Stochastik für die Informatik, Vorlesung 24

Inhalt

- Konvergenz gegen die invariante Verteilung
- Anwendungen des Konvergenztheorems
- Einführung in die Informationstheorie, Präfixcodes

Lernziele

- Die Bedeutung von invarianten Verteilungen kennen
- Anwendungen des Konvergenztheorems kennen

Vorkenntnisse: Markovketten, stochastische Matrizen, Matrizenrechnung, lineare Algebra.

Konvergenz gegen die invariante Verteilung

(Theorem 11.7) Sei $(X_i)_{i\in\mathbb{N}_0}$ eine irreduzible und aperiodische homogene Markov-Kette mit Zustandsraum S. Sei $(\pi_a)_{a\in S}$ eine invariante Verteilung. Dann gilt für alle $a \in S$, unabhängig von der Startverteilung,

$$\lim_{n\to\infty} \mathbb{P}(X_n = a) = \pi_a.$$

- Egal wo man startet: Für große n gilt $\mathbb{P}(X_n = a) \approx \pi_a$.
- Falls die Kette in einer (beliebigen) Startverteilung ν startet, befindet sie sich nach hinreichend langer Zeit in der stationären Verteilung
- Information über Struktur der Kette. Bsp. Netzwerk: Knoten mit vielen Kanten werden oft besucht, für solche Knoten a ist π_a größer als für solche die selten besucht werden.

Konvergenz gegen die invariante Verteilung

(Theorem 11.7) Sei $(X_i)_{i\in\mathbb{N}_0}$ eine irreduzible und aperiodische homogene Markov-Kette mit Zustandsraum S. Sei $(\pi_a)_{a\in S}$ eine invariante Verteilung. Dann gilt für alle $a \in S$, unabhängig von der Startverteilung,

$$\lim_{n\to\infty} \mathbb{P}(X_n = a) = \pi_a.$$

- Egal wo man startet: Für große n gilt $\mathbb{P}(X_n = a) \approx \pi_a$.
- Falls die Kette in einer (beliebigen) Startverteilung ν startet, befindet sie sich nach hinreichend langer Zeit in der stationären Verteilung
- Information über Struktur der Kette. Bsp. Netzwerk: Knoten mit vielen Kanten werden oft besucht, für solche Knoten a ist π_a größer als für solche die selten besucht werden.

Beispiel 11.13 Warteschlangen

- Zeitabschnitte von fester Länge
- Parameter $k \in \mathbb{N}, \mu \in [0,1], \lambda \in [0,1]$, mit $\lambda + \mu \leq 1$.
- In jedem Zeitabschnitt geschieht eines der folgenden Ereignisse:
  - Falls weniger als k Kunden da sind, so kommt ein neuer Kunde mit Wahrscheinlichkeit λ an
  - Falls mindestens ein Kunde da ist, so verlässt ein Kunde die Schlange mit Wahrscheinlichkeit μ
  - Andernfalls geschieht nichts
- X_n:=Anzahl Kunden in der Schlange im $n-$ten Zeitabschnitt
- $(X_n)_{n \in \mathbb{N}_0}$ ist Markov-Kette mit Zustandsraum $S = \{0, ..., k\}$

Beispiel 11.13 Warteschlangen

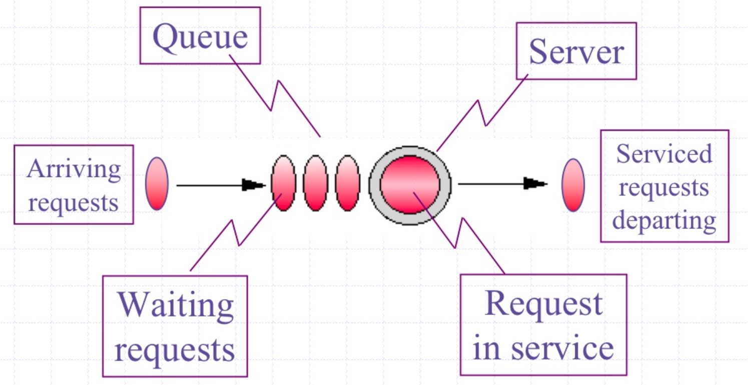

Bild: http://www.perfdynamics.com/

Beispiel 11.13 Warteschlangen

Mögliche Übergänge:

$$\begin{aligned} p_{i,i+1} &= \lambda \text{ falls } i < k \\ p_{i,i-1} &= \mu \text{ falls } i > 0. \end{aligned}$$

d.h.

$$P = \begin{bmatrix} 1-\lambda & \lambda & 0 & \cdots & & & \cdots & 0 \\ \mu & 1-\mu-\lambda & \lambda & 0 & \cdots & & & 0 \\ 0 & \mu & 1-\mu-\lambda & \lambda & 0 & \cdots & & 0 \\ \vdots & & \ddots & \ddots & \ddots & & & \vdots \\ 0 & \cdots & & & 0 & \mu & 1-\mu-\lambda & \lambda \\ 0 & \cdots & & & & 0 & \mu & 1-\mu \end{bmatrix}$$

Frage: Wie viele Kunden sind auf lange Sicht in der Schlange?
Stabilität für $k \to \infty$?

Beispiel 11.13 Warteschlangen

Mögliche Übergänge:

$$\begin{aligned} p_{i,i+1} &= \lambda \text{ falls } i < k \\ p_{i,i-1} &= \mu \text{ falls } i > 0. \end{aligned}$$

d.h.

$$P = \begin{bmatrix} 1-\lambda & \lambda & 0 & \cdots & & & \cdots & 0 \\ \mu & 1-\mu-\lambda & \lambda & 0 & \cdots & & & 0 \\ 0 & \mu & 1-\mu-\lambda & \lambda & 0 & \cdots & & 0 \\ \vdots & & \ddots & \ddots & \ddots & & & \vdots \\ 0 & \cdots & & & 0 & \mu & 1-\mu-\lambda & \lambda \\ 0 & \cdots & & & & 0 & \mu & 1-\mu \end{bmatrix}$$

Frage: Wie viele Kunden sind auf lange Sicht in der Schlange?
Stabilität für $k \to \infty$?

Beispiel 11.13 Warteschlangen

(Satz 11.9) Die invariante Verteilung der beschriebenen Warteschlange ist gegeben durch

$$\pi_i = \frac{1}{\sum_{j=0}^{k}(\lambda/\mu)^j} \cdot \left(\frac{\lambda}{\mu}\right)^i, \quad i = 0, \ldots, k$$

Was passiert für $k \to \infty$?

- Falls $\lambda/\mu < 1$ ist, gilt $\sum_{j=0}^{\infty} \left(\frac{\lambda}{\mu}\right)^j = \frac{\mu}{\mu-\lambda} < \infty$, die Warteschlange hat für $k \to \infty$ eine invariante Verteilung, ist stabil
- Andernfalls ist $\sum_{j=0}^{\infty} \left(\frac{\lambda}{\mu}\right)^j = \infty$, die Warteschlange ist instabil

Kapitel 16: Markov-Ketten Monte Carlo (MCMC)

Ziel: Simulation aus einer vorgegebenen Verteilung $\pi = (\pi_x)_{x \in S}$.

Grundidee: Konstruiere eine Markov-Kette $X = (X_n)_{n \in \mathbb{N}}$, welche π als invariante Verteilung hat. Lasse dann die Markov-Kette lange genug laufen. Nach n Schritten, für n hinreichend groß , hat nach dem Satz auf der letzten Folie X_n annähernd die Verteilung π.

- Konkrete Konstruktion einer solchen Kette?
- Benötigte Zeit für die Konvergenz?

Metropolis-Algorithmus

(Metropolis-Hastings-Algorithmus, Simulated Annealing)

Fragestellung: Konstruktion einer Markov-Kette auf einem endlichen Zustandsraum S mit invarianter Verteilung π.

Vorgehen: Zweistufiges Verfahren

1. Konstruiere eine einfache Referenzkette, welche zulässige Übergänge vorschlägt
2. Entscheide bei jedem vorgeschlagenen Übergang, ob er tatsächlich ausgeführt wird oder nicht. Dabei sollen Übergänge welche in einen Zustand mit höherer Wahrscheinlichkeit bezüglich π führen, bevorzug ausgeführt werden.

Metropolis-Algorithmus

Gegeben Verteilung π auf endlicher Menge S. Ziel: Simulation aus dieser Verteilung.

Algorithmus:

- Sei $Q = (q_{k,l})_{k,l \in S}$ eine beliebige Übergangsmatrix auf S. Die zugeörige Markov-Kette ist die Referenzkette.
- Definiere eine neue Übergangsmatrix $P = (p_{k,l})_{k,l \in S}$ mit Hilfe von Q als

$$p_{k,l} = q_{k,l} \cdot \min\left(1, \frac{\pi(l) q_{lk}}{\pi(k) q_{kl}}\right) \quad \text{falls } k \neq l$$

  und $p_{k,k} = 1 - \sum_{l \neq k} p_{k,l}$ sonst.
- Starte eine Markov-Kette in einem beliebigen Punkt von S, und führe Übergänge entsprechend der Matrix P aus.

Metropolis-Algorithmus

- Q schlägt Übergänge vor, welche mit Wahrscheinlichkeit $\min\left(1, \frac{\pi(l)q_{lk}}{\pi(k)q_{kl}}\right)$ tatsächlich ausgeführt werden
- π ist invariante Verteilung für P (Beweis)
- (Satz 16.3) Falls P irreduzibel und aperiodisch ist, konvergiert die Markov-Kette gegen die gesuchte Verteilung π (Monte-Carlo-Simulation)
- Konkrete Berechnung: Braucht die Verhältnisse $\pi(l)q_{lk}/\pi(k)q_{kl}$ zu kennen, aber nicht die einzelnen Werte
- Wahl von Q kann geschickt ans Problem angepasst werden
- Verbleibende Frage: Wie lange muss man die Markov-Kette laufen lassen, um eine vernünftige Genauigkeit zu erreichen?

Beispiel 16.1: Simulation bedingter Verteilungen

- Gegeben: $Y_1, \dots, Y_N$ unabhängige, identisch verteilte, diskrete Zufallsvariablen mit bekannter Verteilung $r_i, i \in \mathbb{N}_0$.
- Gesucht: Bedingte Verteilung von $(Y_1, \dots, Y_N)$ gegeben $Y_1 + \dots + Y_N = M$, für ein vorgegebenes $M \in \mathbb{N}$.
- Notation: Für $k = (k_1, \dots, k_N) \in \mathbb{N}_0^N$ schreibe

$$\pi(k) := \mathbb{P}(Y_1 = k_1, \dots, Y_N = k_N \mid Y_1 + \dots + Y_N = M)$$

  für die gesuchte bedingte Verteilung von $Y_1, \dots, Y_N$ gegeben $Y_1 + \dots + Y_N = M$ (abhängig von N und M, und der Verteilung $r_i, i \in \mathbb{N}_0$).

Beispiel: Simulation bedingter Verteilungen

In der Situation der letzten Folie hat die gesuchte Verteilung π die Form

$$\pi(k) = \frac{1}{C(N,M)} \cdot r_{k_1} \cdot \ldots . \cdot r_{k_N},$$

für $k = (k_1, ..., k_N) \in \mathbb{N}_0^N, k_1 + ... + k_N = M$

- Da im Metropolis-Algorithmus nur die Verhältnisse $\pi(k)/\pi(l)$ auftreten, braucht man $C(N,M)$ nicht zu bestimmen
- (Konstruiere die Referenzkette)
- (Konstruiere die Metropolis-Kette)

Konvergenzgeschwindigkeit

- Aus dem Theorem über die Konvergenz einer (irreduzibler und aperiodischer) Markov-Kette $(X_n)_{n\in\mathbb{N}}$ gegen ihre invariante Verteilung π wissen wir: Für jedes $a \in S$ und für große n gilt

$$\mathbb{P}(X_n = a) \approx \pi(a).$$

- Frage: Wie kann man große n und $\approx$ quantifizieren?

Konvergenzgeschwindigkeit

(Def. 16.1) Sei $(X_n)_{n\in\mathbb{N}_0}$ eine Markov-Kette mit invarianter Verteilung $\{\pi_a\}_{a\in S}$. Die $\varepsilon-$Mischzeit $\tau(\varepsilon)$ der Kette ist definiert als das kleinste $n > 0$ so dass für alle $a \in S$ und für beliebige Startverteilung gilt:

$$|\mathbb{P}(X_n = a) - \pi(a)| \leq \varepsilon.$$

- In Worten: Für $n = \tau(\varepsilon)$ ist der Abstand von der invarianten Verteilung höchstens ε.
- Man kann sogar beweisen: Für alle n welche größer sind als die $\varepsilon-$Mischzeit ist der Abstand von der invarianten Verteilung höchstens ε.
- Oft auch: $\varepsilon = 1/4$ fest gewählt.
- Berechnung der Mischzeit?

Spektrallücke

(Satz 16.2) Sei S eine endliche Menge mit $|S| = N$. Sei P die Übergangsmatrix einer irreduziblen, aperiodischen Markov-Kette auf S. Dann sind alle Eigenwerte dieser Matrix reell und dem Betrag nach kleiner oder gleich 1.

(Ohne Beweis)

(Def. 16.2) Sei λ^* der dem Betrage nach größte Eigenwert von P, welcher $\neq 1$ ist. Dann heißt

$$\gamma^* := 1 - |\lambda^*|$$

die Spektrallücke von P (bzw. der zugehörigen Markov-Kette).

Spektrallücke und Mischzeit

(Satz 16.3) Sei $(X_n)_{n\in\mathbb{N}}$ eine Markov-Kette auf einem endlichen Zustandsraum S, mit Übergangsmatrix P und invarianter Verteilung $\{\pi_a\}_{a\in S}$, welche zusätzlich $\pi_a p_{a,b} = \pi_b p_{b,a}$ für alle a, b erfüllt. Sei $\bar{\pi}$ das kleinste $\pi_a, a \in S$. Für die ε-Mischzeit gilt dann

$$\tau(\varepsilon) \leq \log(1/(\bar{\pi}\varepsilon)) \cdot \frac{1}{\gamma^*}.$$

(Ohne Beweis)

- Daraus folgt insbesondere für $a \in S, n \in \mathbb{N}$

$$|P(X_n = a) - \pi_a| \leq \frac{1}{\bar{\pi}} e^{-\gamma^* n}$$

- Je größer die Spektrallücke, desto schneller konvergiert die Markov-Kette gegen die invariante Verteilung.

Spektrallücke und Mischzeit

- Um die Aussage $\mathbb{P}(X_n = a) \approx \pi(a)$ präzise zu machen, geht man also folgendermaßen vor:
  - Wähle die Präzision ε.
  - Berechne π und untersuche ob die zusätzliche Bedingung erfüllt ist.
  - Berechne die Eigenwerte von P.
  - Berechne $\bar{\pi}$ und die Spektrallücke γ^*.
  - Dann gilt für alle $n \geq \log(1/(\bar{\pi}\varepsilon))\frac{1}{\gamma^*}$, dass

    $$|\mathbb{P}(X_n = a) - \pi(a)| \leq \varepsilon$$

    für alle a und jede Startverteilung gilt.
- Anwendung: Simulation von Verteilungen mit einer vorgegebenen Präzision.
- Die Bestimmung von π und γ^* ist ein einfaches Problem der linearen Algebra.

Kapitel X: Informationstheorie

Literatur: G. Kersting, A. Wakolbinger: Elementare Stochastik; Kapitel VI: Ideen aus der Informationstheorie.

Verfügbar in der Bibliothek, auch als E-Book.

Inhalt: Stochastische Aspekte der Informationsübermittlung: Codierung, Redundanz, Entropie...

Grundbegriffe

- Ein Alphabet ist eine höchstens abzählbare Menge S.
- Ein Buchstabe ist ein Element eines Alphabets.
- Ein binärer Code ist eine *injektive* Abbildung

$$k : S \to \bigcup_{l \geq 1} \{0,1\}^l,$$

also eine Abbildung, die jedem Buchstaben $a \in S$ eine Folge $k(a) = k_1(a), \ldots, k_l(a)$ der Länge $l = \ell(a)$ mit Folgengliedern $k_i(a) \in \{0,1\}$ als Codewort zuordnet.

Präfixcodes

Länge als Maß für die Güte des Codes: Möglichst kurz, aber auch leicht zu entschlüsseln

- Ein Präfixcode ist ein binärer Code, bei dem kein Codewort Anfangsstück eines anderen Codeworts ist. Das bedeutet: Für $a, b \in S$ mit $a \neq b$ gibt es *keine* Folge $f = (f_1, ..., f_m) \in \{0, 1\}^m$ mit $k(a)f = k(b)$.

- (Code als Fragestrategie)
- (Baumdarstellung)

Fano-Kraft-Ungleichung

(Satz) Sei k ein Präfixcode für ein Alphabet S. Sei $\ell(a)$ die Länge des Codewortes von $a \in S$. Dann gilt

$$\sum_{a \in S} 2^{-\ell(a)} \leq 1$$

- (Beweis)
- (Gleichheit, strikte Ungleichheit)