Prof. Dr. J. Teichmann | ETH Zürich | Sommer 2017

Wahrscheinlichkeit und Statistik
BSc D-INFK

Name:	
Vorname:	
Stud. Nr.:	

Das Folgende bitte nicht ausfüllen!

Aufg.	Summe	Kontr.	Pkte.-Max.
1			10
2			10
3			10
4			10

Punktetotal:	
Vollständigkeit:	

Siehe nächste Seite!

Hinweise zur Prüfung

Prüfungsdauer: 2 Stunden.

Hilfsmittel: 10 A4-Seiten resp. 5 Blätter Zusammenfassung. Kein Taschenrechner!

Bitte beachten Sie folgende Punkte:

- Legen Sie Ihre Legi offen auf den Tisch.
- Tragen Sie Ihre Daten in dieses Deckblatt ein und schreiben Sie auf jedes Blatt Ihren Namen.
- Beginnen Sie jede Aufgabe **auf einem neuen Blatt**.
- Schreiben Sie nicht mit Bleistift, rotem oder grünem Kugelschreiber.
- Um die volle Punktzahl zu erreichen, begründen Sie alle Resultate durch Zwischenschritte und -rechnungen (ausser Aufgabe 1) und vereinfachen Sie die Resultate so weit wie möglich.
- Lesen Sie alle Aufgaben durch, bevor Sie beginnen. Für eine genügende Note wird nicht erwartet, dass Sie alle Aufgaben in der Ihnen zur Verfügung stehenden Zeit lösen können.
- Es dürfen sich nur erlaubte Hilfsmittel auf dem Tisch befinden, d.h. 10 A4-Seiten resp. 5 Blätter Zusammenfassung. Kein Taschenrechner!

Siehe nächste Seite!

Lösung

---

1. Bei den folgenden 10 Fragen ist jeweils genau eine Antwort richtig. Es gibt pro richtig beantwortete Frage 1 Punkt und pro falsche Antwort 1/2 Punkt Abzug. Minimal erhält man für die gesamte Aufgabe 0 Punkte. **Bitte benützen Sie das beiliegende Antwortblatt.**

a) Sei $X \sim Bin(n, \frac{1}{4})$ für ein $n \in \mathbb{N}$. Ferner nehmen wir an, dass $E[X] = 1$. Wie gross ist $P[X \leq 1]$?

1. $P[X \leq 1] = \left(\frac{3}{4}\right)^3 \frac{7}{4}$.
2. $P[X \leq 1] = \left(\frac{3}{4}\right)^3$.
3. $P[X \leq 1] = 4\left(\frac{1}{4}\right)^4$.

Lösung: 1 ist richtig.

$$E[X] = 1 = \frac{n}{4} \Rightarrow n = 4.$$

und

$$\begin{aligned} P[X \leq 1] &= P[X = 0] + P[X = 1] \\ &= \binom{4}{0}\left(\frac{1}{4}\right)^0\left(\frac{3}{4}\right)^4 + \binom{4}{1}\left(\frac{1}{4}\right)^1\left(\frac{3}{4}\right)^3 \\ &= \left(\frac{3}{4}\right)^3\left(1 + \frac{3}{4}\right). \end{aligned}$$

b) Betrachten Sie folgende Tabelle:

X_i	X_1	X_2	X_3	X_4
$\text{Var}[X_i]$	2	?	10	6

wo X_1, X_2, X_3, X_4 unabhängig voneinander sind. Berechnen Sie $\text{Var}[X_2]$, gegeben, dass die Varianz von $Z = 2X_1 + 3(X_2 - X_4)$ gleich 80 ist.

1. $\frac{94}{3}$.
2. $\frac{34}{9}$.
3. 2.

Lösung: 3 ist richtig

$$80 = \text{Var}[Z] = 4\text{Var}[X_1] + 9\text{Var}[X_2] + 9\text{Var}[X_4] = 9\text{Var}[X_2] + 62$$

$$\Rightarrow \text{Var}[X_2] = \frac{80 - 62}{9} = 2.$$

Siehe nächste Seite!

c) Sei $X \sim Geom(p)$ und $Var[X] = 56$. Berechnen Sie p.

1. $p = \frac{1}{8}$.
2. $p = \frac{1}{56}$.
3. $p = \frac{\sqrt{56}}{2}$.

Lösung: 1 ist richtig

$$\text{Var}[X] = \frac{1-p}{p^2} = 56 \Rightarrow 56p^2 + p - 1 = 0$$

Da

$$p = \frac{-1 \pm \sqrt{225}}{112}, \tag{1}$$

ist die positive Lösung von (1) gleich $p_1 = \frac{7}{56}$.

d) Eine Biologiestudentin untersucht die Augenfarbe von Fruchtfliegen. Unter jeweils 1000 Fliegen haben 20 weisse Augen, der Rest hat rote Augen. Wir fangen nun 1000 Fruchtfliegen und nehmen zufällig 200 heraus. Darunter befindet sich eine Anzahl W von Fliegen mit weissen Augen. Wie ist W verteilt?

1. W ist hypergeometrisch verteilt mit Parametern $n = 1000, m = 200$ und $r = 20$.
2. $W \sim NB(r, p)$, mit $r = 20$ und $p = 1/50$.
3. $W \sim Geom(p)$ mit $p = 1/50$.

Lösung: 1 ist richtig.
In diesem Experiment ziehen wir zufällig aus einer Menge von 1000 Fruchtfliegen 200 heraus. W beschreibt die Anzahl der Fliegen mit weissen Augen (Typ 1) in dieser Stichprobe vom Umfang 1000, also ist W hypergeometrisch verteilt.

e) Die Dichte einer Zufallsvariablen X sei gegeben durch

$$f(x) = \begin{cases} c + x & \text{falls } -\frac{c}{2} \leq x \leq 0, \\ c - x & \text{falls } 0 \leq x \leq \frac{c}{2}, \\ 0 & \text{sonst.} \end{cases}$$

Bestimmen Sie die Konstante c.

1. $c = \frac{1}{\sqrt{3}}$.
2. $c = \frac{2}{\sqrt{3}}$.
3. $c = \frac{3}{\sqrt{2}}$.

Siehe nächste Seite!

Lösung: 2 ist richtig.
Da $f(x)$ eine Dichte ist, muss folgende Gleichung erfüllt sein:

$$1 \overset{!}{=} \int_{-\frac{c}{2}}^{0} (c+x)\ dx + \int_{0}^{\frac{c}{2}} (c-x)\ dx = (cx + \frac{1}{2}x^2)\big|_{-\frac{c}{2}}^{0} + (cx - \frac{1}{2}x^2)\big|_{0}^{\frac{c}{2}}$$
$$= \frac{c^2}{2} - \frac{c^2}{8} + \frac{c^2}{2} - \frac{c^2}{8} = \frac{3}{4}c^2.$$

(Alternativ kann man die Fläche unter der Funktion f auch elementargeometrisch bestimmen.) Es folgt $c = \pm\frac{2}{\sqrt{3}}$ und weil $f(x) \geq 0$ sein muss, folgt $c = \frac{2}{\sqrt{3}}$.

f) Sei $(X_k)_k$ eine Folge von i.i.d. Zufallsvariablen mit Dichtefunktion

$$f(x) = \begin{cases} \frac{1}{6}, & \text{für } 0 \leq x < 2, \\ \frac{1}{3}, & \text{für } 2 \leq x \leq 4, \\ 0, & \text{sonst.} \end{cases}$$

Berechne den (P-f.s.) Grenzwert

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} X_k.$$

1. $\frac{1}{2}$.
2. ∞.
3. $\frac{7}{3}$.

Lösung: 3 ist richtig

$$E[X_k] = \frac{1}{6}\int_0^2 x dx + \frac{1}{3}\int_2^4 x dx = \frac{7}{3}.$$

Nach dem starken Gesetz der grossen Zahl gilt mit Wahrscheinlichkeit 1

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} X_k = E[X_1] = \frac{7}{3}.$$

g) Aus einer Urne mit 40 roten und 80 blauen Kugeln wird 27 Mal eine Kugel gezogen und wieder zurückgelegt. Sei A das Ereignis, dass total mindestens 18 rote Kugeln gezogen werden. Welche Aussage ist richtig?

1. $P[A] = \sum_{k=18}^{27} \left(\frac{1}{3}\right)^k \left(\frac{2}{3}\right)^{27-k}$.
2. Wegen der Chebyshev-Ungleichung gilt $P[A] \leq \frac{2}{27}$.
3. Wegen der Chernoff-Schranke gilt $P[A] \geq 2\left(\frac{e^1}{2^2}\right)^9$.

Lösung: 2 ist richtig

Mit Chebyshev-Ungleichung und $E[X] = n \cdot p = 9$, mit $n = 27$ und $p = \frac{40}{120}$:

$$P[A] = P[X \geq 18] = P[X - 9 \geq 9] \leq P[|X - 9| \geq 9] \leq \frac{E[|X - 9|^2]}{9^2}$$
$$\leq \frac{\mathrm{Var}[X]}{9^2} = \frac{np(1-p)}{81} = \frac{6}{81} = \frac{2}{27}.$$

h) Sei X eine exponentialverteilte Zufallsvariable mit Parameter λ, d.h. $X \sim Exp(\lambda)$. Berechnen Sie die momenterzeugende Funktion $M_X(t)$.

1. $M_X(t) = \frac{\lambda}{\lambda - t}$ für $t < \lambda$.
2. $M_X(t) = \frac{\lambda}{t - \lambda}$ für $t > \lambda$.
3. $M_X(t) = \lambda$ für $t \in \mathbb{R}$.

Lösung: 1 ist richtig.

Man rechnet

$$M_X(t) = E[e^{tX}] = \int_0^\infty e^{tx} \lambda e^{-\lambda x} dx = \lambda \int_0^\infty e^{-(\lambda - t)x} dx = -\frac{\lambda}{\lambda - t} e^{-(\lambda - t)x} \Big|_0^\infty$$
$$= \begin{cases} \frac{\lambda}{\lambda - t} & \text{falls } t < \lambda, \\ \infty & \text{falls } t \geq \lambda. \end{cases}$$

Alternativ: Da X positiv ist, ist $M_X(t)$ monoton wachsend. Somit kann man 2. ausschliessen. Ausserdem kann 3. i.A. nicht stimmen, da $M_X(0) = 1$ (für alle λ).

i) Welche Aussage gilt für einen Test zum Niveau α?

1. Das Niveau α, kontrolliert die Wahrscheinlichkeit des Fehlers 2. Art.
2. Je kleiner das Niveau α, desto kleiner die Wahrscheinlichkeit, dass die Nullhypothese H_0 abgelehnt wird, obwohl sie richtig ist.
3. Die Wahrscheinlichkeit, dass die Nullhypothese beibehalten wird, wenn sie richtig ist, beträgt α.

Lösung: 2 ist richtig.

j) Seien $X_1, \ldots, X_n$ unter P_ϑ i.i.d. mit endlichem Erwartungswert. Ferner sei $Y = \sum_{i=1}^n \lambda_i X_i$ für beliebige $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$ mit $\lambda_1 + \cdots + \lambda_n = 1$. Ist dann Y ein erwartungstreuer Schätzer für $E_\vartheta[X]$?

1. Diese Aussage ist falsch.
2. Es sind zu wenig Informationen vorhanden, um eine Aussage zu machen.
3. Diese Aussage ist richtig.

Siehe nächste Seite!

Lösung: 3 ist richtig.

Wegen der Linearität und da $X_1, \ldots, X_n$ unter P_ϑ identisch verteilt sind, gilt

$$E_\vartheta[Y] = \sum_{i=1}^{n} \lambda_i E_\vartheta[X_i] \overset{\text{i.d.}}{=} E_\vartheta[X_1] \sum_{i=1}^{n} \lambda_i = E_\vartheta[X_1].$$

Siehe nächste Seite!

2. (**10 Punkte**)

Seien X und Y zwei diskrete Zufallsvariablen mit folgender gemeinsamer Gewichtsfunktion:

$$p(j,k) = P[X=j, Y=k] = \begin{cases} C\left(\frac{1}{2}\right)^k & \text{für } k = 2,3,\dots \text{ und } j = 1,2,\dots,k-1 \\ 0 & \text{sonst.} \end{cases}$$

a) **(3 Punkte)**Bestimmen Sie die Konstante C.

Lösung:

Da p eine Gewichtsfunktion sein soll, muss gelten

$$1 \overset{!}{=} \sum_{k=2}^{\infty}\sum_{j=1}^{k-1} P[X=j,Y=k] = C\sum_{k=2}^{\infty}\sum_{j=1}^{k-1}\left(\frac{1}{2}\right)^k = C\sum_{j=1}^{\infty}\sum_{k=j+1}^{\infty}\left(\frac{1}{2}\right)^k$$
$$= C\sum_{j=1}^{\infty}\left(\frac{1}{2}\right)^{j+1}\sum_{k=0}^{\infty}\left(\frac{1}{2}\right)^k = C\sum_{j=1}^{\infty}\frac{\left(\frac{1}{2}\right)^{j+1}}{1-\frac{1}{2}} = C\sum_{j=1}^{\infty}\left(\frac{1}{2}\right)^j = C\frac{\frac{1}{2}}{1-\frac{1}{2}} = C.$$

Also ist $C = 1$.

b) **(1 Punkt)**Beweisen Sie dass X geometrisch verteilt mit Parameter $1/2$ ist.

Lösung:

$$p_X(j) = \sum_{k=j+1}^{\infty} p(j,k) = \sum_{k=j+1}^{\infty}\left(\frac{1}{2}\right)^k = \left(\frac{1}{2}\right)^{j+1}\sum_{k=0}^{\infty}\left(\frac{1}{2}\right)^k = \left(\frac{1}{2}\right)^j \quad \text{für } j \geq 1;$$

also ist X geometrisch verteilt mit Parameter $p = \frac{1}{2}$.

c) **(2 Punkte)**Beweisen Sie, dass bedingt auf das Ereignis $\{Y = k\}$ die Zufallsvariable X gleichverteilt auf $\{1,\dots,k-1\}$ ist.

Lösung:

$$p_Y(k) = \sum_{j=1}^{k-1} p(j,k) = (k-1)\left(\frac{1}{2}\right)^k,$$

$$p_{X|Y}(j \mid k) = \frac{p(j,k)}{p_Y(k)} = \frac{\left(\frac{1}{2}\right)^k}{(k-1)\left(\frac{1}{2}\right)^k} = \frac{1}{k-1} \quad \text{für } j = 1,\dots,k-1$$

Siehe nächste Seite!

und sonst $p_{X|Y}(j\,|\,k) = 0$ d.h. gegeben $Y = k$ ist X gleichverteilt auf $\{1, \ldots, k-1\}$.

d) **(2 Punkte)**Berechnen Sie $E[X]$.

Lösung:

$$E[X] = \sum_{j=0}^{\infty} P[X > j] = \sum_{j=0}^{\infty} (1 - \frac{1}{2})^j = \frac{1}{1 - (1 - \frac{1}{2})} = 2.$$

$$E[X] = \sum_{j=1}^{\infty} jP[X = j] = \sum_{j=1}^{\infty} \sum_{k=1}^{j} P[X = j] = \sum_{k=1}^{\infty} \sum_{j=k}^{\infty} P[X = j]$$

$$= \sum_{k=1}^{\infty} P[X \geq k] = \sum_{k=1}^{\infty} P[X > k-1] = \sum_{k=0}^{\infty} P[X > k]$$

$$E[X] = \sum_{j=1}^{\infty} jP[X = j] = \sum_{j=1}^{\infty} j \frac{1}{2^{j-1}} * \frac{1}{2}$$

e) **(2 Punkte)**Berechnen Sie $E[Y]$ und $\mathrm{Var}[X]$.

Hinweis: Mann kann benutzen, dass für eine geometrisch verteilte Zufallsvariable mit Parameter p gilt

$$E[Z(Z-1)] = \frac{2(1-p)}{p^2}.$$

Lösung:

$$p_Y(k) = \sum_{j=1}^{k-1} p(j,k) = (k-1) \left(\frac{1}{2}\right)^k \text{ für } k \geq 2.$$

$$E[Y] = \sum_{k=2}^{\infty} k(k-1) \left(\frac{1}{2}\right)^k = E[X(X-1)] = 4$$

$$\mathrm{Var}[X] = E[X^2] - E[X]^2 = E[X(X-1)] + E[X] - E[X]^2 = 2.$$

3. (**10 Punkte**)

Seien X und Y zwei Zufallsvariablen mit gemeinsamer Dichtefunktion

$$f(x,y) = \begin{cases} \frac{1}{x^2y^2} \text{ für } x \geq 1, y \geq 1 \\ 0 \text{ sonst.} \end{cases}$$

Definiere

$$U := \frac{X}{Y} \quad \text{und} \quad V := XY.$$

a) **(3 Punkte)** Berechnen Sie die Verteilungsfunktion F_U der Zufallsvariable U.

Lösung:

Für $u \leq 0$, $F(u) = 0$

Für $u > 1$

$$\begin{aligned} F(u) = \mathbb{P}(X/Y \leq u) &= \mathbb{P}(X \leq Yu) \\ &= \int_1^\infty \int_1^{yu} \frac{1}{x^2y^2} dxdy \\ &= \int_1^\infty \frac{1}{y^2}\left(1 - \frac{1}{uy}\right) dy \\ &= \int_1^\infty \left(\frac{1}{y^2} - \frac{1}{uy^3}\right) dy \\ &= \int_1^\infty \left(\frac{1}{y^2}\right) dy - \int_1^\infty \left(\frac{1}{uy^3}\right) dy \\ &= 1 - \frac{1}{2u}. \end{aligned}$$

Für $0 < u \leq 1$

$$\begin{aligned} F(u) = \mathbb{P}(X/Y \leq u) &= \mathbb{P}(X \leq Yu) \\ &= \int_{1/u}^\infty \int_1^{yu} \frac{1}{x^2y^2} dxdy \\ &= \int_{1/u}^\infty \frac{1}{y^2}\left(1 - \frac{1}{uy}\right) dy \\ &= \int_{1/u}^\infty \left(\frac{1}{y^2} - \frac{1}{uy^3}\right) dy \\ &= \int_{1/u}^\infty \left(\frac{1}{y^2}\right) dy - \int_{1/u}^\infty \left(\frac{1}{uy^3}\right) dy \\ &= u - \frac{u}{2} = \frac{u}{2}. \end{aligned}$$

Siehe nächste Seite!

b) (3 Punkte) Berechnen Sie die Verteilungsfunktion F_V der Zufallsvariable V.

Lösung:

Für $v < 1$, $F(v) = 0$ und für $v \geq 1$,

$$\begin{aligned}
F(v) &= \mathbb{P}(XY \leq v) = \mathbb{P}(X \leq v/Y) \\
&= \int_1^v \int_1^{v/y} \frac{1}{x^2 y^2} dx dy \\
&= \int_1^v \frac{1}{y^2} \int_1^{v/y} \frac{1}{x^2} dx dy \\
&= \int_1^v \frac{1}{y^2} \left(1 - \frac{y}{v}\right) dy \\
&= \int_1^v \frac{1}{y^2} dy - \int_1^v \frac{1}{yv} dy \\
&= 1 - \frac{1}{v} \left(1 + \log(v)\right).
\end{aligned}$$

c) (2 Punkte) Bestimmen Sie die Dichtefunktion f_U der Zufallsvariable U.

Lösung:

Aus **a)**, für $u > 1$

$$f(u) = \frac{dF(u)}{du} = \frac{1}{2u^2}.$$

und

$$f(u) = \frac{dF}{du} = \frac{1}{2}, 0 < u \leq 1.$$

Ausserdem gilt $f(u) = 0$ für $u \leq 0$.

d) (2 Punkte) Bestimmen Sie die Dichtefunktion f_V der Zufallsvariable V.

Lösung:

$$\frac{dF}{dv} = f(v) = \frac{\log(v)}{v^2},$$

for $v \geq 1$ and $f(v) = 0$ for $v < 1$.

Siehe nächste Seite!

4. (**10 Punkte**)

Wir haben eine Münze mit einer Seite rot und der anderen Seite blau gefärbt, und wir vermuten, dass die Münze gezinkt ist und eher auf der blauen Seite landet. Also machen wir ein Experiment, in dem wir die Münze 10 Mal werfen, und wir beobachten jeweils, ob sie auf blau landet. Gehen Sie davon aus, dass alle Würfe unabhängig voneinander sind. Sei $X_i = 1$, wenn der i-te Wurf auf blau landet, und sonst gleich 0.

Wurf i	1	2	3	4	5	6	7	8	9	10
x_i	0	1	1	0	0	1	0	0	1	1

a) **(9 Punkte)** Führen Sie einen Test zum Signifikanzniveau $\alpha = 2.5\%$ durch, um festzustellen, ob die Münze gezinkt ist.

Geben Sie folgendes an:

1. das Modell,
2. die Nullhypothese,
3. die Alternativhypothese,
4. die Teststatistik,
5. die Verteilung der Teststatistik unter der Nullhypothese,
6. den Verwerfungsbereich,
7. den beobachteten Wert der Teststatistik, sowie
8. den Testentscheid.

Solution

Modell: Unter P_p sind die X_i, i.i.d., $\sim$ Ber(p), $i = 1, ..., 10$, p unbekannt.

Nullhypothese H_0**:** $p = p_0 = 0.5$.

Alternativhypothese H_A**:** $p = p_A > p_0$.

Teststatistik: $T = \sum_{i=1}^{10} X_i$, denn

$$\begin{aligned} R(x_1, \ldots, x_{10}; \lambda_0, \lambda_A) &= \frac{L(x_1, \ldots, x_{10}; \lambda_0)}{L(x_1, \ldots, x_{10}; \lambda_A)} \\ &= \frac{p_0^T (1-p_0)^{10-T}}{p_A^T (1-p_A)^{10-T}} \\ &= \left(\frac{p_0(1-p_A)}{p_A(1-p_0)}\right)^T \left(\frac{1-p_0}{1-p_A}\right)^{10} \end{aligned}$$

Da $\left(\frac{p_0(1-p_A)}{p_A(1-p_0)}\right) < 1$ wird $R(x_1, \ldots, x_{10}; p_0, p_A)$ klein, genau dann, wenn $\sum_{i=1}^{10} x_i$ gross ist. Wir wählen als Teststatistik also

$$T = \sum_{i=1}^{10} X_i.$$

Siehe nächste Seite!

Verteilung der Teststatistik unter H_0: $T \sim \text{Bin}(10, 1/2)$.

Verwerfungsbereich: Der kritische Bereich "Quotient klein" hat die äquivalente Form "Summe gross", also ist der Verwerfungsbereich von der Form $K = (k, \infty)$. Um das Signifikanzniveau einzuhalten, muss gelten

$$P_{p_0}[T \in K] = P_{p_0}[T > k] \leq 2.5\% \Leftrightarrow P_{p_0}[T \leq k] \geq 97.5\%.$$

k	$P_{\lambda_0}[T \leq k]$
0	0.001
1	0.011
2	0.055
3	0.172
4	0.377
5	0.623
6	0.828
7	0.945
8	0.989
9	0.999
10	1.000

Deshalb haben wir als Verwerfungsbereich$= [8, \infty)$.

Beobachteter Wert der Teststatistik: $t = T(\omega) = 6$.

Testentscheid: Da 6 nicht im Verwerfungsbereich liegt, wird die Nullhypothese **nicht** verworfen.

b) **(1 Punkt)**Wie hoch ist das kleinste Niveau, bei welchem die Nullhypothese verworfen wird (dies ist der so genannte P-Wert)?

Solution:

Wir rechnen, dass $P_{H_0}[T > 6] = (1 - P[T \leq 6]) = 1 - 0.828 = 0.172$ gilt.

Viel Erfolg!