Wahrscheinlichkeit und Statistik - Lösung
D-INFK

---

1. (10 Punkte)

Bei den folgenden 10 Fragen gibt es pro richtig beantwortete Frage 1 Punkt. Pro falsche Antwort gibt es 1/2 Punkt Abzug. Minimal erhält man für die gesamte Aufgabe 0 Punkte.

a) Eine faire Münze wird zweimal geworfen. Für $i = 1, 2$ sei K_i das Ereignis, dass beim i-ten Wurf Kopf erscheint. Dann gilt:

1. K_1 und K_2 sind unabhängig.
2. K_1 und K_2 sind disjunkt.
3. K_1 und K_2 sind unabhängig und $K_1 \cup K_2 = \Omega$.

b) Seien A und B zwei Ereignisse mit $P[A] > 0$ und $A \cap B = \emptyset$. Die Aussage $P[B|A] = 0$ ist

1. wahr.
2. nicht wahr.

c) Wir untersuchen eine Bodenprobe auf zwei Schadstoffe X und Y. Wir definieren die Ereignisse A: "Bodenprobe enthält X" und B: "Bodenprobe enthält Y". Es sei $P[A] = 0.1$ und $P[B] = 0.3$. Zusätzlich wissen wir, dass die Wahrscheinlichkeit, dass die Bodenprobe keinen der beiden Schadstoffe enthält, 0.7 beträgt. Dann gilt:

1. $P[A \cap B] = 0.03$.
2. $P[A \cap B] = 0.1$.
3. Es gibt zu wenig Information, um $P[A \cap B]$ zu berechnen.

Siehe nächste Seite!

d) Es sei X eine Zufallsvariable mit $E[X] = 2$, $\text{Var}(X) = 3$. Dann gilt:

1. $E[X^2] = 4$.
2. $E[X^2] = 7$.
3. Es gibt zu wenig Information, um $E[X^2]$ zu berechnen.

e) Wie oft muss man einen fairen Würfel im Schnitt werfen, bis man zum ersten Mal eine 3 erhält?

1. 6 mal.
2. 6^6 mal.
3. $6!$ mal.

f) Sei X eine Zufallsvariable mit Verteilungsfunktion

$$F_X(x) = \begin{cases} 1 - e^{-cx}, & x \geq 0, \\ 0, & x < 0, \end{cases}$$

für ein $c > 0$. Der Median (d.h. $F_X^{-1}(\frac{1}{2})$) dieser Verteilung beträgt

1. $\frac{1}{2}$
2. $\frac{1}{c} \ln 2$
3. $1 - e^{-\frac{1}{2}c}$

g) Angenommen, 10% der Bevölkerung sind Raucher. Die Zufallsvariable X sei die Anzahl der Raucher unter 100 zufällig ausgewählten Personen. Wie gross ist die Standardabweichung von X?

1. 3
2. $\sqrt{10}$
3. 10

h) Seien X und Y unabhängig und lognormalverteilt (eine Zufallsvariable Z ist lognormalverteilt, falls $\ln Z$ normalverteilt ist). Welche Aussage ist korrekt?

1) XY ist lognormalverteilt.
2) XY ist normalverteilt.
3) e^{X+Y} ist normalverteilt.

Siehe nächste Seite!

i) Die stetige Zufallsvariable X habe die kumulative Verteilungsfunktion

$$F(x) = 1 - e^{-2x},\ x \geq 0.$$

Welche der folgenden Aussagen ist richtig?

1. Die Dichte von X ist $f(x) = 2xe^{-2x},\ x \geq 0$.
2. Es ist $P[X \geq 0.5] = e^{-0.5}$.
3. Es ist $P[X = 1] = 1 - e^{-2}$.
4. Für $Y = 2X$ gilt $P[Y \leq 2] = 1 - e^{-2}$.

j) Bei der Berechnung des realisierten p-Wertes eines Tests

1. muss man wissen, ob der Test einseitig oder zweiseitig ist.
2. braucht man die konkreten Daten nicht.
3. muss man das Signifikanzniveau α kennen.

Lösung:

a) 1. Punkt 2. und 3. stimmen nicht, da bei beiden Versuchen Kopf auftreten kann bzw. Kopf nicht auftreten muss.

b) 1. $P[B|A] = P[A \cap B]/P[A] = 0/P[A] = 0$.

c) 2. Es ist gegeben, dass $P[A^c \cap B^c] = 0.7$. Also $P[A \cup B] = 1 - P[(A \cup B)^c] = 1 - P[A^c \cap B^c] = 1 - 0.7 = 0.3$. Weiterhin gilt für beliebige Ereignisse A, B immer $P[A \cup B] = P[A] + P[B] - P[A \cap B]$. Also $0.3 = 0.1 + 0.3 - P[A \cap B]$. Also folgt $P[A \cap B] = 0.1$.

d) 2. $\mathrm{Var}(X) = E[X^2] - (E[X])^2$. Also

$$E[X^2] = \mathrm{Var}(X) + (E[X])^2 = 3 + 2^2 = 7.$$

e) 1. Sei X die Anzahl der Würfe, bis zum ersten Mal eine 3 erscheint. Dann ist $X \sim \mathrm{Geom}(1/6)$. Also ist $E[X] = 1/(1/6) = 6$.

f) 2. Wegen $F_X(x) = 1 - \exp(-cx) = z \iff x = -\ln(1-z)/c$ gilt, dass $F_X^{-1}(z) = -\ln(1-z)/c$. Somit folgt $F_X^{-1}(0.5) = \ln(2)/c$.

g) 1. Es gilt $X \sim \mathrm{Bin}(100, 0.1)$. Daher ergibt sich

$$\mathrm{Var}(X) = 100 \cdot 0.1 \cdot 0.9 = 9.$$

Somit ist $\mathrm{sd}(X) = \sqrt{\mathrm{Var}(X)} = 3$.

h) 1. $\ln(XY) = \ln X + \ln Y$ ist als Summe der unabhängigen normalverteilten Zufallsvariablen $\ln X$ und $\ln Y$ normalverteilt.

Siehe nächste Seite!

i) 4.

1. Falsch. Die Dichte ist $f(x) = 2e^{-2x},\ x \geq 0$.
2. Falsch. $P[X \geq 0.5] = e^{-2 \cdot 0.5} = e^{-1}$.
3. Falsch. Für stetige Verteilungen gilt immer $P[X = x] = 0$ für alle x.
4. Richtig. $P[Y \leq 2] = P[2X \leq 2] = P[X \leq 1] = 1 - e^{-2}$.

j) 1. Man muss wissen, was es bedeutet, dass die Teststatistik T einen extremeren Wert als den beobachteten Wert $t = T(\omega)$ annimmt. Bei zweiseitigen Tests betrachtet man demnach das Ereignis $\{|T| < t\}$ (bzw. $\{|T| > t\}$), während für einseitige Tests das Ereignis $\{T < t\}$ (bzw $\{T > t\}$) für die Berechnung des p-Wertes herangezogen wird.

Siehe nächste Seite!

2. (8 Punkte)

Eine Zufallsvariable Y heisst χ^2-verteilt mit Freiheitsgrad $\nu \in \mathbb{N}$ (geschrieben $Y \sim \chi^2_\nu$), falls

$$Y = \sum_{k=1}^{\nu} Z_k^2,$$

wobei $Z_1, \ldots, Z_\nu$ i.i.d. $\mathcal{N}(0,1)$-verteilt sind.

a) **(3 Punkte)** Zeige, dass

$$E[Y] = \nu \qquad \text{und} \qquad \mathrm{Var}[Y] = 2\nu$$

gilt.

b) **(1.5 Punkte)** Gib mit Hilfe der Chebyshev-Ungleichung eine untere Schranke für die Wahrscheinlichkeit

$$\mathbb{P}\left[\left|\frac{Y}{\nu} - 1\right| \le \frac{3}{4}\right]$$

an. Wie lautet der konkrete Wert der Schranke für $\nu = 12$?

c) **(3.5 Punkte)** Leite eine Annäherung für die obige Wahrscheinlichkeit mit Hilfe des zentralen Grenzwertsatzes her. Wie lautet der konkrete Wert der Approximation für $\nu = 12$?

Lösung:

a) Für ganzzahlige Werte von ν kann Y geschrieben werden als $Y = \sum_{k=1}^{\nu} Z_k^2$ mit $Z_1, \ldots, Z_\nu$ i.i.d. $\mathcal{N}(0,1)$.

Daraus folgt einerseits

$$E[Y] = \sum_{k=1}^{\nu} E[Z_k^2] = \nu \cdot 1 = \nu. \qquad \text{(1P)}$$

Andererseits ist wegen der Unabhängigkeit und identischen Verteilung der Z_k

$$E[Y^2] = E[(Z_1^2 + Z_2^2 + \cdots + Z_\nu^2)^2] = \nu E[Z_1^4] + \nu(\nu-1)(E[Z_1^2])^2, \qquad \text{(0.5P)}$$

oder alternativ

$$\mathrm{Var}[Y] = \mathrm{Var}\left[\sum_{k=1}^{\nu} Z_k^2\right] = \sum_{k=1}^{\nu} \mathrm{Var}\left[Z_1^2\right] = \nu\left(E[Z_1^4] - E[Z_1^2]^2\right). \qquad \text{(0.5P)}$$

Mit $E[Z_1^2] = 1$ und

Siehe nächste Seite!

$$E[Z_1^4] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} Y x^4 e^{-x^2/2}\, dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} Y x^3 x e^{-x^2/2}\, dx$$
$$= \frac{1}{\sqrt{2\pi}} \left(-x^3 e^{-x^2/2}\right)\Big|_{-\infty}^{\infty} + \frac{3}{\sqrt{2\pi}} \int_{-\infty}^{\infty} Y x^2 e^{-x^2/2}\, dx = 3 \qquad \text{(1P)}$$

erhalten wir schliesslich $E[Y^2] = 3\nu + \nu(\nu-1) = \nu^2 + 2\nu$, und daraus

$$\text{Var}[Y] = (\nu^2 + 2\nu) - \nu^2 = 2\nu, \qquad \text{(0.5P)}$$

beziehungsweise

$$\text{Var}[Y] = \nu(3-1) = 2\nu. \qquad \text{(0.5P)}$$

b) Die Chebyshev-Ungleichung lautet $P[|Y-\nu| > c] \leq \text{Var}[Y]/c^2 = \frac{2\nu}{c^2}$. Somit folgt

$$P\left[\left|\frac{Y}{\nu} - 1\right| \leq \frac{3}{4}\right] = P\left[\left|\frac{Y-\nu}{\nu}\right| \leq \frac{3}{4}\right] = 1 - P\left[|Y-\nu| > \frac{3\nu}{4}\right]$$
$$\geq 1 - \frac{2\nu}{9\nu^2/16} = 1 - \frac{32}{9\nu}. \qquad \text{(1P)}$$

Für $\nu = 12$ erhalten wir

$$P\left[\left|\frac{Y}{\nu} - 1\right| \leq \frac{3}{4}\right] = \frac{19}{27} (\approx 0.7037). \qquad \text{(0.5P)}$$

c) Die $Y_i = Z_i^2$ sind i.i.d $\sim \chi_1^2$ (0.5P); also ist $E[Y_i] = 1$ und $\text{Var}[Y_i] = 2$, $i = 1, \ldots, \nu$. (0.5P)
Der zentrale Grenzwertsatz für $Y = \sum_{i=1}^{\nu} Y_i$ liefert

$$Z = \frac{Y - \nu E[Y_i]}{\sqrt{\nu \text{Var}[Y_i]}} = \frac{Y-\nu}{\sqrt{2\nu}} \overset{\text{approx.}}{\sim} \mathcal{N}(0,1). \qquad \text{(05.P)}$$

Eine Approximation der Wahrscheinlichkeit ergibt sich daher durch

$$P\left[\left|\frac{Y}{\nu} - 1\right| \leq \frac{3}{4}\right] = P\left[\left|\frac{Y-\nu}{\nu}\right| \leq \frac{3}{4}\right] = P\left[\left|\frac{Y-\nu}{\sqrt{2\nu}}\right| \leq \frac{3}{4}\sqrt{\frac{\nu}{2}}\right] \qquad \text{(0.5P)}$$
$$\approx \Phi\left(\frac{3}{4}\sqrt{\frac{\nu}{2}}\right) - \Phi\left(-\frac{3}{4}\sqrt{\frac{\nu}{2}}\right). \qquad \text{(0.5P)}$$

Für $\nu = 12$ erhalten wir

$$P\left[\left|\frac{Y}{\nu} - 1\right| \leq \frac{3}{4}\right] \approx \Phi\left(\frac{3}{4}\sqrt{6}\right) - \Phi\left(-\frac{3}{4}\sqrt{6}\right)$$
$$= 2\Phi\left(\frac{3}{4}\sqrt{6}\right) - 1 \qquad \text{(0.5P)}$$

Siehe nächste Seite!

und mit Hilfe der Abschätzung $\sqrt{6} \approx \sqrt{6.25} = 2.5$ folgt $\frac{3}{4}\sqrt{6} \approx \frac{15}{8} = 1.875$, wodurch wir den konkreten Wert

$$P\left[\left|\frac{Y}{\nu} - 1\right| \leq \frac{3}{4}\right] \approx 2\Phi\left(1.875\right) - 1 \approx 2 \cdot 0.9696 - 1 = 0.9392 \qquad (0.5\text{P})$$

erhalten.

3. (8 Punkte)

Peter und Hans sitzen pro Woche die zufälligen Zeiten P und H über ihren Stochastikserien. Peters Zeitaufwand P kann durch eine Uniform(0,1)-verteilte Zufallsvariable beschrieben werden. Hans, der gewieftere Stochastiker, braucht gegeben Peters Zeit P die zufällige Zeit H, die uniform auf (0,P) verteilt ist.

a) **(2.5 Punkte)** Bestimme die gemeinsame Dichte von H und P, und skizziere den Bereich, auf dem diese Dichte positiv ist.

b) **(1.5 Punkt)** Bestimme die Randdichte von H.

c) **(2 Punkte)** Bestimme den Erwartungswert von H. Vergleiche dies mit dem Erwartungswert von P.

d) **(2 Punkte)** Bestimme die Kovarianz zwischen H und P. Sind H und P unabhängig?

Lösung:

a) Nach Aufgabenstellung sind $P \sim \mathrm{U}(0,1)$ und $H|P \sim \mathrm{U}(0,P)$, d.h. die bedingte Dichte von H gegeben P ist durch

$$f_{H|P}(h|p) = \frac{1}{p} 1_{(0,p)}(h) \qquad (0.5\text{P})$$

gegeben. Also verschwindet $f_{H|P}$ ausserhalb des Dreiecks

$$D := \{\, (p,h) \mid 0 < h < p \leq 1 \,\}.$$

Für die gemeinsame Dichte von P und H ergibt sich:

$$f_{P,H}(p,h) = f_{H|P}(h|p) f_P(p) \; (0.5\text{P}) = \frac{1}{p} 1_{(0,p)}(h) 1_{[0,1]}(p). \qquad (0.5\text{P})$$

Der Bereich D, wo $f_{H,P}$ positiv ist, ist in Abbildung 1 skizziert. (1P)

b) Die Randdichte von H berechnet sich für $0 < h \leq 1$ als Integral über die gemeinsame Dichte, also

$$\begin{aligned} f_H(h) &= \int_{-\infty}^{\infty} f_{P,H}(p,h)\, dp \qquad (0.5\text{P}) \\ &= \int_0^1 \frac{1}{p} 1_{(0,p)}(h)\, dp = \int_h^1 \frac{1}{p}\, dp \qquad (0.5\text{P}) \\ &= -\log h. \qquad (0.5\text{P}) \end{aligned}$$

c) Mit partieller Integration folgt

$$\begin{aligned} E[H] &= -\int_0^1 h \log h \, dh \qquad (0.5\text{P}) \\ &= -\frac{h^2}{2}\log h\Big|_0^1 + \int_0^1 \frac{h^2}{2}\frac{1}{h}\,dh = 0 + \int_0^1 \frac{h^2}{2}\frac{1}{h}\,dh \qquad (0.5\text{P}) \\ &= \frac{h^2}{4}\Big|_0^1 = \frac{1}{4}. \qquad (0.5\text{P}) \end{aligned}$$

Andererseits ist $E[P] = 1/2$, also $E[H] = E[P]/2$. (0.5P)

Bemerkung:

$$\begin{aligned} -\frac{h^2}{2}\log h\Big|_0^1 &= -\frac{1}{2}\underbrace{\log 1}_{=0} + \lim_{h\to 0}\frac{h^2}{2}\log h \\ &= \lim_{h\to 0}\frac{\log h}{2h^{-2}} \qquad \Big(\text{unbestimmter Ausdruck des Typs `}\frac{\infty}{\infty}\text{`}\Big) \\ &= \lim_{h\to 0}\frac{1/h}{-4h^{-3}} = \lim_{h\to 0} -h^2/4 = 0, \end{aligned}$$

wobei wir die l'Hôpitalsche Regel benutzt haben.

Alternativ: Es gilt $E[H] = E[E[H|P]]$. (0.5P) Laut Aufgabenstellung ist H gegeben P U$(0, P)$-verteilt, so dass $E[H|P] = \frac{1}{2}P$ (0.5P) ist und damit

$$E[H] = \frac{1}{2}E[P] \; (0.5\text{P}) = \frac{1}{4}. \qquad (0.5\text{P})$$

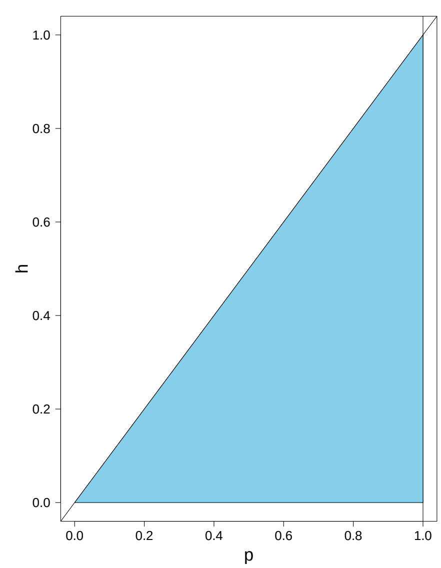

Abbildung 1: Bereich D, auf dem die gemeinsame Dichte von H und P positiv ist.

Siehe nächste Seite!

d) Man hat

$$\begin{aligned} E[PH] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} ph\, f_{P,H}(p,h)\, dh\, dp = \int_0^1 dp \int_0^p h\, dh\, dp = \int_0^1 \frac{p^2}{2}\, dp = \frac{1}{6}, \qquad \text{(1P)} \\ \mathrm{Cov}(P,H) &= E[PH] - E[P]E[H] = \frac{1}{6} - \frac{1}{2}\frac{1}{4} = \frac{1}{24} (\approx 0.0417). \qquad \text{(0.5P)} \end{aligned}$$

H und P sind somit nicht unabhängig. (0.5P)

4. (8 Punkte)

Unterhalb einer Kläranlage wurden 16 unabhängige Wasserproben aus einem Fluss entnommen und jeweils deren Ammoniumkonzentration X_i (angegeben in μgNH$_4$-N/ℓ) mit einem Messgerät bestimmt. Der Mittelwert der Proben ergab $\overline{x} = 204.2$.

Wir wollen nun wissen, ob mit diesem Experiment eine Überschreitung des Grenzwerts von 200 μgNH$_4$-N/ℓ nachgewiesen werden kann (auf dem 5%-Niveau).

a) **(2 Punkte)** Nimm an, die Standardabweichung der Messungen sei im Voraus aufgrund früherer Studien bekannt. Sie betrage 10 μgNH$_4$-N/ℓ. Finde einen geeigneten statistischen Test, um zu überprüfen, ob eine Grenzwertüberschreitung nachgewiesen werden kann. Wie lauten die Modellannahmen?

b) **(3.5 Punkte)**

Führe den Test aus Punkt a) durch. Gib dazu folgendes explizit an: die Null- und Alternativhypothesen H_0 und H_A, die Teststatistik, den realisierten Wert der Teststatistik, den Verwerfungsbereich und das Testergebnis.

c) **(1.5 Punkte)** Wie wahrscheinlich ist es, dass man mit 16 unabhängigen Wasserproben eine Grenzwertüberschreitung nachweisen kann, wenn die wahre Ammoniumkonzentration tatsächlich über dem Grenzwert und zwar bei 205 μgNH$_4$-N/ℓ liegt?

d) **(1 Punkt)** Wie wahrscheinlich ist es, dass man mit 16 unabhängigen Wasserproben fälschlicherweise eine Grenzwertüberschreitung nachweist, obwohl die wahre Ammoniumkonzentration bei 200 μgNH$_4$-N/ℓ liegt und den Grenzwert somit genau einhält?

Lösung:

a) Da die Standardabweichung der Messungen als bekannt vorausgesetzt wird, entscheiden wir uns für einen z-Test. (0.5P) Wir nehmen an, dass die Ammoniumkonzentrationen X_i unabhängig (0.5P) und identisch normalverteilt (0.5P) sind für $i = 1, \ldots, 16$, mit $\sigma^2 = 100$ und μ unbekannt. (0.5P)

Siehe nächste Seite!

b) Nullhypothese H_0: X_i i.i.d. $\sim \mathcal{N}(\mu_0, \sigma^2)$ mit $\mu_0 = 200, \sigma = 10$ (0.5P)

Alternative H_A: X_i i.i.d. $\sim \mathcal{N}(\mu, \sigma^2)$ mit $\mu > 200, \sigma = 10$ (einseitig) (0.5P)

Teststatistik: $Z = \frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1)$ unter H_0 (0.5P)

Verwerfungsbereich: Aus der Normalverteilungstabelle:

$\mathcal{K} = \{z \geq z_{1-\alpha}\} = [1.64, \infty)$.

(Dies entspricht dem Verwerfungsbereich $[204.1, \infty)$

auf der Skala von $\overline{X}$.) (1P)

Wert der Teststatistik: $z = \frac{\overline{x}-\mu_0}{\sigma/\sqrt{n}} = \frac{204.2-200}{\sigma/\sqrt{16}} = \frac{4.2}{2.5} = 1.68$ (0.5P)

Testentscheid: $1.68 \in \mathcal{K}$, also wird die Nullhypothese verworfen.

Die Überschreitung des Grenzwertes ist auf dem

5%-Niveau signifikant. (0.5P)

c) Aus a) folgt: Die Nullhypothese kann verworfen werden, falls der Mittelwert aller Messungen grösser als 204.1 ist,

$$\overline{X} > 204.1.$$

Um die Wahrscheinlichkeit zu berechnen, dass eine Grenzwertüberschreitung nachgewiesen werden kann (H_0 verworfen werden kann), geht man wieder zu einer standardisierten normalverteilten Zufallsvariable über. Mit $\mu_A = 205$ und $\sigma = 10$ erhält man

$$\begin{aligned} P_{\mu_A}[\overline{X} > 204.1] \quad \text{(0.5P)} \quad &= P_{\mu_A}\left[\frac{\overline{X}-\mu_A}{\sigma/\sqrt{n}} > \frac{204.1-\mu_A}{\sigma/\sqrt{n}}\right] \\ &= P_{\mu_A}\left[\frac{\overline{X}-\mu_A}{\sigma/\sqrt{n}} > -0.36\right] \quad \text{(0.5P)} \\ &= P[Z > -0.36]. \end{aligned}$$

Dies entspricht also der Wahrscheinlichkeit, dass eine normalverteilte Zufallsvariable Z mit Varianz 1,

$$Z \sim \mathcal{N}(0,1),$$

einen Wert grösser als -0.36 annimmt. Diese Wahrscheinlichkeit ist wegen der Symmetrie der Normalverteilung gleich

$$P[Z \leq 0.36] = 0.6406, \qquad \text{(0.5P)}$$

Siehe nächste Seite!

wie aus der Tabelle entnommen werden kann. Die gesuchte Wahrscheinlichkeit (die Macht des Tests an der Stelle $\mu = \mu_A = 205$) ist also ungefähr 64%.

d) Dies ist genau das Niveau des Testes und war als 5% vorgegeben. (1P)