LÖSUNG

Diskrete Wahrscheinlichkeitstheorie – Endterm

Beachten Sie: **Soweit nicht anders angegeben, sind alle Antworten zu begründen!**

Aufgabe 1 — je 2P = 10P

(a) Seien X und Y unabhängig und Bernoulli-verteilt mit Erfolgsw'keit p, d.h., $X \sim \text{Bin}(1;p)$ und $Y \sim \text{Bin}(1;p)$.

Geben Sie die Dichte von $X - Y$ an. (Keine Begründung verlangt.)

(b) Seien A und B Ereignisse mit $\Pr[A \mid B] = \Pr[B \mid A]$ und $\Pr[A \cup B] = 1$ und $\Pr[A \cap B] > 0$.

Zeigen Sie: $\Pr[A] > \frac{1}{2}$.
Hinweis: Zeigen Sie zunächst, dass $\Pr[A] = \Pr[B]$.

(c) Seien X und Y unabhängige Zufallsvariablen mit $\mathbb{E}[X] = 1$, $\text{Var}[X] = 4$, $\mathbb{E}[Y] = 3$ und $\text{Var}[Y] = 8$.

Bestimmen Sie $\mathbb{E}[(X+Y)^2]$ und $\mathbb{E}[(X-Y)^2]$.

(d) Es gelte $X \sim \text{Bin}(n;p)$.

Geben Sie Werte für n und p an, so dass die Standardabweichung von X echt größer dem Erwartungswert von X ist.

(e) Es seien X und Y unabhängig und geometrisch verteilt mit Erfolgsw'keit p.

Bestimmen Sie die Dichte der Zufallsvariablen $\min\{X, Y\}$.

Lösungsvorschlag:

(a) $\Pr[X - Y = 1] = \Pr[X - Y = -1] = p(1-p)$
$\Pr[X - Y = 0] = p^2 + (1-p)^2$
$\Pr[X - Y = z] = 0$ für $z \notin \{-1, 0, 1\}$.

(b) Nach Definition von bedingter W'keit gilt $\Pr[A \mid B] = \frac{\Pr[A\cap B]}{\Pr[B]}$ und $\Pr[B \mid A] = \frac{\Pr[A\cap B]}{\Pr[A]}$.
Wegen $\Pr[A \mid B] = \Pr[B \mid A]$ und $\Pr[A \cap B] > 0$ folgt $\Pr[A] = \Pr[B]$.
Mit der Siebformel gilt $1 = \Pr[A \cup B] = \Pr[A] + \Pr[B] - \Pr[A \cap B] = 2\Pr[A] - \Pr[A \cap B]$, also $\Pr[A] = \frac{1+\Pr[A\cap B]}{2} > \frac{1}{2}$.

(c)

$$
\begin{aligned}
\text{Var}[X+Y] &= \text{Var}[X] + \text{Var}[Y] && \text{da unabh.} \\
&= 4 + 8 = 12 \\
\text{Var}[X+Y] &= \mathbb{E}[(X+Y)^2] - \mathbb{E}[X+Y]^2 \\
&= \mathbb{E}[(X+Y)^2] - (\mathbb{E}[X] + \mathbb{E}[Y])^2 && \text{Linearität} \\
&= \mathbb{E}[(X+Y)^2] - 16 \\
\mathbb{E}[(X+Y)^2] &= \text{Var}[X+Y] + (\mathbb{E}[X] + \mathbb{E}[Y])^2 = 12 + 16 = 28 \\
\mathbb{E}[(X-Y)^2] &= \mathbb{E}[X^2 - 2XY + Y^2] \\
&= \mathbb{E}[X^2] - 2\mathbb{E}[X]\mathbb{E}[Y] + \mathbb{E}[Y^2] && \text{Linearität} \\
&= \mathbb{E}[X^2] + 2\mathbb{E}[X]\mathbb{E}[Y] + \mathbb{E}[Y^2] - 4\mathbb{E}[X]\mathbb{E}[Y] \\
&= \mathbb{E}[(X+Y)^2] - 12 && \text{Linearität und Def. Var} \\
&= 16
\end{aligned}
$$

(d)
$$\sqrt{np(1-p)} > np \Leftrightarrow np(1-p) > n^2p^2 \Leftrightarrow 1-p > np \Leftrightarrow p < \frac{1}{n+1}.$$
Man beachte, dass $n > 0, p > 0$ gelten muss.

(e)
$$\Pr[\min\{X,Y\} > k] = \Pr[X > k, Y > k] = \Pr[X > k]^2 = (1-p)^{2k}.$$
Also
$$\Pr[\min\{X,Y\} = k] = \Pr[\min\{X,Y\} > k-1] - \Pr[\min\{X,Y\} > k] = (1-p)^{2k-2} - (1-p)^{2k}.$$
Folglich gilt $X \sim \mathrm{Geo}(1-(1-p)^2)$ (nicht verlangt).

Aufgabe 2 **1P+3P+3P=7P**

Bei einer Geburt ist das Kind mit W'keit 1/2 ein Mädchen und mit W'keit 1/2 ein Junge. Wir interessieren uns für die W'keit, dass bei n Geburten mindestens doppelt so viele Mädchen wie Jungen geboren werden. Wenn X die Zahl der geborenen Mädchen bezeichnet, dann interessieren wir uns also für $\Pr\left[X \geq \frac{2}{3}n\right]$.

(a) Geben Sie den Namen der Verteilung von X an, sowie $\mathbb{E}X$ und $\mathrm{Var}X$.

(b) Zeigen Sie mit der Chebyshev-Ungleichung:
$$\Pr\left[X \geq \frac{2}{3}n\right] \leq \frac{9}{2n} \text{ für alle } n \geq 1$$
Hinweis: Die W'keit, dass mindestens doppelt so viele Mädchen wie Jungen geboren werden, ist gleich groß wie die W'keit, dass mindestens doppelt so viele Jungen wie Mädchen geboren werden.

(c) Sie dürfen jetzt X mit der Normalverteilung (zentraler Grenzwertsatz) approximieren. Bestimmen Sie n, sodass gilt:
$$\Pr\left[X \geq \frac{2}{3}n\right] \approx 0.001$$
Hinweis: $\Phi(3) \approx 0.999$

Lösungsvorschlag:

(a) Es gilt $X \sim \mathrm{Bin}(n; \frac{1}{2})$. Folglich $\mathbb{E}X = \frac{n}{2}$ und $\mathrm{Var}X = \frac{n}{4}$.

(b)
$$\begin{aligned}
\Pr\left[X \geq \frac{2}{3}n\right] &= \frac{1}{2} \cdot 2 \cdot \Pr\left[X \geq \frac{2}{3}n\right] \\
&= \frac{1}{2} \cdot \left(\Pr\left[X \geq \frac{2}{3}n\right] + \Pr\left[X \leq \frac{1}{3}n\right]\right) \\
&= \frac{1}{2} \cdot \left(\Pr\left[X \geq \frac{2}{3}n \;\text{ oder }\; X \leq \frac{1}{3}n\right]\right) \\
&= \frac{1}{2}\Pr\left[\left|X - \frac{n}{2}\right| \geq \frac{n}{6}\right] \\
&\leq \frac{1}{2} \cdot \frac{n/4}{(n/6)^2} = \frac{9}{2n} && \text{(Chebyshev)}
\end{aligned}$$

(c) Mit dem ZGWS gilt:
$$\Pr\left[X \geq \frac{2}{3}n\right] = \Pr\left[\frac{X - \frac{n}{2}}{\sqrt{n/4}} \geq \frac{n/6}{\sqrt{n/4}}\right] \approx 1 - \Phi(\frac{1}{3}\sqrt{n})$$
Gleichsetzen mit 0.001 liefert zusammen mit dem Hinweis: $\frac{1}{3}\sqrt{n} \approx 3$. Folglich wähle $n \approx 81$.

Aufgabe 3 **2P+2P+4P=8P**

Es seien X und Y unabhängig und jeweils exponential-verteilt mit Parameter λ.

(a) Bestimmen Sie die Dichte der Zufallsvariable $aY + b$ für beliebige Konstanten $a, b \in \mathbb{R}$, wobei $a < 0$ gelte.

(b) Berechnen Sie den Erwartungswert und die Varianz der Zufallsvariablen $X - Y$.

(c) Leiten Sie die Dichte von $X - Y$ her.

Lösungsvorschlag:

(a)

$$\Pr[aY+b \le t] = \Pr\left[Y \ge \frac{t-b}{a}\right] = 1 - F_Y(\frac{t-b}{a}).$$

Damit folgt für die Dichte:

$$f_{aY+b}(t) = -\frac{1}{a} f_Y(\frac{t-b}{a}) = \frac{\lambda}{-a} e^{\lambda \frac{t-b}{-a}} I_{(-\infty,b]}(t).$$

(b)

$$\mathbb{E}[X-Y] = \mathbb{E}[X] - \mathbb{E}[Y] = \frac{1}{\lambda} - \frac{1}{\lambda} = 0$$

$$\mathrm{Var}[X-Y] \overset{\text{unabh.}}{=} \mathrm{Var}[X] + (-1)^2 \mathrm{Var}[Y] = \frac{1}{\lambda^2} + \frac{1}{\lambda^2} = \frac{2}{\lambda^2}$$

(c) *Berechnung mit Hilfe der Faltung*:

Dichte von $-Y$ nach (a) mit $a = -1$ und $b = 0$:

$$f_{-Y}(t) = f_Y(-t) = \lambda \cdot e^{\lambda \cdot t} \cdot I_{(-\infty,0]}(t).$$

Damit

$$\begin{aligned}
f_{X-Y}(t) &= \int_{z=-\infty}^{\infty} f_X(z) f_{-Y}(t-z) dz \\
&= \int_{z=-\infty}^{\infty} f_X(z) f_Y(z-t) dz \\
&= \int_{z=\max(0,t)}^{\infty} \lambda e^{-\lambda z} \cdot \lambda e^{-\lambda(z-t)} dz \\
&= \int_{z=\max(0,t)}^{\infty} \lambda^2 e^{-2\lambda z} \cdot e^{\lambda t} dz \\
&= -\frac{\lambda}{2} e^{\lambda t} \int_{z=\max(0,t)}^{\infty} -2\lambda e^{-2\lambda z} dz \\
&= -\frac{\lambda}{2} e^{\lambda t} [e^{-2\lambda z}]_{z=\max(0,t)}^{\infty} \\
&= \frac{\lambda}{2} e^{\lambda t} e^{-2\lambda \max(0,t)} \\
&= \begin{cases} \frac{\lambda}{2} e^{-\lambda t} & \text{falls } t \ge 0 \\ \frac{\lambda}{2} e^{\lambda t} & \text{falls } t < 0 \end{cases}
\end{aligned}$$

Berechnung mittels der Verteilungsfunktion von $X - Y$:

$X, Y \sim \exp(\lambda)$. Dichte von $X - Y$:

$$\begin{aligned}
\Pr[X - Y \le t] &= \Pr[X \le t + Y] \\
&= \int_{y=-\infty}^{\infty} f_Y(y) F_X(t+y) dy
\end{aligned}$$

Falls $t + y < 0$, gilt $F_X(t+y) = 0$:

$$\begin{aligned}
&= \int_{y=-t}^{\infty} f_Y(y) \cdot (1 - e^{-\lambda \cdot (t+y)}) dy \\
&= \int_{y=-t}^{\infty} f_Y(Y) dy - e^{-\lambda \cdot t} \cdot \int_{y=-t}^{\infty} f_Y(y) \cdot e^{-\lambda y} dy \\
&= (1 - F_Y(-t)) - e^{-\lambda \cdot t} \int_{y=\max(0,-t)}^{\infty} \lambda \cdot e^{-2\lambda y} dy \\
&= (1 - F_Y(-t)) + e^{-\lambda \cdot t} \frac{1}{2} \int_{y=\max(0,-t)}^{\infty} -2\lambda e^{-2\lambda y} dy \\
&= (1 - F_Y(-t)) + e^{-\lambda \cdot t} \frac{1}{2} [e^{-2\lambda y}]_{y=\max(0,-t)}^{y=\infty} \\
&= (1 - F_Y(-t)) - \frac{1}{2} \cdot e^{-\lambda \cdot t} \cdot e^{-2\lambda \max(0,-t)}.
\end{aligned}$$

Also:

$$\mathbf{F}_{X-Y}(t) = \begin{cases} \lambda e^{\lambda t} - \frac{1}{2} \cdot e^{\lambda \cdot t} = \frac{1}{2} \cdot e^{\lambda \cdot t} & \text{falls } t < 0 \\ 1 - \frac{1}{2} \cdot e^{-\lambda \cdot t} & \text{falls } t \geq 0. \end{cases}$$

Für die Dichte folgt somit:

$$f_{X-Y}(t) = \begin{cases} \frac{\lambda}{2} \cdot e^{\lambda \cdot t} & \text{falls } t < 0 \\ \frac{\lambda}{2} \cdot e^{-\lambda \cdot t} & \text{falls } t \geq 0. \end{cases}$$

Aufgabe 4

1P+2P+3P=6P

Es seien $X_1, \ldots, X_n$ unabhängige Zufallsvariablen. Alle haben die Dichte

$$f(x) = \begin{cases} \lambda^2 x e^{-\lambda x} & x > 0 \\ 0 & x < 0 \end{cases} \qquad \text{wobei } \lambda > 0.$$

(a) Es sei $\vec{k} = (k_1, \ldots, k_n)$ ein Stichprobenvektor.

Geben Sie die Likelihood-Funktion $L(\vec{k}, \lambda)$ an.

(b) Bestimmen Sie mit Hilfe von $L(\vec{k}, \lambda)$ den ML-Schätzer für den Parameter λ.

Hinweise:

- Verwenden Sie $\ln L(\vec{k}, \lambda)$.
- Die Betrachtung der zweiten Ableitung wird nicht verlangt.

(c) Zeigen Sie, dass der Schätzer aus (b) für $n = 1$ *nicht* erwartungstreu ist.

Lösungsvorschlag:

(a) Es gilt

$$L(\vec{k}, \lambda) = \prod_{i=1}^{n} f(k_i) = \prod_{i=1}^{n} \lambda^2 k_i e^{-\lambda k_i} = \lambda^{2n} \cdot \prod_{i=1}^{n} k_i \cdot \prod_{i=1}^{n} e^{-\lambda k_i}\,.$$

(b) Es gilt

$$\ln L(\vec{k}, \lambda) = 2n \ln \lambda + \sum_{i=1}^{n} \ln k_i - \lambda \cdot \sum_{i=1}^{n} k_i\,.$$

Um L zu maximieren, genügt es $\ln L$ zu maximieren. Ableiten nach λ führt zu

$$2n \cdot \frac{1}{\lambda} - \sum_{i=1}^{n} k_i\,.$$

Gleich 0 setzen führt zu

$$\lambda = \frac{2n}{\sum_{i=1}^{n} k_i}\,.$$

Der gesuchte Schätzwert ist also $\dfrac{2n}{\sum_{i=1}^{n} k_i}$.

Damit ergibt sich der Schätzer $\frac{2n}{\sum_{i=1}^{n} X_i}$.

(c) Es ist zu zeigen, dass $\mathbb{E}[\frac{2}{X_1}] \neq \lambda$ gilt.

$$\begin{aligned} \mathbb{E}[\frac{2}{X_1}] &= \int_{-\infty}^{\infty} \frac{2}{x} f(x) dx \\ &= \int_{0}^{\infty} \frac{2}{x} \lambda^2 x e^{-\lambda x} dx \\ &= \int_{0}^{\infty} 2\lambda^2 e^{-\lambda x} dx \\ &= -2\lambda \int_{0}^{\infty} -\lambda e^{-\lambda x} dx \\ &= -2\lambda [e^{-\lambda x}]_0^{\infty} \\ &= 2\lambda. \end{aligned}$$

Da $\lambda > 0$ vorausgesetzt ist, folgt $\mathbb{E}[\frac{2}{X_1}] \neq \lambda$.

Aufgabe 5 **2P+2P+3P+3P+2P=12P**

Sie empfangen Mails von Server A und von Server B.

- Mails von Server A kommen in einem Poisson-Prozess mit Rate $\frac{1}{2}$ (Mails pro Stunde).
  Das bedeutet, die Wartezeiten $T_1, T_2, \ldots$ zwischen zwei A-Mails sind exponential-verteilt mit Parameter $\frac{1}{2}$.
- Mails von Server B kommen in einem Poisson-Prozess mit Rate 1.
  Das bedeutet, die Wartezeiten $U_1, U_2, \ldots$ zwischen zwei B-Mails sind exponential-verteilt mit Parameter 1.

Die Wartezeiten $T_1, U_1, T_2, U_2, \ldots$ sind unabhängig.

(a) Sei $M = \min\{T_1, U_1\}$. Berechnen Sie $\mathbb{E}M$.

(b) Sei N_A die Zahl der A-Mails, die in den ersten zwei Stunden kommen. Berechnen Sie $\Pr[N_A = 0]$.

(c) Nehmen wir (nur für diese Teilaufgabe) an, in den ersten zwei Stunden kommt genau 1 Mail. Wie groß ist, unter dieser Bedingung, die bedingte W'keit, dass es eine A-Mail ist?

(d) Zeigen Sie: $\Pr[T_1 < U_1] = \frac{1}{3}$.

(e) Wie groß ist die W'keit, dass die ersten drei Mails alle A-Mails sind?

Lösungsvorschlag:

(a) Laut Vorlesung ist M exponential-verteilt mit Rate $\frac{1}{2} + 1 = \frac{3}{2}$. Es gilt $\mathbb{E}M = \frac{2}{3}$, also 40 Minuten.

(b) Die Zufallsvariable N_A ist Poisson-verteilt mit Parameter $\frac{1}{2} \cdot 2 = 1$. Es folgt

$$\Pr[N_A = 0] = e^{-1} \cdot \frac{1^0}{0!} = e^{-1}\,.$$

(c) Seien N_A und N_B die Zahl der A-Mails bzw. der B-Mails in den ersten zwei Stunden.
Es gilt $N_A \sim \text{Poi}(1)$ und $N_B \sim \text{Poi}(2)$. Gesucht ist $\Pr[N_A = 1 \mid N_A + N_B = 1]$. Es gilt

$$\begin{aligned}
\Pr[N_A = 1 \mid N_A + N_B = 1] &= \frac{\Pr[N_A = 1, N_B = 0]}{\Pr[N_A = 1, N_B = 0] + \Pr[N_A = 0, N_B = 1]} \\
&= \frac{\Pr[N_A = 1] \cdot \Pr[N_B = 0]}{\Pr[N_A = 1] \cdot \Pr[N_B = 0] + \Pr[N_A = 0] \cdot \Pr[N_B = 1]} \\
&= \frac{e^{-1} \cdot e^{-2}}{e^{-1} \cdot e^{-2} + e^{-1} \cdot 2e^{-2}} \\
&= \frac{1}{3}
\end{aligned}$$

(d) Seien f_T und f_U die Dichten von T_1 bzw. U_1. Sei F_T die Verteilung von T_1. Dann gilt:

$$f_T(t) = \begin{cases} \frac{1}{2}e^{-t/2} & t \geq 0 \\ 0 & t < 0 \end{cases} \qquad f_U(u) = \begin{cases} e^{-u} & u \geq 0 \\ 0 & u < 0 \end{cases} \qquad F_T(t) = \begin{cases} 1 - e^{-t/2} & \geq 0 \\ 0 & t \leq 0 \end{cases}$$

Es folgt

$$\begin{aligned}
\Pr[T_1 \leq U_1] &= \int_{t \leq u} f_U(u) f_T(t)\ dt\ du \\
&= \int_{u=-\infty}^{\infty} f_U(u) \underbrace{\int_{t=-\infty}^{u} f_T(t)\ dt}_{F_T(u)}\ du = \int_0^\infty e^{-u} \cdot (1 - e^{-u/2})\ du \\
&= \int_0^\infty \left(e^{-u} - e^{-\frac{3}{2}u}\right)\ du = \left[-e^{-u} + \frac{2}{3}e^{-\frac{3}{2}u}\right]_0^\infty = 0 + 0 + 1 - \frac{2}{3} = \frac{1}{3}
\end{aligned}$$

(e) Aus der Gedächtnislosigkeit und der vorherigen Teilaufgabe folgt: Die gesuchte W'keit ist $\frac{1}{3} \cdot \frac{1}{3} \cdot \frac{1}{3} = \frac{1}{27}$.

Aufgabe 6 **3P+2P+2P=7P**

Gegeben sei ein Zufallszahlengenerator, der Zufallszahlen $X_1, X_2, \ldots$ erzeugt. Die Zufallszahlen X_i sind unabhängig und stetig gleichverteilt auf einem Intervall $[0, a]$, wobei $a > 0$.

(a) Sei $T := \max\{X_1, X_2, X_3\}$. Berechnen Sie Verteilung $F_T(t)$, Dichte $f_T(t)$ und Erwartungswert von T (jeweils in Abhängigkeit von a).

(b) Wir entwickeln einen Test für den Parameter a. Es soll getestet werden, ob $a \leq 10$ oder $a \geq 10$ gilt. Der Parameter a soll nicht fälschlicherweise zu klein geschätzt werden und die Fehlerw'keit 1. Art soll auf 0.01 begrenzt werden. Wie muss die Nullhypothese H_0 gewählt werden und warum?

(c) Als Testgröße wird $T := \max\{X_1, X_2, X_3\}$ gewählt. Berechnen Sie einen Ablehnungsbereich, sodass die Fehlerw'keit 1. Art 0.01 beträgt.

Lösungsvorschlag:

(a) Seien F_X und f_X Verteilung und Dichte einer auf $[0, a]$ stetig gleichverteilten Zufallsvariable. Für $0 \leq t \leq a$ gilt:

$$\begin{aligned}
F_T(t) &= \Pr[T \leq t] = \Pr[\max\{X_1, X_2, X_3\} \leq t] = \Pr[X_1 \leq t, X_2 \leq t, X_3 \leq t] = \Pr[X_1 \leq t] \cdot \Pr[X_2 \leq t] \cdot \Pr[X_3 \leq t] \\
&= F_X(t)^3 = \left(\frac{t}{a}\right)^3 = \frac{t^3}{a^3} \\
f_T(t) &= F_T'(t) = \frac{3t^2}{a^3} \\
\mathbb{E}T &= \int_{-\infty}^{\infty} t f_T(t)\, dt = \int_0^a \frac{3}{a^3} t^3\, dt = \frac{3}{4a^3}\left[t^4\right]_0^a = \frac{3}{4}a
\end{aligned}$$

(b) Die Fehlerw'keit 1. Art ist begrenzt, das ist die W'keit, dass H_0 fälschlicherweise abgelehnt wird. Gleichzeitig soll a nicht fälschlicherweise zu klein geschätzt werden. Folglich ist H_0 die Hypothese, dass $a \geq 10$ (oder $a > 10$) ist.

(c) Bei der gewählten Nullhypothese ist als Ablehnungsbereich nur ein Intervall $K = [0, k]$ sinnvoll. Für die Fehlerw'keit 1. Art gilt:

$$\sup_{a \in H_0} \Pr_a[T \in K] = \sup_{a \geq 10} \Pr_a[T \leq k] = \Pr_{a=10}[T \leq k] = F_T(k) = \frac{k^3}{1000}$$

Gleichsetzen mit 0.01 liefert $k = \sqrt[3]{10} \approx 2.15$.