

# 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

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

Ein Beispiel ist das *Irrfahrtmodell* mit absorbierenden Schranken.

Absorbierende MK spielen auch in der Demographie eine Rolle, etwa im folgenden

*Heiratsmodell:*

Mit  $x = 1, 2, \dots, w$  werde das Alter eines Ledigen, gemessen in Geburtstagen, bezeichnet. Dabei sei  $w$  eine Altersobergrenze, über die hinaus wir die Person nicht mehr verfolgen. Wir betrachten folgende Zustände der Person:

$$\left. \begin{array}{l} \text{I Heirat (vor Tod)} \\ \text{II Tod als Junggeselle} \\ \text{s im Alter } w + 1 \text{ noch ledig} \end{array} \right\} \text{absorbierende Zustände}$$

$x = 0, 1, 2, \dots, w$  im Alter  $x$  ledig.

Für einen  $x$ -jährigen Ledigen bestehen im Altersintervall  $(x, x + 1]$  drei Möglichkeiten. Er kann  
 mit Wahrscheinlichkeit  $q_{x,I}$  in  $(x, x + 1]$  heiraten  
 mit Wahrscheinlichkeit  $q_{x,II}$  in  $(x, x + 1]$  als Junggeselle sterben  
 mit Wahrscheinlichkeit  $p_x$  als Lediger das Alter  $x + 1$  erreichen.

Mit diesen Übergangswahrscheinlichkeiten lässt sich der Heiratsprozess als MK modellieren mit der ÜM

$$\begin{array}{c} I \quad II \quad s \quad 0 \quad 1 \quad \dots \quad \dots \quad w \\ \begin{pmatrix} I & 1 & 0 & 0 & & & & 0 \\ II & 0 & 1 & 0 & & 0 & & \\ s & 0 & 0 & 1 & & & & \\ 0 & q_{0,I} & q_{0,II} & 0 & p_0 & 0 & \dots & 0 \\ 1 & \vdots & \vdots & 0 & \vdots & p_1 & & \vdots \\ \vdots & \vdots & \vdots & \vdots & & & & p_{w-1} \\ w & q_{w,I} & q_{w,II} & p_w & 0 & & & 0 \end{pmatrix} \end{array}$$

- (c) *Marshall/Goldhammer* benutzten ein einfaches Markovmodell zur Darstellung des Prozesses, der zur ersten Einweisung in ein Nervenkrankenhaus führt. Sie nahmen – bei einer Zeiteinheit von einem Jahr – einen Prozess mit den Zuständen  $S_4$  gesund,  $S_3$  leichte Erkrankung, nicht im Krankenhaus;  $S_2$  schwere Erkrankung, nicht im Krankenhaus,  $S_1$  krank, im Krankenhaus,  $S_0$

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  $P_i^\infty$  berechnen sich aus (3.3), (3.4),  $L^\infty$  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  $\mu_{jj}$ . 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  $\mu_{jj}$  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  $\pi_i, i \in S$ , 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  $\pi$  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  $\pi$  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  $\beta_{11}$ . 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  $\theta$  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  $\theta$  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  $\theta$ :

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

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.

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

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

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

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 ( $\stackrel{a}{\sim} \chi^2(1)$ ):

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

$\Rightarrow$  **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 unfaiem Würfel mit Wahrscheinlichkeit 0.05, und von unfaiem 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,  $\epsilon_t$  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  $\pi^*$  auf  $(S, \mathcal{S})$  mit Dichte  $\pi$  (bezüglich des dominierenden Maßes  $\mu$ ) 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  $\pi^*$  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  $\pi^*$ . Dann ist  $\pi^*$  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  $\pi$  bzw.  $\pi^*$  bestimmt werden. Jetzt betrachtet man das umgekehrte Problem: Finde zu gegebener Verteilung  $\pi$  eine MK mit Übergangskern  $P$  und  $\pi$  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  $\pi^*$  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  $\pi^*$  als invariante Verteilung besitzt, so konvergieren Iterationen des Kerns gegen  $\pi^*$ , 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  $\pi$  bzw.  $\pi^*$  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  $\pi^*$  angesehen werden. Eigenschaften von  $\pi^*$  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  $\pi^*$  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.