Prüfungsvorbereitungskurs Höhere Mathematik 3

Stochastik

Marco Boßle   Jörg Hörner

Mathematik–Online

Frühjahr 2011

Zusammenfassung

Wahrscheinlichkeitsraum (WR):

- Menge der Elementarereignisse Ω (Ergebnismenge)
- Ereignismenge $\mathcal{M} \subseteq P(\Omega)$ (σ-Algebra)
- Wahrscheinlichkeitsverteilung $p : \mathcal{M} \to [0, 1]$
  - Endlicher oder abzählbarer WR:
    Elementarwahrscheinlichkeiten $p(\{\omega\})$

    $$p(M) = \sum_{\omega \in M} p(\{\omega\}), \quad p(\Omega) = \sum_{\omega \in \Omega} p(\{\omega\}) = 1$$

  - $\Omega \subseteq \mathbb{R}^n$: Verteilungsfunktion F oder Dichte f

    $$F(x) = F(x_1, ..., x_n) = p((-\infty, x_1) \times \cdots \times (-\infty, x_n))$$

    $$F(x) = \int_{-\infty}^{x_1} \ldots \int_{-\infty}^{x_n} f(y)\, dy_n \cdots dy_1, \quad \int_{\mathbb{R}^n} f(y)\, dy = 1$$

Wahrscheinlichkeitsverteilungen:

- Laplaceverteilung $L(\Omega)$: $p(M) = \dfrac{\#(M)}{\#(\Omega)}$ (Gleichverteilung)
- Binomialverteilung $B(n,p)$:
  $\Omega = \{0,\dots,n\}\,,\; p(\{\ell\}) = \binom{n}{\ell} p^\ell (1-p)^{n-\ell}$
- Hypergeometrische Verteilung $H(n,m,k)$:
  $\Omega = \{0,\dots,k\}\,,\; p(\{\ell\}) = \dfrac{\binom{n}{\ell}\binom{m}{k-\ell}}{\binom{n+m}{k}}$
- Geometrische Verteilung $G(p) : \Omega = \mathbb{N}\,,\; p(\{\ell\}) = (1-p)^{\ell-1} p$
- Poissonverteilung $P(\lambda) : \Omega = \mathbb{N}_0\,,\; p(\{\ell\}) = e^{-\lambda}\dfrac{\lambda^\ell}{\ell!}$
- Gleichverteilung: $\Omega = [a,b]\,,\; p([c,d]) = (d-c)/(b-a)$
- Exponentialverteilung: $\Omega = \mathbb{R}_0^+\,,\; f(x) = \lambda e^{-\lambda x}$
- Normalverteilung: $N(\mu,\sigma^2)\,, \Omega = \mathbb{R}\,,\; f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)}$

Andere Verteilungen

- Urnenexperimente
  Kombinationsmöglichkeiten bei Auswahl von k aus n Elementen:

Reihenfolge	relevant	nicht relevant
mit Wiederholungen	n^k	$\binom{n+k-1}{k}$
ohne Wiederholungen	$n(n-1)\cdots(n-k+1)$	$\binom{n}{k}$

- Modellierung durch Zufallsvariablen

$$X : \Omega \to \tilde{\Omega}\,, \quad p_X(A) = p(\{\omega \in \Omega | X(\omega) \in A\})$$

Bedingte Wahrscheinlichkeit:

$$p(A|B) = \frac{p(A \cap B)}{p(B)}, \quad p(B|A) = \frac{p(A \cap B)}{p(A)}$$

Formel von der totalen Wahrscheinlichkeit:

- $A_1 \cup A_2 = \Omega\,,\ A_1 \cap A_2 = \emptyset:\quad p(B) = p(B|A_1)p(A_1) + p(B|A_2)p(A_2)$
- A_k: Zerlegung von Ω in disjunkte Teilmengen:

$$p(B) = \sum_k p(B|A_k)p(A_k)$$

Formel von Bayes:

$$p(A_1|B) = \frac{p(B|A_1)p(A_1)}{p(B|A_1)p(A_1) + p(B|A_2)p(A_2)}, \; p(A_\ell|B) = \frac{p(B|A_\ell)p(A_\ell)}{\sum_k p(B|A_k)p(A_k)}$$

Unabhängigkeit von n Ereignissen $A_1, \ldots, A_n$:

$$p\left(\bigcap_{k \in K} A_k\right) = \prod_{k \in K} p(A_k)\,, \quad \forall K \subseteq \{1, \ldots, n\}$$

Momente von reellwertigen Zufallsvariablen ($\tilde{\Omega} \subseteq \mathbb{R}$)

- Erwartungswert: (Erstes Moment)

$$E(X) = \sum_{x \in \tilde{\Omega}} x\, p(\{x\}), \quad E(X) = \int_{-\infty}^{\infty} x\, f(x)\, dx$$

- Varianz: (Zweites zentrales Moment)

$$D(X) = E((X - E(X))^2) = E(X^2) - (E(X))^2$$

- Rechenregeln:

$$E(\alpha X + \beta Y) = \alpha E(X) + \beta E(Y), \qquad D(\alpha X + \beta) = \alpha^2 D(X)$$

Erwartungswert und Varianz bei speziellen Verteilungen:

- Binomialverteilung $B(n, p) : E(X) = np\,,\; D(X) = np(1-p)$
- Geometrische Verteilung $G(p) : E(X) = 1/p\,,\; D(X) = (1-p)/p^2$
- Poissonverteilung $P(\lambda) : E(X) = \lambda\,,\; D(X) = \lambda$
- Gleichverteilung auf $[a, b] : E(X) = (a+b)/2\,,\; D(X) = (b-a)^2/12$
- Exponential-λ-Verteilung: $E(X) = 1/\lambda\,,\; D(X) = 1/\lambda^2$
- Normalverteilung $N(\mu, \sigma^2) :\; E(X) = \mu\,,\; D(X) = \sigma^2$

Zufallsvektoren $X = (X_1, \ldots, X_n)$

- Erwartungswerte: (erste Momente)

$$E(X_k) = \int\limits_{\mathbb{R}^n} x_k\, f(x_1, \ldots, x_n)\, dx$$

- Kovarianz: (Zweite zentrale Momente)

$$\operatorname{cov}(X_j, X_k) = E((X_j - E(X_j))(X_k - E(X_k)))$$

- Rechenregeln bei unabhängigen X_k:

$$\text{Verteilfunktion: } F_X(x) = F_{X_1}(x_1) \cdot \cdots \cdot F_{X_n}(x_n)$$

$$\text{Dichte: } f_X(x) = f_{X_1}(x_1) \cdot \cdots \cdot f_{X_n}(x_n)$$

$$E(X_j X_k) = E(X_j) E(X_k)\,, \quad D(X_j + X_k) = D(X_j) + D(X_k)$$

Approximation der $B(n,p)$ Verteilung:

- Seltenes Ereignis (kleines p)

  Poissonverteilung $P(\lambda)$ mit $\lambda = np$: $p(k) \approx e^{-np}\frac{(np)^k}{k!}$
- Häufige Wiederholung (großes n)
  Normalverteilung $N(\mu,\sigma^2)$ mit $\mu = np$, $\sigma^2 = np(1-p)$:

  $$p(\{k,\ldots,\ell\}) \approx p(a \leq (X-\mu)/\sigma < b) \approx \Phi(b) - \Phi(a)$$

  mit $a = (k - 1/2 - \mu)/\sigma$, $b = (\ell + 1/2 - \mu)/\sigma$

Parameterschätzung bei Beobachtungen $x_1, \ldots, x_n$:

- Erwartungswert: $E(X) \approx \bar{x} = \frac{1}{n}\sum_{k=1}^{n} x_k$
- Varianz: $D(X) \approx \tilde{s}^2 = \frac{1}{n-1}\sum_{k=1}^{n}(x_k - \bar{x})^2$
- $B(n,p)$-Verteilung: $p \approx \bar{x}/n$
- Poisson-Verteilung: $\lambda \approx \bar{x}$
- Exponential-Verteilung: $\lambda \approx 1/\bar{x}$
- $N(\mu,\sigma^2)$-Verteilung: $\mu \approx \bar{x}, \sigma^2 \approx \tilde{s}^2$
- Maximum-Likelihood-Schätzung:
  Wähle u so, dass $p(X = x, u)$ maximal ist.

Aufgabe 4.1 (I1438)

In einer Urne befinden sich drei rote und drei grüne Kugeln, in einer zweiten Urne befinden sich drei rote und drei blaue Kugeln.
Aus diesen Urnen wird (ohne zurücklegen) zweimal eine Kugel gezogen, indem zunächst zufällig eine Urne ausgewählt und dann dieser Urne zufällig eine Kugel entnommen wird.
Bestimmen Sie die Wahrscheinlichkeiten $p(A)$ und $p(B)$ folgender Ereignisse:
A: beide Kugeln sind rot.
B: beide Kugeln haben dieselbe Farbe.
Vergleichen Sie die Ergebnisse mit den Wahrscheinlichkeiten $p_2(A)$ und $p_2(B)$ bei einer Ziehung von zwei Kugeln ohne Zurücklegen aus einer Urne mit allen 12 Kugeln.

Lösung 4.1 (I1438)

Ereignisse:
A: beide Kugeln sind rot
B: beide Kugeln haben dieselbe Farbe
C_1: beide Kugeln aus gleicher Urne $\quad p(C_1) = 1/2$
C_2: Kugeln aus verschiedenen Urnen $\quad p(C_2) = 1/2$
Bedingte Wahrscheinlichkeiten:

$$\begin{aligned} p(A|C_1) &= \binom{3}{2} \Big/ \binom{6}{2} = \frac{3 \cdot 2}{6 \cdot 5} = 1/5 \\ p(A|C_2) &= \frac{3}{6}\frac{3}{6} = \frac{1}{4} \end{aligned}$$

$p(B|C_1)$: andere Farbe gleiche Fall wie rot $\quad p(B|C_1) = 2p(A|C_1) = 2/5$
$p(B|C_2)$: Kugeln müssen rot sein $p(B|C_2) = p(A|C_2) = 1/4$

Formel der totalen Wahrscheinlichkeit:

$$p(A) = p(A|C_1)p(C_1) + P(A|C_2)p(C_2) = \frac{1}{5}\,\frac{1}{2} + \frac{1}{4}\,\frac{1}{2} = \frac{9}{40}$$

$$p(B) = p(B|C_1)p(C_1) + P(B|C_2)p(C_2) = \frac{2}{5}\,\frac{1}{2} + \frac{1}{4}\,\frac{1}{2} = \frac{13}{40}$$

Alle Kugeln in einer Urne:

$$p_2(A) = \binom{6}{2} \Big/ \binom{12}{2} = \frac{6 \cdot 5}{12 \cdot 11} = 5/22$$

Ereignis D: beide Kugeln sind grün

$$p_2(D) = \binom{3}{2} \Big/ \binom{12}{2} = \frac{3 \cdot 2}{12 \cdot 11} = 1/22$$

Analog für blau , daher

$$p_2(B) = p_2(A) + 2p_2(D) = 7/22$$

Vergleich:

$$p(A) = 9/40 = 99/440 < 100/440 = 5/22 = p_2(A)$$

$$p(B) = 13/40 = 143/440 > 140/440 = 7/22 = p_2(A)$$

Aufgabe 4.4 (I1451)

Sei

$$f_\alpha(x) = \begin{cases} \alpha x_1 x_2 & \text{für } x = (x_1, x_2) \in [0,2]^2\,, \\ 0 & \text{sonst}\,, \end{cases}$$

und

$$A: \ \min(x_1, x_2) > 1\,, \quad B: (x_1 - 1)^2 + (x_2 - 1)^2 \leq 1\,.$$

Bestimmen Sie α so, dass $f_\alpha(x)$ die Dichte einer Zufallsvariablen ist, und bestimmen Sie für diese Dichte die Wahrscheinlichkeiten $p(A)$, $p(B)$ und $p(A \cap B)$.

Sind A und B stochastisch unabhängig?

Lösung 4.4 (I1451)

Bedingung für Wahrscheinlichkeitsdichte

$$f(x) \geq 0\,, \quad \int\limits_{\mathbb{R}^2} f(x)dx = 1$$

Erste Bedingung erfüllt für $\alpha > 0$.
Zweite Bedingung:

$$\int\limits_0^2 \int\limits_0^2 \alpha x_1 x_2 \, dx_1 \, dx_2 = \alpha \frac{1}{2}\, 4\, \frac{1}{2}\, 4 = 4\alpha \overset{!}{=} 1 \Rightarrow \alpha = 1/4\,.$$

Wahrscheinlichkeiten der Teilmengen:

$$p(A) = \int\limits_A f(x)\, dx = \int\limits_1^2 \int\limits_1^2 \frac{x_1 x_2}{4}\, dx_1\, dx_2 = \frac{1}{4} \frac{4-1}{2} \frac{4-1}{2} = 9/16 = .5625$$

Kreis mit (verschobenen) Polarkoordinaten

$$x_1 = 1 + r\cos(\varphi)\,, \quad x_2 = 1 + r\sin(\varphi)\,, \quad r \in [0,1]\,,\ \varphi \in [0,2\pi)\,,$$

$$\begin{aligned}
p(B) &= \frac{1}{4}\int_0^1\int_0^{2\pi}(1 + r\cos\varphi)(1 + r\sin\varphi)r\,d\varphi\,dr \\
&= \frac{1}{4}\int_0^1\int_0^{2\pi} r + r^2\cos\varphi + r^2\sin\varphi + r^3\sin\varphi\cos\varphi\,d\varphi\,dr \\
&= \frac{1}{4}\Big([r^2/2]_0^1[\varphi]_0^{2\pi} + [r^3/3]_0^1[\sin\varphi]_0^{2\pi} + [r^3/3]_0^1[-\cos\varphi]_0^{2\pi} \\
&\qquad + [r^4/4]_0^1[\tfrac{1}{2}\sin^2\varphi]_0^{2\pi}\Big) \\
&= \frac{1}{4}(\pi + 0 + 0 + 0) = \pi/4 \approx .7854
\end{aligned}$$

Schnittmenge Viertekreis $\rightarrow \varphi \in [0, \pi/2)$

$$p(A \cap B) \;=\; \frac{1}{4}\left(\pi/4 + 1/3 + 1/3 + 1/8\right) = \pi/16 + 19/96 \approx .3943$$

$p(A)p(B) = 9\pi/64 \neq \pi/16 + 19/96 \Rightarrow$ nicht unabhängig.

Aufgabe 4.6 (I1618)

Zur Analyse der Dauer von Arbeitslosigkeit wird der Zusammenhang zwischen Ausbildungsniveau und Dauer der Arbeitslosigkeit untersucht. Unter 152 Arbeitslosen ohne Ausbildung waren 98 kurzzeitig, 32 mittelfristig und 22 längerfristig arbeitslos.

a) Schätzen Sie jeweils die Wahrscheinlichkeiten dafür, dass ein Arbeitsloser ohne Ausbildung kurzfristig, mittelfristig oder langfristig arbeitslos ist und geben Sie für jede der Schätzungen ein (approximatives) 97%-Konfidenzintervall an.

b) Wieviel größer müßte die Stichprobe sein, um die Länge der Konfidenzintervalle zu vierteln?

Lösung 4.6 (I1618)

Schätzung von p der $B(n,p)$-Verteilung: $p = x/n$

$$\begin{aligned} p_k &= 98/152 = 49/76 \approx 0.6447\,, \\ p_m &= 32/152 = 4/19 \approx 0.2105\,, \\ p_l &= 22/152 = 11/76 \approx 0.1447\,. \end{aligned}$$

Berechnung der Intervalle:
Approximative-Normalverteilung: $\mu = np\,,\ \sigma^2 = p(1-p)n$
Gesucht: $I_p = [p - \delta/n, p + \delta/n]$ bzw. $I = [\mu - \delta, \mu + \delta]$ mit

$$\begin{aligned} 97\% &\overset{!}{=} P(X \in I) = P\left(|X - \mu| < \delta\right) = P\left(\left|\frac{X-\mu}{\sigma}\right| < \frac{\delta}{\sigma}\right) \\ &= \Phi\left(\frac{\delta}{\sigma}\right) - \Phi\left(-\frac{\delta}{\sigma}\right) = 2\Phi(\delta/\sigma) - 1 \Rightarrow \Phi(\delta/\sigma) = 1.97/2 = 0.985 \end{aligned}$$

Tabelle der Φ-Funktion $\to \delta/\sigma \approx 2.17 \Rightarrow \delta \approx 2.17\sqrt{p(1-p)n}$

Somit

$$
\begin{aligned}
\delta_k &\approx 2.17\sqrt{98/152 \cdot 54/152 \cdot 152} \approx 12.8 \\
&\Rightarrow p_k \in [(98-12.8)/152, (98+12.8)/152] = [.5605, 0.7290] \\
\delta_m &\approx 2.17\sqrt{32/152 \cdot 120/152 \cdot 152} \approx 10.907 \\
&\Rightarrow p_m \in [(32-10.907)/152, (32+10.907)/152] = [.1388, 0.2823] \\
\delta_l &\approx 2.17\sqrt{22/152 \cdot 130/152 \cdot 152} \approx 9.413 \\
&\Rightarrow p_l \in [(22-9.413)/152, (22+9.413)/152] = [.0828, 0.2067]\,.
\end{aligned}
$$

Länge der Intervalle: $2\delta/n = 2 \cdot 2.17\sqrt{p(1-p)/n} \Rightarrow \tilde{n} = 16n$.