Kapitel 3
Markov-Ketten

3.1 Grundlegende Eigenschaften, Beispiele

- Markov-Ketten werden in vielen statistischen Modellen von realen Prozessen verwendet und sind außerdem in MCMC-Verfahren ganz zentral (siehe 3.8).

- 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) $:\Leftrightarrow \forall t \in \mathbb{N}_0,\ j, i, i_{t-1}, \ldots, i_0 \in S$

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

- Beispiel: Einfache Irrfahrt

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

$$\begin{aligned} P(Z_t = 0 | X_t = i) &= r_i, \\ P(Z_t = -1 | X_t = i) &= q_i, \\ P(Z_t = +1 | X_t = i) &= p_i, \\ p_i + r_i + q_i &= 1 \end{aligned}$$

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

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

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 :⇔

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

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

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

$\Longrightarrow$ Computerintensive Verfahren, Generalisierte Regression.

- Definition: Übergangsmatrix

  Für homogene Markov-Ketten heißt

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

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

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

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

- P ist endlich, falls S endlich ist.

- Eigenschaften der Übergangsmatrix

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

- Folgerungen aus der Definition einer Markov-Kette:

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

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

Ist zusätzlich eine Anfangsverteilung

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

gegeben, so gilt

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

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

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

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

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

- 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.

    $\Rightarrow$ 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.

Beweis:

Beweis zu 3.: FKO S. 17.

Beweis zu 1.:

$$
\begin{aligned}
& P\{X_{t+s} = i_s, X_{t+s-1} = i_{s-1}, \ldots, X_{t+2} = i_2, X_{t+1} = i_1 \mid X_t = i_0\} \\
= \quad & P\{X_{t+s} = i_s \mid X_{t+s-1} = i_{s-1}, \ldots, X_{t+2} = i_2, X_{t+1} = i_1, X_t = i_0\} \\
& \cdot P\{X_{t+s-1} = i_{s-1}, \ldots, X_{t+2} = i_2, X_{t+1} = i_1 \mid X_t = i_0\} \\
\stackrel{\text{hom. ME}}{=} \quad & p_{i_{s-1},i_s} P\{X_{t+s-1} = i_{s-1}, \ldots, X_{t+2} = i_2, X_{t+1} = i_1 \mid X_t = i_0\} \\
\stackrel{\text{iterativ}}{=} \quad & p_{i_{s-1},i_s} p_{i_{s-2},i_{s-1}} \cdot \ldots \cdot p_{i_0,i_1}.
\end{aligned}
$$

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

$$P\{X_{t+2} = i_{t+2}, X_{t+1} = i_{t+1} \mid X_t = i_t, V_{t-1}\}$$

$$\begin{aligned}
&= P\{X_{t+2} = i_{t+2} \mid X_{t+1} = i_{t+1}, X_t = i_t, V_{t-1}\} \cdot P\{X_{t+1} = i_{t+1} \mid X_t = i_t, V_{t-1}\} \\
&\overset{\text{ME}}{=} P\{X_{t+2} = i_{t+2} \mid X_{t+1} = i_{t+1}\} \cdot P\{X_{t+1} = i_{t+1} \mid X_t = i_t\} \\
&\overset{3.}{=} P\{X_{t+2} = i_{t+2} \mid X_{t+1} = i_{t+1}, X_t = i_t\} \cdot P\{X_{t+1} = i_{t+1} \mid X_t = i_t\} \\
&= P\{X_{t+2} = i_{t+2}, X_{t+1} = i_{t+1} \mid X_t = i_t\}.
\end{aligned}$$

$\square$

- Definition: Mehrschrittige Übergangswahrscheinlichkeiten

$$p_{ij}^{(t)} := P(X_{t+s} = j | X_s = i) = P(X_t = j | X_0 = i), \quad t \geq 0$$

heißt t-schrittige Übergangswahrscheinlichkeit von i nach j.

- Chapman–Kolmogorov–Gleichungen

$$p_{ij}^{(t+s)} = \sum_{k \in S} p_{ik}^{(s)} p_{kj}^{(t)} \qquad (1)$$

$$(p_{ij}^{(t)})_{i,j \in S} = P^t = \underbrace{P \cdot \ldots \cdot P}_{t-mal} \qquad (2)$$

$$P^{t+s} = P^s P^t. \qquad (3)$$

Beweis:

Beweis zu (1):

$$
\begin{aligned}
p_{ij}^{(t+s)} &= P\{X_{t+s} = j \mid X_0 = i\} \\
&\overset{\text{Satz d. tot. Wkt.}}{=} \sum_{k \in S} P\{X_{t+s} = j, X_s = k \mid X_0 = i\} \\
&= \sum_{k \in S} P\{X_{t+s} = j \mid X_s = k, X_0 = i\} \cdot P\{X_s = k \mid X_0 = i\} \\
&\overset{\text{ME}}{=} \sum_{k \in S} P\{X_{t+s} = j \mid X_s = k\} \cdot P\{X_s = k \mid X_0 = i\} \\
&\overset{\text{Homogenität}}{=} \sum_{k \in S} p_{ik}^{(s)} p_{kj}^{(t)}
\end{aligned}
$$

Beweis zu (2):

$$P^t = P \cdot \ldots \cdot P \text{ folgt (iterativ) aus } P^{t+s} = P^s \cdot P^t$$

Beweis zu (3): (3) entspricht Aussage (1) in Matrixform.

$\square$

- 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

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

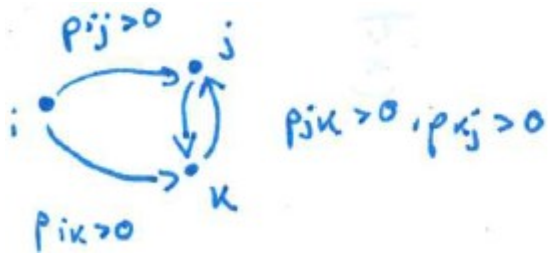
bzw.

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

Beweis über den Satz von der totalen Wahrscheinlichkeit:

$P(X_t = j) = \sum_{i \in S} P(X_t = j | X_0 = i) \cdot P(X_0 = i)$

- Darstellung von Markov-Ketten durch gerichtete Graphen.



Beispiele:

(a) Irrfahrtmodelle FKO, S. 22

(b) Markov-Ketten auf einem DNA-Strang

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

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

(angeordnet in Form einer Doppel-Helix).

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

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

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 $AGCT$

$\rightarrow$ 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

$$P = \begin{array}{r} \\ A \\ C \\ G \\ T \\ AG \\ AGC \\ AGCT \end{array} \begin{array}{c} \begin{array}{ccccccc} A & C & G & T & AG & AGC & AGCT \end{array} \\ \begin{pmatrix} .32 & .18 & 0 & .27 & .23 & 0 & 0 \\ .37 & .23 & .05 & .35 & 0 & 0 & 0 \\ .30 & .21 & .25 & .24 & 0 & 0 & 0 \\ .23 & .19 & .25 & .33 & 0 & 0 & 0 \\ .30 & .0 & .25 & .24 & 0 & .21 & 0 \\ .37 & .23 & .05 & 0 & 0 & 0 & .35 \\ .23 & .19 & .25 & .33 & 0 & 0 & 0 \end{pmatrix} \end{array} .$$

Für Sequenzanalysen eignen sich MK nur bedingt, insbesondere ist die Annahme der Homogenität verletzt. Sie sind jedoch wesentlicher Baustein in komplexeren Modellen wie "Hidden Markov Modellen" (3.6).

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

Ein Gerät werde zu den diskreten Zeitpunkten $t = 0, 1, 2, \ldots$ überprüft. Falls ein Fehler gefunden wird, wird es durch ein neues ersetzt, anderfalls 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

$$P\{Z \geq i+1 \mid Z \geq i\} = p_i\,,\ i \geq 0. \tag{4}$$

Sei nun X_t das Alter des zu Beginn der t-ten Periode arbeitenden Gerätes. Mit Wahrscheinlichkeit p_i geht dann also gemäß (4) 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 seit Beginn der ersten Arbeitslosigkeit in Monaten.
$p_{01}(t), \quad t = 1, 2, \ldots,$ Wahrscheinlichkeit für Übergang in Zustand Arbeit.
Analog: $p_{10}(t)$ ÜW von Arbeit nach Arbeitslosigkeit.
I.a. inhomogen (und i.a. abhängig von persönlichen Merkmalen des Arbeitslosen)
Fraglich: überhaupt MK, d.h. gilt die Markoveigenschaft?

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

Bei einigen Beispielen ist die Markov-Eigenschaft zweifelhaft.

- Markov-Ketten höherer Ordnung

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

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

- Modellierung z.B. durch konditionale Logit-Modelle

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