AlMa II - Überblick Stochastik

Jochen Garcke

universitätbonn

Fraunhofer SCAI

Grundbegriffe

- Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ beschreibt Zufallssituation
- Realisierung, Stichprobe, Elementarereignis $\omega \in \Omega$ Ergebnisraum
- zufälliges Ereignis $A \subseteq \Omega$
- $\mathcal{A}$ als σ − *Algebra* gibt wohlgestellte Operationen für Ereignisse
- Wahrscheinlichkeitsvert. P auf $(\Omega, \mathcal{A})$: positiv, normiert, σ-additiv
- Rechenregeln basieren auf Mengenoperationen
- zum Rechnen brauchen wir Kombinatorik:
  Formeln für Anzahl der Möglichkeiten
  - mit/ohne zurücklegen
  - mit/ohne Berücksichtigung der Reihenfolge
- Vorsicht bei der Modellierung von realen Aufgabenstellungen

Bedingte Wahrscheinlichkeiten

bedingte Wahrscheinlichkeit von A gegeben B

$$P(A|B) := \frac{P(A \cap B)}{P(B)} \qquad \text{für } P(B) \neq 0$$

- Vorsicht bei der Modellierung von realen Aufgabenstellungen
- Multiplikationsregeln
- totale Wahrscheinlichkeit (mittels Zerlegung) $P(A) = \sum P(A|B_i)P(B_i)$
- mehrstufige Experimente entlang Pfaden

Bayes

$$P(B_j|A) := \frac{P(A|B_j)P(B_j)}{P(A)}$$

Stochastische Unabhängigkeit

A und B stochastisch unabhängig

$$P(A \cap B) = P(A)P(B)$$

- stochastisch Ausage, A und B können sich real beeinflussen
- für stoch. Unab. von $\{A_i\}$ muss das für jede Teilauswahl gelten
- Vorsicht bei der Modellierung von realen Aufgabenstellungen
- einfaches Modell ist Produktraum: unabhängige Hintereinanderausführung von Experimenten

Zufallsvariable

- betrachten $X : \Omega \rightarrow R$
- ordnet Elementarereignissen Zahlen zu
- sowie W-keiten darüber $P(X = x_i) = P(\{\omega \in \Omega\} | X(\omega) = x_i)$
- untersuchen wie häufig Werte von X auftreten
- insbesondere diskrete ZV: bis zu abzählbar unendlich viele Werte
- Wahrscheinlichkeitsfunktion $p_X(x_i) = p_i = P(X = x_i)$
- mehrdimensional
  $p_X(x) = p_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = P(X_1 = x_1, \ldots, X_n = x_n)$
- marginale Wahrscheinlichkeit:
  $p_{X_2,X_4,X_7}(x_2, x_4, x_7)$ wegsummieren der "anderen" $X_i, i \notin 2, 4, 7$
- Rechenregeln basieren wieder auf Mengenoperationen

Erwartungswert und Varianz

- Erwartungswert $E(X) = \sum x_i p_i$, $X \in \mathcal{L}^1$
- Transformationssatz (diskret): $E(g(X)) = \sum g(x_i)p_i$
- Varianz $Var(X) = E([X - E(X)]^2) = E(X^2) - E(X)^2$, $\sigma_X = \sqrt{Var(X)}$, $X \in \mathcal{L}^2$
- Kovarianz $Cov(X, Y) = E([X - E(X)][y - E(Y)]) = E(XY) - E(X)E(Y)$, $X, Y \in \mathcal{L}^2$
- unkorreliert falls $Cov(X, Y) = 0$
- elementare Rechenregeln basierend auf den vorherigen Aussagen
- stoch. unabhängig $P(X_1 = x_1, \ldots, X_n = x_n) = P(X_1 = x_1) \cdots P(X_n = x_n)$ ist einfacher, gilt direkt für Teilfamilien (Gegensatz zu Ereignismengen)
- X, Y stoch. unabhängig g.d.w. $f(X), g(Y)$ unkorreliert für alle $f(X), g(Y) \in \mathcal{L}^2$

Konvergenzsätze

Tschebyscheff-Ungleichung

$$P(|X - E(X)| \geq \epsilon) \leq \frac{Var(X)}{\epsilon^2}$$

oder

$$P(|X - E(X)| \geq k\sigma_x) \leq \frac{1}{k^2}$$

schwaches Gesetz der grossen Zahlen

$X_1, \ldots, X_n$ paarweise unkorreliert, $M_n = \max Var(X_i)$, $S_n = \sum X_i$. Dann gilt für alle $\epsilon > 0$

$$P(|S/n - E(S_n)/n| \geq \epsilon) \leq \frac{M_n}{\epsilon^2 n}$$

Falls M_n/n Nullfolge und $E(X_i) = S\ \forall i$, so konvergiert $\{S_n/n\}$ stochastisch gegen S, d.h. für $n \to \infty$

$$P(|S_n/n - S| \geq \epsilon) \to 0,$$

Monte Carlo-Verfahren

Monte Carlo-Schätzer

Eine Approximation für den Erwartungswert von g

$$E_p(g) = \sum g(a)p(a)$$

ist mit X_i mit Wahrscheinlichkeitsfunktion P

$$\eta_n(\omega) := \frac{1}{n} \sum g(X_i(\omega))$$

Es gilt

$$E(|\eta_n - E_p(g)|^2) = E(|\eta_n - E(\eta_n)|^2) = Var(\eta_n) = \frac{1}{n} Var_p(g)$$

bzw.

$$P(\eta_n - E_p(g)| \geq \epsilon) \leq \frac{1}{n\epsilon^2} Var_p(g)$$

Pseudozufallszahlen

gleichverteilte Pseudozufallszahlen

Rekurrenzrelationen vom Typ

$$x_{n+1} = (ax_n + b) \mod m, n = 0, 1, 2, \dots$$

heißen lineare Kongruenzrelationen und liefern als Generatorvorschrift gleichverteilte Pseudozufallszahlen zwischen in 0 und m-1
(erster "richtiger" Algorithmus in der Vorlesung)

Pseudozufallszahlen verteilt nach P

Auf $A = \{a_1, \dots\}$ haben wir W-verteilung P mit $p_i = p(a_i)$. Mit der kommulativen Verteilungsfunktion $s_n = \sum_{i=1}^{n} p_i$ und einer in $[0, 1)$ gleichverteilten Zufallsvariable U setzen wir

$$X(\omega) = a_i \text{ falls } s_{i-1} < U(\omega) \leq s_i$$

Dann ist $X \sim P$
(zweiter "richtiger" Algorithmus in der Vorlesung)

Direkte Simulation

Algorithm 1: Schritt eines linearen Kongruenzgenerator, **muss** mit Ergebnis wieder aufgerufen werden

Data: Stichprobe xn, Konstanten a, b, m
Result: Pseudozufallsstichprobe x gleichverteilt zwischen 0 und m-1
return (a xn + b) mod m

Algorithm 2: direkte Simulation einer Pseudozufallsstichprobe x von

Data: Ereignisse a(1), a(2),..., Gewichte p(1), p(2), ...,
Result: Pseudozufallsstichprobe x von *P*
n = 1
s = p(1)
erzeuge in (0,1) gleichverteilte Pseudozufallsstichprobe u
▷ `z.B. linearer Kongruenzgenerator`
while $u > s$ **do** ▷ `wenn endlicher Zustandsraum auch n <= nmax`
    n = n +1
    s = s + p(n)
return x = a(n)

Diskrete Verteilungen, deren Eigenschaften und Anwendungen

- diskrete Gleichverteilung $G(n)$ hat

$$P(X = k) = 1/n$$

- Bernoulli-Verteilung / Null-Eins-Verteilung

$$P(X = 1) = p, P(X = 0) = q = 1 - p$$

- Geometrische Verteilung $Geo(p)$ hat

$$P(X = k) = (1 - p)^k p$$

- Binomialverteilung $Bin(n, p)$ hat

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

- Poissonverteilung $Pos(\mu t)$ hat

$$P(X_t = k) = \frac{(\mu t)^k}{k!} e^{-\mu t}$$

- Hypergeometrische Verteilung $H(n, N, M)$

Stetige Verteilungen, deren Eigenschaften und Anwendungen

Zufallsvariable X ist stetig falls integrierbares (z.B. stetiges) $f(x)$ (Dichtefunktion) existiert mit

$$P(a \leq X \leq b) = \int_a^b f(x)dx$$

- Erwartungswert, Varianz verallgemeinern sich entsprechend
- Stetige Gleichverteilung

$$f(x) = \frac{1}{b-a} \quad \text{falls } x \in [a, b], \text{ sonst } 0$$

- Exponentialverteilung $Ex(\lambda)$ mit Parameter $\lambda > 0$

$$f(x) = \lambda e^{-\lambda x} \quad \text{falls } x \geq 0, \text{ sonst } 0$$

- Normalverteilung $N(\mu, \sigma^2)$ mit Erwartungswert μ und Varianz σ^2

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Markov-Ketten

Markov-Eigenschaft

Vergangenheit spielt bei Übergangswahrscheinlichkeit einer Folge von Zufallsvariable über einem (abzählbar un)endlichen Zustandsraum keine Rolle, d.h.

$$P(X_{n+1} = s_j | X_0 = s_{i_0}, X_1 = s_{i_1}, \ldots, X_{n-1} = s_{i_{n-1}}, X_n = s_i) = P(X_{n+1} = s_j | X_n = s_i)$$

Wert heißt Übergangswahrscheinlichkeit p_{ij}, und kann in Übergangsmatrix P organisiert werden.

- stochastische Matrix, d.h. Zeilensumme 1, nicht-negative Einträge
- rechnen mit P, z.B. $\mu^{(n)} = \mu^{(0)} P^n$ mit Startvert. $\mu^{(0)}$ nach n Schritten
- Chapman-Kolmogorov:

$$(P^{n-l})_{tu} = \sum_{i=1}^{k} (P^{n-m})_{iu} (P^{m-l})_{ti}$$

Stationäre Verteilung einer Markov-Ketten

- stationäre Verteilung / Gleichverteilung ist stochastischer Zeilenvektor mit $\pi = \pi P$
- z.B. π reversibel bzgl. P / Markov-Kette $\pi_i p_{ij} = \pi_j p_{ji}$
- Minorisierungsbedingung $(P^r)_{ij} \geq \delta \pi_j \; \forall i, j$ für ein $\delta \in (0, 1]$, $r \in \mathbb{N}$ liefert Konvergenz für jede Startverteilung gegen (eindeutige) stationäre Verteilung
- für irreduzible Matrix $(P^n)_{ij} > 0$ für s_i, s_j gibt es $n \in \mathbb{N}$ derart
  - stark zusammenhängender Graph
  - von jedem s_i kommt man zu jedem s_j in endlich vielen Schritten
- zusammen mit aperiodisch gilt für irreduzible Matrizen / Ketten die Minorisierungsbedingung, also Konvergenz gegen stationäre Vert.
- für aperiodisch, irreduzible Matrizen / Ketten kann man weiterhin auch Existenz einer stationären Verteilung zeigen

Algorithm 3: Vereinfachter MCMC: $\tilde{Q}$ nur Kantengewichtet

Data: Startwert $i \in S$, ungerichtete, zusammenhängende Graphmatrix $\tilde{Q} = (S, E)$

Result: Realisier. $(x[0], x[1], \dots)$ der Markov-Kette $(X_0, X_1, \dots)$ mit Vert. p

x(0) = i

n = 0

repeat

  bestimme $j \in S$ gleichverteilt unter Nachbarn $(\tilde{q}_{ij})$

  ▷ `wähle Nachbarknoten zufällig gleichverteilt`

  bestimme $\delta \in \{0, 1\}$ gemäß Bernoulli-Verteilung mit $p = a_{ij}$

  ▷ `Metropolis:` $a_{ij} = \min(1, p_j/p_i)$, `Gibbs:` $a_{ij} = p_j/(p_i + p_j)$

  if $\delta = 1$ **then**

    x(n+1) = j

  else

    x(n+1) = x(n)

  n = n+1

until $n > max$