7. Markov-Ketten

Benannt nach Andrei A. Markov [1856-1922]
Einige Stichworte:

- Markov-Ketten
  - Definition
  - Eigenschaften
  - Konvergenz
- Hidden Markov Modelle
  - Motivation und Inferenz
  - Baum-Welch-Algorithmus
  - Viterbi-Algorithmus

7.1 Definition und Eigenschaften von Markov-Ketten

Sei $\mathbf{X} = (X_0, X_1, X_2, \ldots)$ eine Folge von diskreten Zufallsvariablen, die alle Ausprägungen in einer endlichen bzw. abzählbaren Menge S haben.
S heißt der **Zustandsraum** und $s \in S$ ein **Zustand**.
$\mathbf{X}$ heißt **Markov-Kette** (MK), falls

$$\begin{aligned} & P(X_n = s | X_0 = x_0, X_1 = x_1, \ldots, X_{n-1} = x_{n-1}) \\ = \; & P(X_n = s | X_{n-1} = x_{n-1}) \end{aligned}$$

für alle $n \geq 1$ und alle $s, x_0, x_1, \ldots, x_{n-1} \in S$.

Interpretation

Bedingt auf die gesamte Vergangenheit $X_0, X_1, \ldots, X_{n-1}$ des Prozesses $\mathbf{X}$ hängt X_n nur vom letzten Wert X_{n-1} ab.
Darstellung mit Pfeilen in einem **graphischen Modell**:

$$X_0 \to X_1 \to X_2 \to \ldots \to X_{n-1} \to X_n$$

Anders ausgedrückt:

Bedingt auf die **Gegenwart** (X_{n-1}) ist die **Zukunft** des Prozesses $(X_n, X_{n+1}, \ldots)$ unabhängig von seiner **Vergangenheit** $(X_0, X_1, \ldots, X_{n-2})$.
$\rightarrow$ Begriff der **bedingten Unabhängigkeit**

Bedingt unabhängige Zufallsvariablen

Zwei diskrete Zufallsvariablen X und Y heißen **bedingt unabhängig** gegeben Z, wenn für die entsprechend definierten Wahrscheinlichkeitsfunktionen gilt:

$$f_{X,Y|Z}(x, y|z) = f_{X|Z}(x|z) \cdot f_{Y|Z}(y|z)$$

für alle x, y und z.

Übergangswahrscheinlichkeiten

Die Entwicklung einer Markov-Kette $\mathbf{X}$ ist gekennzeichnet durch die (Ein-Schritt) **Übergangswahrscheinlichkeiten**

$$P(X_{n+1} = j | X_n = i)$$

für alle $i, j \in S$.
Man nennt eine Markov-Kette **homogen**, wenn diese nicht von n abhängen und definiert

$$p_{ij} = P(X_{n+1} = j | X_n = i) = P(X_1 = j | X_0 = i)$$

für alle $n \geq 1$ und alle $i, j \in S$.

Die Übergangsmatrix

Die Werte p_{ij} werden in einer $|S| \times |S|$-Matrix **P** zusammengefasst, der sogenannten **Übergangsmatrix**. **P** beschreibt also die Kurzzeitentwicklung einer homogenen Markov-Kette **X**.

P ist eine **stochastische Matrix**, d.h. sie hat folgende Eigenschaften:

1. $p_{ij} \geq 0 \quad$ für alle $i, j \in S$
2. $\sum_j p_{ij} = 1 \quad$ für alle $i \in S \quad \Rightarrow$ "Zeilensummen gleich eins"

Beispiele

1.Beispiel: Telephon besetzt / frei mit $S = \{0, 1\}$ und

$$\mathbf{P} = \begin{pmatrix} 0.9 & 0.1 \\ 0.4 & 0.6 \end{pmatrix}$$

2.Beispiel: Der Zustandsraum umfasst die vier Basen der DNA (Adenin, Cytosin, Guanin, Thymin): $S = \{A, C, G, T\}$.
Die geschätzte Übergangsmatrix beim "Ablaufen" der DNA ist

$$\mathbf{P} = \begin{pmatrix} 0.300 & 0.205 & 0.285 & 0.210 \\ 0.322 & 0.298 & 0.078 & 0.302 \\ 0.248 & 0.246 & 0.298 & 0.208 \\ 0.177 & 0.239 & 0.292 & 0.292 \end{pmatrix}$$

Durbin *et al.* (1998) "Biological sequence analysis", p. 50

Gerichtete Graphen

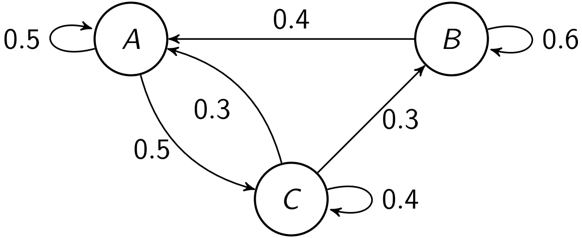
Der von der Markov-Kette $\mathbf{X}$ mit Zustandsraum S und Übergansmatrix $\mathbf{P}$ erzeugte gerichtete Graph G besteht aus $|S|$ Knoten und maximal $|S|^2$ Kanten. Jeder Knoten bezieht sich auf einen Zustand und jede Kante (i, j) auf p_{ij} für $p_{ij} > 0$. Der Graph beschreibt also die Struktur aller direkten Verbindungen von Zustand i zu Zustand j $\forall_{i,j}$.

Darstellung von MKs mit gerichteten Graphen

Ein Zusandsraum $S = \{A, B, C\}$ und die dazugehörige Übergangsmatrix

$$\mathbf{P} = \begin{pmatrix} 0.5 & 0 & 0.5 \\ 0.4 & 0.6 & 0 \\ 0.3 & 0.3 & 0.4 \end{pmatrix}$$

erzeugen den folgenden gerichteten Graphen:

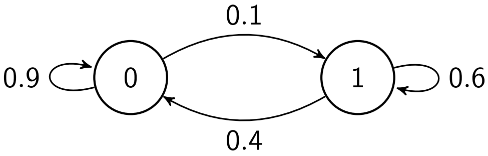


Darstellung von MKs mit gerichteten Graphen II

1. Beispiel: Aus Telephon besetzt / frei mit $S = \{0, 1\}$ und

$$\mathbf{P} = \begin{pmatrix} 0.9 & 0.1 \\ 0.4 & 0.6 \end{pmatrix}$$

folgt der folgende gerichtete Graph:



Die n-Schritt-Übergangsmatrix

Die Langzeitentwicklung einer Markov-Kette $\mathbf{X}$ ist durch die n-**Schritt-Übergangsmatrix** $\mathbf{P}(m, m+n)$ mit Elementen

$$\begin{aligned} p_{ij}(m, m+n) &= P(X_{m+n} = j | X_m = i) \\ \text{bzw.} \quad p_{ij}(n) &= P(X_n = j | X_0 = i) \end{aligned}$$

gegeben, wobei die letzte Gleichung für homogene Markov-Ketten gilt. In diesem Fall ergibt sich natürlich

$$\mathbf{P}(m, m+1) = \mathbf{P}$$

und wir schreiben $\mathbf{P}_n = \mathbf{P}(m, m+n)$.

Die Chapman-Kolmogorov-Gleichungen

Für eine homogene Markov-Kette $\mathbf{X}$ gilt:

$$p_{ij}(m, m+n+r) = \sum_k p_{ik}(m, m+n) p_{kj}(m+n, m+n+r) \quad (1)$$

$$\mathbf{P}(m, m+n+r) = \mathbf{P}(m, m+n)\mathbf{P}(m+n, m+n+r) \quad (2)$$

$$\mathbf{P}_n = \mathbf{P}(m, m+n) = \mathbf{P}^n \quad (3)$$

Dabei ist (2) nur (1) in Matrizenform und (3) folgt durch Iteration.
$\mathbf{P}^n$ bezeichnet die n-te Potenz von $\mathbf{P}$.
Beweis zu (1) in Vorlesung.

Fortsetzung des 1. Beispiels

Zwei-Schritt-Übergangsmatrix:

$$\mathbf{P}_2 = \mathbf{P}^2 = \begin{pmatrix} 0.9 & 0.1 \\ 0.4 & 0.6 \end{pmatrix} \begin{pmatrix} 0.9 & 0.1 \\ 0.4 & 0.6 \end{pmatrix} = \begin{pmatrix} 0.85 & 0.15 \\ 0.6 & 0.4 \end{pmatrix}$$

Zum Beispiel ist also der Eintrag $p_{21}(2)$ in $\mathbf{P}_2$ gleich

$$\begin{aligned} p_{21}(2) &= P(X_{n+2} = 1 | X_n = 2) \\ &= P(X_{n+1} = 2 | X_n = 2) \cdot P(X_{n+2} = 1 | X_{n+1} = 2) \\ &\quad + P(X_{n+1} = 1 | X_n = 2) \cdot P(X_{n+2} = 1 | X_{n+1} = 1) \\ &= 0.6 \cdot 0.4 + 0.4 \cdot 0.9 = 0.6 \end{aligned}$$

Die Zeit bis zum Zustandswechsel

Eine Markov-Kette $\mathbf{X}$ mit Übergangsmatrix $\mathbf{P}$ sei zu einem Zeitpunkt t im Zustand $i \in S$.
Dann ist die Dauer bis zum nächsten Zustandswechsel Z_i geometrisch verteilt mit Parameter $1 - p_{ii}$.
Vergleiche: Gedächtnislosigkeit der geometrischen Verteilung:

$$P(Z_i = n + k | Z_i > n) = P(Z_i = k)$$

Nützliche Eigenschaft zum Test auf Modellanpassung:
Vergleich der beobachteten und theoretischen Dauern bis zum nächsten Zustandswechsel.

Die Anfangsverteilung

Schließlich hat jede Markov-Kette auch eine **Anfangsverteilung** für X_0. Die entsprechende Wahrscheinlichkeitsfunktion bezeichnet man mit dem Zeilenvektor $\boldsymbol{\mu}^{(0)}$ mit Elementen

$$\mu_i^{(0)} = P(X_0 = i).$$

Die (unbedingte) Wahrscheinlichkeitsfunktion von X_n fasst man entsprechend in dem Zeilenvektor $\boldsymbol{\mu}^{(n)}$ zusammen und es gilt:

$$\begin{aligned} \boldsymbol{\mu}^{(m+n)} &= \boldsymbol{\mu}^{(m)} \cdot \mathbf{P}^n \quad \text{und daher} \\ \boldsymbol{\mu}^{(n)} &= \boldsymbol{\mu}^{(0)} \cdot \mathbf{P}^n \end{aligned}$$

Folgerung

Durch $\boldsymbol{\mu}^{(0)}$ und $\mathbf{P}$ ist also die Entwicklung einer Markov-Kette vollständig bestimmt.
Die **gemeinsame** Wahrscheinlichkeitsverteilung (wichtig für Likelihood-Inferenz!) von $X_0, \ldots, X_n$ ist gegeben durch

$$\begin{aligned} & P(X_0 = x_0, X_1 = x_1, \ldots, X_n = x_n) \\ = \quad & P(X_0 = x_0) \prod_{t=1}^{n} P(X_t = x_t | X_{t-1} = x_{t-1}) \\ = \quad & \mu_{x_0}^{(0)} \prod_{t=1}^{n} p_{x_{t-1}, x_t} \end{aligned}$$

Beispiel: Inzucht

Eine Pflanze mit Genotyp aus $S = \{aa, ab, bb\}$ wird mit sich selbst gekreuzt. Die Zufallsvariable X_n gibt den Genotyp in der n-ten Generation an. Daher:

$$\mathbf{P} = \begin{pmatrix} 1 & 0 & 0 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ 0 & 0 & 1 \end{pmatrix}$$

$$\mathbf{P}^n = \begin{pmatrix} 1 & 0 & 0 \\ \frac{1}{2} - \left(\frac{1}{2}\right)^{n+1} & \left(\frac{1}{2}\right)^n & \frac{1}{2} - \left(\frac{1}{2}\right)^{n+1} \\ 0 & 0 & 1 \end{pmatrix} \xrightarrow{n\to\infty} \begin{pmatrix} 1 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 1 \end{pmatrix}$$

$\Rightarrow$ Letztendlich bleiben nur die Genotypen *aa* und *bb* übrig, die sich dann deterministisch reproduzieren.

Beispiel: Genhäufigkeit in Population konstanter Größe N

$X_n = i$: Anzahl der Individuen einer Population mit bestimmtem Genotyp zum Zeitpunkt n

Einfaches Modell:

Zu jedem Zeitpunkt stirbt ein zufällig ausgewähltes Mitglied. Das "nachrückende" Mitglied hat den Genotyp mit Wahrscheinlichkeit $\frac{i}{N}$.

$$\rightsquigarrow p_{ij} = \begin{cases} \frac{i(N-i)}{N^2} & \text{für } j = i \pm 1 \\ 1 - 2\frac{i(N-i)}{N^2} & \text{für } j = i \\ 0 & \text{sonst} \end{cases}$$

Beachte: $\mathbf{P}$ ist "tridiagonal". Formulierung beinhaltet Grenzfälle $X_n = 0$ und $X_n = N$.

Beispiel: Modelle für Epidemien $\rightarrow$ Verzweigungsprozesse

$X_n :=$ Anzahl der Infizierten in einer Generation n

$S \quad = \{0, 1, \ldots\}$

Idee: Jeder Infizierte "erzeugt" (unabhängig von den anderen) eine zufällige Anzahl Infizierter in der nächsten Generation mit Erwartungswert λ.
Die Anzahl könnte z.B. Poissonverteilt sein. Dann ist $X_n | X_{n-1} \sim \mathcal{P}(\lambda \cdot X_{n-1})$.

Theorem:

- Für $\lambda < 1$ wird die Epidemie mit Wahrscheinlichkeit 1 irgendwann aussterben.
- Für $\lambda > 1$ ist die Wahrscheinlichkeit, dass die Epidemie explodiert, echt größer 0.

7.2 Klassifikation von Zuständen und Markov-Ketten

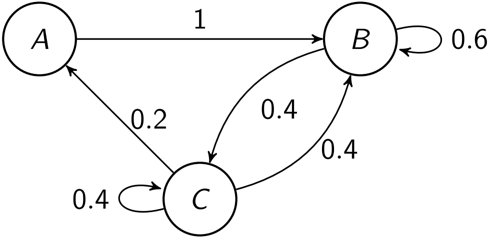
Ein Zustand $i \in S$ heißt **rekurrent** oder auch **persistent**, falls für die Rekurrenzzeit $T_i = min\{n : X_n = i | X_0 = i\}$ gilt

$$P(T_i < \infty) = 1$$

Wenn $P(T_i < \infty) < 1$ heißt der Zustand **transient**. Eine Markov-Kette $\mathbf{X}$ kehrt also in einen rekurrenten Zustand mit Wahrscheinlichkeit 1 zurück. Ein Zustand i heißt **absorbierend**, falls dieser nicht mehr verlassen werden kann, also $p_{ii} = 1$.

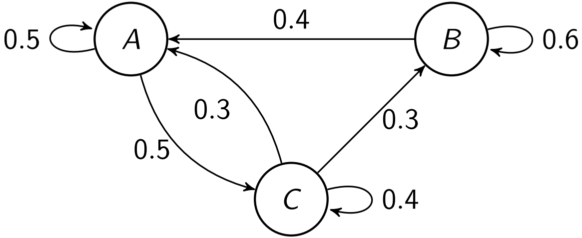
Graphische Beispiele für Zustandsklassen

1. Die Rekuzzenzzeit von Zustand A ist $T_A = 3$:



Die minimale Anzahl an Schritten bis **X** von A nach A zurückkehrt ist 3.

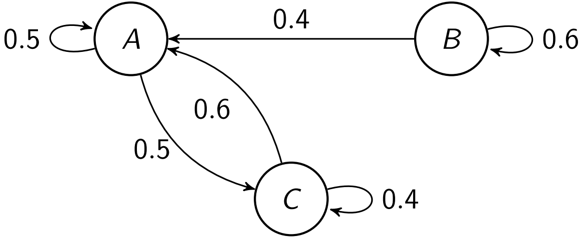
2. Alle Zustände sind rekurrent/persistent:



Die Markov-Kette kehrt mit Wahrscheinlichkeit 1 in alle Zustände zurück.

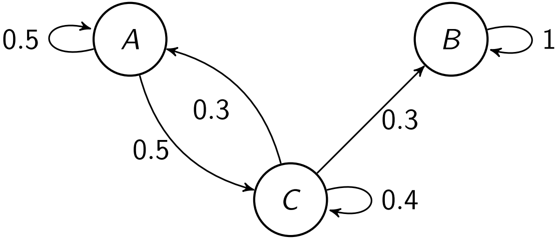
Graphische Beispiele für Zustandsklassen II

3. Zustand B ist transient:



Die Rückkehr nach B aus anderen Zuständen ist nicht möglich.

4. Zustand B ist absorbierend:



Die Markov-Kette verlässt den Zustand B nicht mehr.

Beispiel

Betrachte den Poisson-Verzweigungsprozess

$$X_n | X_{n-1} \sim \mathcal{P}(\lambda \cdot X_{n-1})$$

Dann ist der Zustand 0 **rekurrent**, ja sogar **absorbierend**, da $\mathbf{X}$ diesen Zustand nie verläßt.
Alle anderen Zustände sind **transient**.

Die erwartete Rekurrenzzeit

Man definiert die **erwartete Rekurrenzzeit** eines Zustands i wie folgt:

$$\mu_i = E(T_i) = \begin{cases} \sum_n n f_i(n) & \text{falls } i \text{ rekurrent ist} \\ \infty & \text{falls } i \text{ transient ist} \end{cases}$$

mit $f_i(n) = P(X_1 \neq i, ..., X_{n-1} \neq i, X_n = i | X_0 = i)$. Ein rekurrenter Zustand i heißt **nicht-leer**, falls seine erwartete Rekurrenzzeit endlich ist. Ansonsten heißt er **leer**.

Zusammenhang zu den Übergangswahrscheinlichkeiten

Ein rekurrenter Zustand i ist genau dann leer, wenn

$$p_{ii}(n) \to 0 \text{ für } n \to \infty$$

Dann gilt sogar

$$p_{ji}(n) \to 0 \text{ für } n \to \infty \text{ für alle } j \in S$$

Die Periode

Die **Periode** eines Zustandes i ist der größte gemeinsame Teiler der Menge

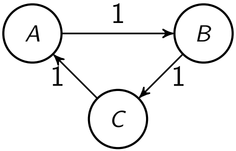
$$\{n : p_{ii}(n) > 0\}$$

Man nennt den Zustand i **periodisch**, falls dessen Periode größer eins ist, ansonsten heißt i **aperiodisch**.
Haben alle Zustände einer Markov-Kette Periode 1, so heißt sie aperiodisch.
Ein Zustand i heißt **ergodisch**, falls er rekurrent, nicht-leer und aperiodisch ist. Sind alle Zustände von $\mathbf{X}$ ergodisch, so heißt $\mathbf{X}$ ergodisch.

Beispiel

Eine Markov-Kette habe die Zustände $S = \{A, B, C\}$ und folgende Übergangsmatrix:

$$\mathbf{P} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \quad \Leftrightarrow$$



Es gilt nun:

- jeder Zustand hat Periode 3
- die Markov-Kette ist nicht aperiodisch

Irreduzible Zustände

Zustand j heißt **erreichbar** durch Zustand i, falls

$$p_{ij}(n) > 0$$

für mindestens ein $n \in \mathbb{N}$. (Schreibweise: $i \rightarrow j$)
Zwei Zustände $i \neq j$ einer Markov-Kette **X kommunizieren** miteinander, falls die Zustände **gegenseitig erreichbar** sind, also falls

$$p_{ij}(n) > 0 \text{ und } p_{ji}(n') > 0$$

für mindestens ein $n \in \mathbb{N}_0$ bzw. $n' \in \mathbb{N}_0$. (Schreibweise: $i \leftrightarrow j$)

Ein Zustand i kommuniziert per Definition immer mit sich selbst: $i \leftrightarrow i$
Die angenommenen Zustände zwischen i und j werden **Pfad** genannt.

Irreduzible Zustände II

Es gilt:

- Wenn $i \leftrightarrow k$ und $k \leftrightarrow j$, dann auch $i \leftrightarrow j$
- Ist i rekurrent und $i \leftrightarrow j$, so ist auch j rekurrent
- Ist i transient und $i \leftrightarrow j$, so ist auch j transient
- Erreichbarkeit entspricht einer Äquivalenzrelation
  $\Rightarrow$ Es können Äquivalenzklassen gebildet werden

Eine Menge $C \subset S$ heißt **irreduzibel**, falls $i \leftrightarrow j$ für alle $i, j \in C$. Eine

Menge $C \subset S$ heißt **geschlossen**, falls $p_{ij} = 0$ für alle $i \in C$ und $j \in \bar{C}$.

Eine MK heißt **stark verbunden** falls $i \leftrightarrow j$, also wenn für mindestens ein $n \in \mathbb{N}$ gilt: $p_{ij}(n) > 0 \ \forall_{i,j}$.

Eine MK ist also irreduzibel, g.d.w. sie stark verbunden ist.

Beispiele

Markov-Ketten mit reduziblem Zustandsraum S

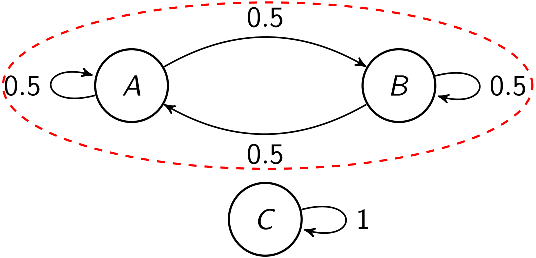
$$\mathbf{P_1} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \mathbf{P_2} = \begin{pmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ 0 & 0 & 1 \end{pmatrix}$$

Im folgenden Beispiel ist S irreduzibel:

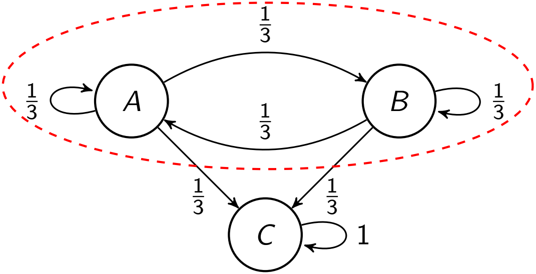
$$\mathbf{P_3} = \begin{pmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

Beispiele für reduzible Zustandsräume in graphischer Form

$\mathbf{P_1}$:



$\mathbf{P_2}$:

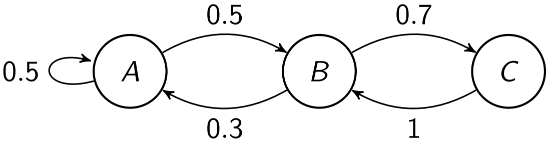


Zustandsraum $S = \{A, B, C\}$ kann auf $C = \{A, B\}$ (rot markiert) reduziert werden, C ist dann irreduzibel.

Essentielle/finale Zustände

- Zustand i heißt **essentiell/final**, falls $\forall_j \; j \leftrightarrow i$, also wenn er mit allen andernen Zuständen kommuniziert.

Graphisches Beispiel:



Alle Zunstände sind essentiell, da sie miteinander kommunizieren.
Eine irreduzible bzw. stark verbundene **MK** besteht also aus essentiellen/finalen Zuständen.

Der Zerlegungssatz

Der Zustandsraum S einer Markov-Kette $\mathbf{X}$ lässt sich zerlegen in

1. eine Menge T mit transienten Zuständen
2. Mengen C_k, die irreduzibel und geschlossen sind

$$\Rightarrow S = T \cup C_1 \cup C_2 \cup \ldots$$

Ferner gilt folgendes Lemma:

Wenn S endlich ist, dann ist mindestens ein Zustand rekurrent und alle rekurrenten Zustände sind nicht-leer.

Beispiel

Eine Markov-Kette mit Zustandsraum $S = \{1, 2, 3, 4, 5, 6\}$ habe die Übergangsmatrix

$$\mathbf{P} = \begin{pmatrix} 0.5 & 0.5 & 0 & 0 & 0 & 0 \\ 0.25 & 0.75 & 0 & 0 & 0 & 0 \\ 0.25 & 0.25 & 0.25 & 0.25 & 0 & 0 \\ 0.25 & 0 & 0.25 & 0.25 & 0 & 0.25 \\ 0 & 0 & 0 & 0 & 0.5 & 0.5 \\ 0 & 0 & 0 & 0 & 0.5 & 0.5 \end{pmatrix}$$

Man bestimme:

- die Periode jedes Zustands
- die Zerlegung des Zustandsraumes
- transiente Zustände
- ergodische Zustände

7.3 Die stationäre Verteilung und das Grenzwerttheorem

Eine Wahrscheinlichkeitsverteilung $\boldsymbol{\pi}$ (Zeilenvektor) mit Einträgen $(\pi_j : j \in S)$ heißt **stationäre Verteilung** einer Markov-Kette $\mathbf{X}$ mit Übergangsmatrix $\mathbf{P}$, falls gilt:

$$\pi_j = \sum_i \pi_i p_{ij}$$

oder in Matrixnotation:

$$\boldsymbol{\pi} = \boldsymbol{\pi} \cdot \mathbf{P}$$

Interpretation der stationären Verteilung

Hat die Markov-Kette $\mathbf{X}$ die Verteilung $\boldsymbol{\pi}$ zu einem gewissen Zeitpunkt n, dann auch im nächsten Zeitpunkt $n+1$ und sogar in allen nachfolgenden Zeitpunkten $i = n+2, n+3, \ldots$
Betrachte z.B. $i = n+2$:

$$\boldsymbol{\pi} \cdot \mathbf{P}^2 \;=\; (\boldsymbol{\pi}\mathbf{P})\mathbf{P} = \boldsymbol{\pi}\mathbf{P} = \boldsymbol{\pi}$$

Oft wählt man für die Anfangsverteilung $\boldsymbol{\mu}_0$ die stationäre Verteilung, d.h. $\boldsymbol{\mu}_0 = \boldsymbol{\pi}$.
Im Folgenden betrachten wir ausschließlich irreduzible Markov-Ketten, d.h. Markov-Ketten mit irreduziblem Zustandsraum S.

Satz über die stationäre Verteilung

Eine irreduzible Markov-Ketten hat eine **stationäre Verteilung $\boldsymbol{\pi}$** mit $\pi_i > 0 \; \forall_i$.
Dann ist $\boldsymbol{\pi}$ eindeutig und gegeben durch

$$\pi_i = 1/\mu_i$$

wobei μ_i die erwartete Rekurrenzzeit des Zustands i ist.
Unter diesen Voraussetzungen gilt bei endlichem Zustandsraum:

$$\boldsymbol{\pi} = \mathbf{1}(\mathbf{I} - \mathbf{P} + \mathbf{U})^{-1}$$

wobei **I** die Einheitsmatrix und **1** ein Zeilenvektor mit Einsen ist und **U** nur Elemente gleich eins hat.

Bestimmung der stationären Verteilung bei $|S| = 2$

Stationäre Verteilung und Übergangsmatrix haben bei Markov-Ketten mit zwei Zuständen folgende allgemeine Formen:

$$\boldsymbol{\pi} = (\pi_1, 1 - \pi_1) \qquad \mathbf{P} = \begin{pmatrix} 1 - p_{12} & p_{12} \\ p_{21} & 1 - p_{21} \end{pmatrix}$$

Die erste Spalte der Gleichung $\boldsymbol{\pi} = \boldsymbol{\pi} \cdot \mathbf{P}$ ist

$$\pi_1 = \pi_1 - \pi_1 \cdot p_{12} + p_{21} - \pi_1 \cdot p_{21}$$

$$\Rightarrow \pi_1 = \frac{p_{21}}{p_{12} + p_{21}} \text{ und } \pi_2 = \frac{p_{12}}{p_{12} + p_{21}}$$

Die zweite Spalte ergibt die gleiche Lösung.

Beispiel

Wie lautet die stationäre Verteilung für die Markov-Kette mit folgender Übergangsmatrix:

$$\mathbf{P} = \begin{pmatrix} 0.9 & 0.1 \\ 0.4 & 0.6 \end{pmatrix}$$

Mit der eben bestimmten Formel erhält man:

$$\boldsymbol{\pi} = (\pi_1, \pi_2) = \left(\frac{0.4}{0.5}, \frac{0.1}{0.5}\right) = (0.8\,,\,0.2)$$

Die erwarteten Rekurrenzzeiten sind demnach $\mu_1 = 5/4$ für Zustand 1 und $\mu_2 = 5$ für Zustand 2.
Wie verhält es sich bei

$$\mathbf{P} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}?$$

Reversible Markov-Ketten

Sei $\mathbf{X} = (X_0, \ldots, X_N)$ eine reguläre Markov-Kette mit Übergangsmatrix $\mathbf{P}$ und stationärer Verteilung $\boldsymbol{\pi}$, die $\mathbf{X}$ auch zu jedem Zeitpunkt $n = 0, \ldots, N$ besitze.
Definiere nun $\mathbf{Y} = (X_N, \ldots, X_0)$ mit $Y_n = X_{N-n}$.
Dann ist $\mathbf{Y}$ auch eine Markov-Kette mit Übergangswahrscheinlichkeiten

$$P(Y_{n+1} = j | Y_n = i) = (\pi_j / \pi_i) p_{ji}$$

Reversible Markov-Ketten II

Man sagt nun **X** ist **reversibel**, falls X und Y identische Übergangswahrscheinlichkeiten haben, d.h. falls

$$\pi_i p_{ij} = \pi_j p_{ji}$$

für alle $i, j \in S$ gilt.

Beispiele für reversible Markov-Ketten

1. Alle irreduziblen Markov-Ketten mit zwei Zuständen sind reversibel (Beweis in Vorlesung).

2. Markov-Ketten mit tri-diagonaler Übergangsmatrix **P** sind reversibel, z.B.

   - der **random walk** auf endlichem Zustandsraum $S = \{0, 1, \ldots, b\}$
   - der Prozess aus Beispiel 2 (Stichwort: Genhäufigkeit)

Satz über die stationäre Verteilung

Sei $\mathbf{X}$ eine irreduzible Markov-Kette mit Übergangsmatrix $\mathbf{P}$. Ferner gebe es eine Verteilung $\boldsymbol{\pi}$ mit $\pi_i p_{ij} = \pi_j p_{ji}$ für alle $i, j \in S$.
Dann ist $\boldsymbol{\pi}$ die stationäre Verteilung und $\mathbf{X}$ ist bzgl. $\boldsymbol{\pi}$ reversibel.
Beweis:

$$\sum_i \pi_i p_{ij} = \sum_i \pi_j p_{ji} = \pi_j \sum_i p_{ji} = \pi_j$$

Das Grenzwerttheorem

Eine irreduzible und aperiodische Markov-Kette konvergiert gegen ihre stationäre Verteilung $\boldsymbol{\pi}$

$$p_{ij}(n) \longrightarrow \pi_j = \mu_j^{-1} \quad \text{für } n \to \infty \text{ und alle } i$$

bzw.

$$\mathbf{P}_n = \mathbf{P}^n \longrightarrow \begin{pmatrix} \cdots & \boldsymbol{\pi} & \cdots \\ \cdots & \boldsymbol{\pi} & \cdots \\ \vdots & \vdots & \vdots \\ \cdots & \boldsymbol{\pi} & \cdots \end{pmatrix}$$

Daher gilt $\boldsymbol{\mu}^{(0)}\mathbf{P}_n \longrightarrow \boldsymbol{\pi}$ für alle $\boldsymbol{\mu}^{(0)}$.

Ein Gegenbeispiel

Eine Markov-Kette $\mathbf{X}$ mit Übergangsmatrix

$$\mathbf{P} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$$

hat zwar die stationäre Verteilung $\boldsymbol{\pi} = (1/3, 1/3, 1/3)$, konvergiert aber nicht gegen diese, da die Kette periodisch ist.

7.4 Inferenz für Markov-Ketten

- Schätzung der Übergangswahrscheinlichkeiten
- Test auf Modellanpassung
- Allgemeinere Markov-Modelle:
  - Markov-Ketten höherer Ordnung
  - Hidden-Markov Modelle

Schätzung der Übergangswahrscheinlichkeiten

Ziel: Schätzung der Übergangswahrscheinlichkeiten basierend auf einer (oder mehrerer) Realisationen einer Markov-Kette $\mathbf{X}$.

Es erscheint plausibel, die Übergangswahrscheinlichkeiten p_{ij} durch die entsprechenden **Übergangshäufigkeiten**

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

zu schätzen, wobei n_{ij} die Anzahl der beobachteten Übergänge von i nach j ist und $n_i = \sum_j n_{ij}$.

Im Folgenden zeigen wir, dass dies auch die ML-Schätzer sind.

ML-Schätzung der Übergangswahrscheinlichkeiten

Grundlage: Realisation $X_0 = x_0, X_1 = x_1, \ldots, X_N = x_N$ einer Markov-Kette $\mathbf{X}$.

Die Likelihood ist somit (vergleiche Folie S. 13)

$$L(\mathbf{P}) = \mu_{x_0}^{(0)} \prod_{t=1}^{n} p_{x_{t-1},x_t} = \mu_{x_0}^{(0)} \prod_{i,j} p_{ij}^{n_{ij}}$$

wobei n_{ij} die Anzahl der beobachteten Übergänge von i nach j ist.
Die Log-Likelihood ist dann

$$l(\mathbf{P}) = \log(\mu_{x_0}^{(0)}) + \sum_{i,j} n_{ij} \log(p_{ij})$$

ML-Schätzung der Übergangswahrscheinlichkeiten II

Problem: Es sind mehrerer Parameter in $\boldsymbol{\theta} = \mathbf{P}$ durch Maximierung der Log-Likelihood $l(\mathbf{P})$ zu schätzen, wobei noch die Restriktion

$$\sum_j p_{ij} = 1$$

für alle i zu berücksichtigen ist.

$\Rightarrow$ *Lagrangesche Multiplikatorenmethode*:

Maximiere

$$l^*(\mathbf{P}) = \log(\mu_{x_0}^{(0)}) + \sum_{i,j} n_{ij} \log(p_{ij}) - \sum_i \lambda_i \left(\sum_j p_{ij} - 1 \right)$$

ML-Schätzung der Übergangswahrscheinlichkeiten III

Die partiellen Ableitungen nach p_{ij} sind somit

$$\frac{dl^*(\mathbf{P})}{p_{ij}} = \frac{n_{ij}}{p_{ij}} - \lambda_i$$

Nullsetzen liefert $n_{ij} = \lambda_i p_{ij}$. Durch Summation über j folgt

$$\lambda_i = \sum_j n_{ij} = n_i$$

und schließlich

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

Beispiel: Regen bei den Snoqualmie Wasserfällen

Über eine Zeitreihe von $N = 13149$ Tagen wurde an den Snoqualmie Wasserfällen registriert, ob es am jeweiligen Tag geregnet hat oder nicht.

Erstellung eines Markov-Modells:

Der Zustandsraum ist $S = \{0, 1\}$ mit

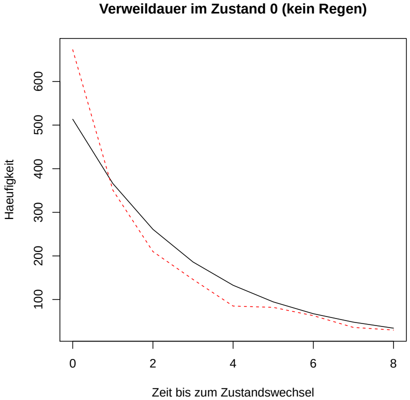
$$\left.\begin{array}{ll} 0: & \text{Kein Regen} \\ 1: & \text{Regen} \end{array}\right\} \text{ am Tag } t = 1, \ldots, N$$

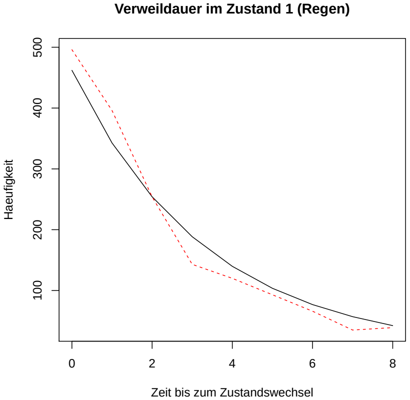
Es ergibt sich $n_0 = 6229$ (Tage ohne Regen) und $n_1 = 6920$.
Übergangsmatrix mit relativen Übergangshäufigkeiten (*ML*-Schätzer):

$$\hat{\mathbf{P}} = \begin{pmatrix} 0.713 & 0.287 \\ 0.258 & 0.742 \end{pmatrix}$$

Frage: Passt sich das Markov-Modell den Daten gut an?
→ Betrachtung der “Verweildauern” (Wartezeiten bis zum nächsten Zustandswechsel)

Verweildauern bei den Snoqualmie Wasserfällen





Theoretische (schwarz) und empirische (rot) Verweildauern in den beiden Zuständen.

Markov-Ketten höherer Ordnung

Zum Beispiel Markov-Kette *zweiter* Ordnung:

Die Regenwahrscheinlichkeit hängt nun von den letzten *zwei* Tagen ab. Die Übergangswahrscheinlichkeiten sind

$$P(X_n = s | X_{n-1} = x_{n-1}, X_{n-2} = x_{n-2}).$$

Diese Markov-Kette kann auch als Markov-Kette erster Ordnung dargestellt werden, indem man den Zustandsraum zu $S = \{00, 01, 10, 11\}$ erweitert, wobei der erste Eintrag x_{n-2} und der zweite Eintrag x_{n-1} darstellt.
$\rightarrow$ In der Übergangsmatrix ergeben sich dann **strukturelle Nullen**.

Strukturelle Nullen

Im Beispiel ergibt sich:

$$\mathbf{P} = \begin{array}{c||c} & \begin{matrix} 00 & 01 & 10 & 11 \end{matrix} \\ \hline\hline \begin{matrix} 00 \\ 01 \\ 10 \\ 11 \end{matrix} & \begin{pmatrix} & & 0 & 0 \\ 0 & 0 & & \\ & & 0 & 0 \\ 0 & 0 & & \end{pmatrix} \end{array}$$

da z.B. auf $X_{n-2} = 0$ und $X_{n-1} = 0$ nicht $X_{n-1} = 1$ und $X_n = 0$ folgen kann etc.

$\rightarrow$ die Anzahl der Spalten kann reduziert werden

Regen bei den Snoqualmie Wasserfällen

Mit reduzierten Spalten ergibt sich im Beispiel

$$\hat{\mathbf{P}} = \left(\begin{array}{c||cc} & 0 & 1 \\ \hline\hline 00 & 0.749 & 0.251 \\ 01 & 0.277 & 0.723 \\ 10 & 0.624 & 0.376 \\ 11 & 0.252 & 0.748 \end{array}\right)$$

Alternativer Ansatz: Hidden Markov-Modell

7.5 Hidden Markov Modelle

- **Mischverteilungsmodell** mit zusätzlichem Zeitreihencharakter
- **Anwendungen** in verschiedenen Bereichen:
  - Ökonometrie,
  - Genetik (DNA-Sequenzierung, Stammbaumanalyse),
  - Spracherkennung, ...

Das Modell

Latente Parameter $\mathbf{X} = (X_1, \ldots, X_N)$ folgen einer (homogenen) **Markov-Kette** mit diskretem Zustandsraum S:

- Übergangsmatrix $\mathbf{P}$ mit $p_{ij} = P(X_t = j | X_{t-1} = i)$
- Anfangsverteilung $\boldsymbol{\pi}$ mit $\pi_i = P(X_0 = i)$
  (oft impliziert durch $\boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{P}$)

Die Beobachtungen $y_t | x_t = s$ sind **bedingt unabhängig** aus einer Verteilung mit Wahrscheinlichkeitsfunktion $f_s(y_t)$ mit Parametern $\boldsymbol{\theta}_s$.
Beispielsweise:

- **diskret** mit Missklassifizierungswahrscheinlichkeiten p_s
- **Poisson** mit Raten λ_s

Beispiel: Ein verrauschtes binäres Signal

Daten:

$$\mathbf{y} = (2\ 2\ 2\ 1\ 2\ 2\ 1\ 1\ 1\ 1\ 1\ 2\ 1\ 1\ 1\ 2\ 1\ 2\ 2\ 2)$$

$$\begin{array}{l} |S| = 2 \\ N = 20 \end{array} \qquad \mathbf{P} = \begin{pmatrix} 0.75 & 0.25 \\ 0.25 & 0.75 \end{pmatrix} \qquad \boldsymbol{\pi} = \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix}$$

$$\begin{array}{ll} f(y_t = 1 | x_t = 1) = 0.8 & f(y_t = 1 | x_t = 2) = 0.2 \\ f(y_t = 2 | x_t = 1) = 0.2 & f(y_t = 2 | x_t = 2) = 0.8 \end{array}$$

Ziel der statistischen Inferenz: **Restauration** der Sequenz $\mathbf{X}$

Inferenz bei festen Hyperparametern

- Hyperparameter: $\mathbf{P}$, $\boldsymbol{\pi}$, $\boldsymbol{\theta}$ fest
- Ziel: **Schätzung** der latenten Zustände $\mathbf{x} = (x_1, \dots, x_N)$
- **Posteriori-Verteilung** $f(\mathbf{x}|\mathbf{y}) = f(\mathbf{x}, \mathbf{y})/f(\mathbf{y})$ mit:

$$f(\mathbf{x}|\mathbf{y}) \propto f(\mathbf{x}, \mathbf{y}) = \underbrace{\pi_{x_1} \prod_{t=2}^{N} p_{x_{t-1}x_t}}_{f(\mathbf{x})} \cdot \underbrace{\prod_{t=1}^{N} f_{x_t}(y_t)}_{f(\mathbf{y}|\mathbf{x})}$$

- **Problem**: Es gibt S^N (!) unterschiedliche Sequenzen $\mathbf{x}$.

Posteriori-Modus Schätzung

- **Viterbi-Algorithmus** [Viterbi (1967)]
  - Rekursiver Algorithmus zur Maximierung von $f(\mathbf{x}, \mathbf{y})$ bzgl. $\mathbf{x}$
  - liefert (einen) MAP-(posteriori Modus)-Schätzer von $f(\mathbf{x}|\mathbf{y}) \propto f(\mathbf{x}, \mathbf{y})$
  - Algorithmus ist numerisch effizient: $O(|S|^2 \cdot N)$
- **Simulated annealing** [Kirkpatrick, Gelatt & Vecchi (1983)]

Rekonstruktion des verrauschten binären Signals

Das empfangene Signal war

$$\mathbf{y} = \quad (2\ 2\ 2\ 1\ 2\ 2\ 1\ 1\ 1\ 1\ 1\ 2\ 1\ 1\ 1\ 2\ 1\ 2\ 2\ 2)$$

Daraus kann man Schätzer für das zugrundeliegende Signal $\mathbf{x}$ berechnen:

Schätzer von $\mathbf{x}$	post. Wkeit $P(\mathbf{x}\|\mathbf{y})$
$\hat{\mathbf{x}}_{MAP_1} = (2\ 2\ 2\ 2\ 2\ 2\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 2\ 2\ 2)$	0.0304
$\hat{\mathbf{x}}_{MAP_2} = (2\ 2\ 2\ 2\ 2\ 2\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 2\ 2\ 2\ 2\ 2)$	0.0304
$\hat{\mathbf{x}}_{MPM} = (2\ 2\ 2\ 2\ 2\ 2\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 2\ 1\ 2\ 2\ 2)$	0.0135
$\mathbf{y} = (2\ 2\ 2\ 1\ 2\ 2\ 1\ 1\ 1\ 1\ 1\ 2\ 1\ 1\ 1\ 2\ 1\ 2\ 2\ 2)$	0.0027

Hierbei bezeichnet $\hat{\mathbf{x}}_{MPM}$ den marginalen Posteriori Modus, d.h. jedes x_i, $i = 1, \ldots, 20$, in $\hat{\mathbf{x}}_{MPM}$ hat marginale Posteriori-Wahrscheinlichkeit $P(x_i|\mathbf{y}) > 0.5$.

Inferenz bei unbekannten Hyperparametern

Seinen nun bestimmte Hyperparameter $\boldsymbol{\theta}$ unbekannt, wie beispielsweise die Übergangsmatrix $\mathbf{P}$ oder die Verteilung $f(y_i|x_i)$.

- **Klassische** Likelihood-Ansätze maximieren die (marginale) Likelihood $L(\boldsymbol{\theta}) = f(\mathbf{y}|\boldsymbol{\theta})$ bezüglich $\boldsymbol{\theta}$, z.B. mit dem Baum-Welch-Algorithmus, der einen Spezialfall des EM-Algorithmus darstellt.

  Problem: Berechnung von $f(\mathbf{y}|\boldsymbol{\theta}) = \sum_{\mathbf{x}} f(\mathbf{x}, \mathbf{y}|\boldsymbol{\theta})$

- **Bayesianische** Ansätze verwenden zusätzliche priori-Verteilungen $f(\boldsymbol{\theta})$ und simulieren aus der posteriori-Verteilung

  $$f(\mathbf{x}, \boldsymbol{\theta}|\mathbf{y}) \propto f(\mathbf{x}, \mathbf{y}|\boldsymbol{\theta}) \cdot f(\boldsymbol{\theta})$$

  mit Markov-Ketten Monte Carlo (MCMC) Verfahren.

Beispiel zur Likelihood-Inferenz

Seinen nun die Einträge der Übergangsmatrix **P** unbekannt. Die marginale Likelihood $L(p_{11}, p_{22})$ der Diagonalelemente p_{11} und p_{22} ist in folgender Graphik dargestellt. Die ML-Schätzungen sind $\hat{p}_{11} = 0.85$ und $\hat{p}_{22} = 0.78$.