

- grafische Darstellung der Samplingpfade,
- Berechnung der Autokorrelationsfunktion der gezogenen Zufallszahlen.