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!

Aufgaben

---

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) Ein Kasten enthält insgesamt 10 Äpfel, r Äpfel sind rot und der Rest ist grün. Wir nehmen zwei Äpfel aus dem Kasten gleichmässig zufällig ohne Zürucklegen. Die Wahrscheinlichkeit, dass zwei rote Äpfel gezogen werden, ist 2/9. Wie viele rote Äpfel enthält der Kasten?

1. 10
2. 5
3. 2

Solution:

2 ist richtig. Die Wahrscheinlichkeit, dass zwei rote Äpfel gezogen werden ist

$$\frac{r(r-1)}{10 \cdot 9} = \frac{2}{9} \Rightarrow r^2 - r - 20 = 0.$$

$$r = \frac{1 + \sqrt{81}}{2} = 5.$$

b) Betrachten Sie eine diskrete Zufallsvariable X mit der Gewichtsfunktion p_X gemäss folgender Tabelle:

x_i	1	-1	2
$p_X(x_i)$	$\frac{1}{4}$	$\frac{1}{4}$	?

Berechnen Sie $\mathrm{Var}[2X+5]$.

1. 6
2. 8
3. 3

Solution: 1 ist richtig.

$$p_X(1) + p_X(-1) + p_X(2) = \frac{1}{4} + \frac{1}{4} + p_X(2) = 1 \Rightarrow p_X(2) = \frac{1}{2}.$$

$$E[X] = 1, E[X^2] = \frac{5}{2}.$$

Siehe nächste Seite!

$$\mathrm{Var}[2X+5] = 4\mathrm{Var}[X] = 4\left(\frac{5}{2}-1\right) = 6.$$

c) Sei $X \sim Geom(1/4)$. Wie gross ist die Wahrscheinlichkeit $P(X > 30|X > 20)$?

1. $\left(\frac{3}{4}\right)^{10}$
2. $\left(\frac{1}{4}\right)^{10}$
3. $\left(\frac{3}{4}\right)^{30}$

Solution 1 ist richtig

$$\begin{aligned} &P(X > 30|X > 20) \\ &= \frac{P(X > 30, X > 20)}{P(X > 20)} = \frac{P(X > 30)}{P(X > 20)} = \frac{(1-1/4)^{30}}{(1-1/4)^{20}} = (1-1/4)^{10} = \left(\frac{3}{4}\right)^{10}. \end{aligned}$$

d) Eine Biologiestudentin untersucht die Augenfarbe von Fruchtfliegen. Unter jeweils 1000 Fliegen haben 20 weisse Augen, der Rest hat rote Augen. W sei die Anzahl nötiger Untersuchungen (mit Zurücklegen), um 20 Fliegen mit weissen Augen zu finden. Was ist die Verteilung von W?

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

Solution: 3. ist richtig.

Hier ist die Wartezeit auf den r-ten Erfolg mit $r = 20$ gesucht und $p = \frac{20}{1000}$. Somit handelt es sich um eine Verallgemeinerung der geometrischen Verteilung, nämlich der negativbinomialen Verteilung.

e) Die gemeinsame Dichte $f(x,y)$ der Zufallsvariablen X und Y sei auf dem Rechteck $R := \{(x,y) : -1/2 \leq x \leq 1/2, -1 \leq y \leq 1\}$ konstant gleich c und Null ausserhalb. Welche der folgenden Aussagen trifft zu?

1. Man muss $c = 1$ wählen, damit $f(x,y)$ wirklich eine Dichte ist.
2. Wenn $c = \frac{1}{2}$, dann sind die Zufallsvariablen unabhängig.
3. Keine der obigen beiden Behauptungen ist richtig.

Solution: 2 ist richtig.

$$1 = \int_{-1}^{1}\int_{-1/2}^{1/2} c\,dxdy = 2 \Rightarrow c = \frac{1}{2}.$$

Sogar,

Siehe nächste Seite!

$$f(x,y) = \frac{1}{2}1_{\{-1/2\le x\le 1/2, -1\le y\le 1\}} = 1_{\{-1/2\le x\le 1/2\}} \cdot \frac{1}{2}1_{\{-1\le y\le 1\}} = f_X(x)\cdot f_Y(y),$$

da X und Y unabhängig sind.

f) Die Anzahl Turbinen, die pro Woche in einer Fabrik hergestellt werden, sei eine Zufallsvariable mit Erwartungswert $\mu = 40$ und Varianz $\sigma^2 = 80$. Aus diesen Informationen man kann schliessen:

1. $P[|X-40| < 10] \geq 0.20$.
2. X ist normalverteilt.
3. Keine dieser Aussagen ist möglich.

Solution: 1 ist richtig correct.

Bei Chebyshev ungleichung, wo $E[X] = 40$ und $\mathrm{Var}[X] = 80$, gilt

$$P[|X-40| \geq 10] \leq \frac{\mathrm{Var}[X]}{10^2} = \frac{80}{100} = 0.8.$$

da $P[|X-40| < 10] = 1 - \mathbb{P}[|X-40| \geq 10] \geq 0.2$.

g) Sei $(X_i)_{i\in\mathbb{N}}$ eine Folge von Zufallsvariablen, die alle den gleichen Erwartungswert $E[X_i] = \mu$ und die gleiche Varianz $Var[X_i] = \sigma^2$ haben. Sei $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$. Dann gilt:

1. $\bar{X}_n$ konvergiert für $n \to \infty$ in Wahrscheinlichkeit gegen $\mu = E[X_1]$.
2. $\bar{X}_n$ konvergiert für $n \to \infty$ P-fast sicher gegen $\mu = E[X_1]$.
3. Keine der obigen Aussagen ist im Allgemeinen richtig.

Solution: 3 ist richtig, da die Annahme von Unabhängigkeit oder unkorreliert Zufallsvariablen benötigt ist.

h) Sei $\Phi(t) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^t e^{-\frac{1}{2}s^2}ds$. Was gilt dann für $\Phi(-t)$?

1. $\Phi(-t) = -\Phi(t)$.
2. $\Phi(-t) = 1 - \Phi(t)$.
3. $\Phi(-t) = \Phi(t)$.

Solution: 2 ist richtig.

Für X Normalverteilt, $\mathcal{N}(0,1)$. Es gilt $P(X \leq t) = \Phi(t)$.

$$\Phi(-t) = P(X \leq -t) = P(-X \leq -t) = P(X \geq t) = 1 - P(X \leq t) = 1 - \Phi(t).$$

i) Sei X eine Bernoulliverteilte Zufallsvariable mit Parameter $1/2$. Die momenterzeugende Funktion $M_X(t)$ von X ist:

1. $M_X(t) = \frac{1}{2} + \frac{e^{-t}}{2}$ für $t \in \mathbb{R}$.

2. $M_X(t) = \frac{e^t}{2}$ für $t \in \mathbb{R}$.
3. $M_X(t) = \frac{1}{2} + \frac{e^t}{2}$ für $t \in \mathbb{R}$.

Solution: 3 ist richtig Für $t \in \mathbb{R}$,

$$M_X(t) = e^t \cdot P(X = 1) + e^0 \cdot P(X = 0) = \frac{1}{2} + \frac{e^t}{2}.$$

j) Seien $X_1, \ldots, X_n$ unter P_ϑ i.i.d. $\sim$ Poisson(ϑ) mit $\vartheta > 0$. Bestimmen Sie den Maximum-Likelihood-Schätzer für ϑ.

1. Der Maximum-Likelihood-Schätzer für ϑ lautet $\frac{1}{n}\sum_{i=1}^n X_i$.
2. Der Maximum-Likelihood-Schätzer für ϑ lautet $\sum_{i=1}^n X_i$.
3. Der Maximum-Likelihood-Schätzer für ϑ lautet $n\sum_{i=1}^n X_i$.

Solution: 1 ist richtig

$$l(x_1, \ldots, x_n, \vartheta) = \log\left(\prod_{i=1}^n e^{-\vartheta}\frac{\vartheta^{x_i}}{x_i!}\right)$$

$$= \log\left(e^{-\vartheta n}\prod_{i=1}^n \frac{\vartheta^{x_i}}{x_i!}\right) = -\vartheta n + \log(\vartheta)\sum_{i=1}^n x_i - \sum_{i=1}^n \log(x_i!).$$

$$\frac{\partial}{\partial\vartheta} l(x_1, \ldots, x_n, \vartheta) = -n + \frac{\sum_{i=1}^n x_i}{\vartheta} = 0 \Rightarrow \vartheta = \frac{\sum_{i=1}^n x_i}{n}.$$

Siehe nächste Seite!

2. (**10 Punkte**) Sie haben sich im Nationalpark von Guarinabo verirrt. TouristInnen machen 2/3 aller BesucherInnen im Park aus. Wird eine TouristIn nach dem Weg gefragt, gibt sie eine korrekte Antwort mit Wahrscheinlichkeit 3/4. Die übrigen ParkbesucherInnen sind Einheimische, und wenn sie nach dem Weg gefragt werden geben sie stets eine falsche Antwort (Antworten auf eine wiederholte Frage sind unabhängig, sogar wenn sie mehrmals an die gleiche Person gestellt wird).

a) **(2 Punkte)**Sie fragen eine zufällig vorbeilaufende BesucherIn, ob der Ausgang aus dem Park in östlicher oder westlicher Richtung liegt (es gibt nur einen Ausgang). Wie gross ist die Wahrscheinlichkeit, dass ihre Antwort korrekt ist?

Lösung:

Bezeichnen Sie mit C das Ereignis $\{$ Die Antwort ist richtig$\}$ und mit T das Ereignis
$\{$der Parkebuscher ist ein Tourist$\}$. Durch der Satz der totalen Wahrscheinlichkeit,

$$P[C] = P[C|T]P[T] + P[C|T^c]P[T^c] = \left(\frac{3}{4} \times \frac{2}{3}\right) + \left(0 \times \frac{1}{3}\right) = \frac{1}{2}.$$

b) **(2 Punkte)**Wie gross ist die Wahrscheinlichkeit, dass die BesucherIn, die Sie gefragt haben, eine TouristIn war, wenn Sie wissen, dass ihre Antwort falsch war?

Lösung: Mit Bayes Formel gilt,

$$P[T|C^c] = \frac{P[C^c|T] \cdot P[T]}{P[C^c]} = \frac{\frac{1}{4} \cdot \frac{2}{3}}{1 - 1/2} = \frac{1}{3}.$$

Angenommen, die Anzahl N aller BesucherInnen im Nationalpark an einem gegebenen Tag sei eine Poissonverteilte Zufallsvariable mit Parameter $\lambda > 0$ (i.e. $\mathbb{P}(N = k) = e^{-\lambda}\lambda^k/k!$, für $k \in \{0, 1, 2, \dots\}$).

c) **(1.5 Punkte)**Jede Person, die den Park betritt, ist männlich mit Wahrscheinlichkeit p und weiblich mit Wahrscheinlichkeit $1 - p$, unabhängig von allen anderen BesucherInnen. Seien X, resp. Y die Gesamtanzahl männlicher, resp. weiblicher BesucherInnen im Nationalpark von Guarinabo. Bestimmen Sie

$$P(X = i|N = i + j),$$

in Abhängigkeit von i und j für alle ganzzahligen $i, j \geq 0$. Begründen Sie Ihre Antwort.

Siehe nächste Seite!

Lösung:

Wir wissen dass jede Person die den Park betritt ist männlich mit Wahrscheinlichkeit p und weiblich mit Wahrscheinlichkeit $1-p$, unabhängig von allen anderen BesucherInnen. Für jede Person haben wir also ein 0-1 Experiment. Der Erfolgsparameter ist dabei p. Wenn insgesamt $i+j$ Personen in den Park betreten, so haben wir eine Summe von $i+j$ unabhängigen 0-1 Experimenten mit demselben Erfolgsparameter.

$$P(X=i, Y=j|N=i+j) = \binom{i+j}{i} p^i (1-p)^j.$$

d) **(2 Punkte)** Bestimmen Sie die gemeinsame Gewichtsfunktion von X und Y, $P(X=i, Y=j)$, $i, j \geq 0$.

Lösung:

$$\begin{aligned} &P(X=i, Y=j) = P(X=i, Y=j|N=i+j) \cdot P(N=i+j) \\ &+ \underbrace{P(X=i, Y=j|N \neq i+j)}_{=0} \cdot P(N \neq i+j) = \binom{i+j}{i} p^i (1-p)^j e^{-\lambda} \frac{\lambda^{i+j}}{(i+j)!} \\ &= \frac{e^{-\lambda p} (\lambda p)^i}{i!} \times \frac{e^{-\lambda(1-p)} (\lambda(1-p))^j}{j!}. \end{aligned}$$

e) **(2.5 Punkte)** Was ist die Verteilung von X und Y? Sind sie unabhängig? Begründen Sie Ihre Antwort.

Lösung:

$$P(X=i) = \sum_{j=0}^{\infty} P(X=i, Y=j) = \frac{e^{-\lambda p} (\lambda p)^i}{i!} \sum_{j=0}^{\infty} \frac{e^{-\lambda(1-p)} (\lambda(1-p))^j}{j!} = \frac{e^{-\lambda p} (\lambda p)^i}{i!}.$$

Analog,

$$P(Y=j) = \frac{e^{-\lambda(1-p)} (\lambda(1-p))^j}{j!}.$$

Furthermore, we have that

$$P(X=i, Y=j) = \frac{e^{-\lambda p} (\lambda p)^i}{i!} \times \frac{e^{-\lambda(1-p)} (\lambda(1-p))^j}{j!} = P(X=i) P(Y=j).$$

Da X und Y unabhängig sind.

Siehe nächste Seite!

3. (**10 Punkte**)

Seien X und Y zwei unabhängige Zufallsvariablen, beide exponentialverteilt mit Parameter $\lambda > 0$. Definiere

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

Seien f_U und f_V die zu U und V gehörigen Dichtefunktionen.

a) **(3 Punkte)** Berechnen Sie die Dichtefunktion f_U und Verteilungsfunktion F_U der Zufallsvariable U

Lösung:

Zuerst bemerken wir, dass U Werte in $(0,1)$ hat. Dazu fixieren wir $u \in (0,1)$.

$$\begin{aligned}
P[U \le u] &= \lambda^2 \int_0^\infty e^{-\lambda x} \left(\int_0^\infty 1_{\{\frac{x}{x+y} \le u\}} e^{-\lambda y} dy \right) dx \\
&= \lambda^2 \int_0^\infty e^{-\lambda x} \left(\int_0^\infty 1_{\{x(u^{-1}-1) \le y\}} e^{-\lambda y} dy \right) dx \\
&= \lambda \int_0^\infty e^{-\lambda x} \left(\int_{x(u^{-1}-1)}^\infty \lambda e^{-\lambda y} dy \right) dx \\
&= \lambda \int_0^\infty e^{-\lambda x} e^{-\lambda x(u^{-1}-1)} dx \\
&= \lambda \int_0^\infty e^{-\lambda u^{-1} x} dx \\
&= u.
\end{aligned}$$

Das bedeuted U ist uniform auf $(0,1)$ verteilt und somit ist $f_U(u) = 1_{\{0 \le u \le 1\}}$.

b) **(3 Punkte)** Berechnen Sie die Dichtefunktion f_V und Verteilungsfunktion F_V der Zufallsvariable V.

Solution:

Für $v > 0$,

Siehe nächste Seite!

$$
\begin{aligned}
P[V \leq v] &= \lambda^2 \int_0^\infty e^{-\lambda x} \left(\int_0^\infty 1_{\{x+y \leq v\}} e^{-\lambda y} dy \right) dx \\
&= \lambda^2 \int_0^\infty e^{-\lambda x} 1_{\{x \leq v\}} \left(\int_0^\infty 1_{\{y \leq v-x\}} e^{-\lambda y} dy \right) dx \\
&= \lambda \int_0^v e^{-\lambda x} \left(\int_0^{v-x} \lambda e^{-\lambda y} dy \right) dx \\
&= \lambda \int_0^v e^{-\lambda x} \left(1 - e^{-\lambda(v-x)}\right) dx \\
&= 1 - e^{-\lambda v} - \lambda v e^{-\lambda v}.
\end{aligned}
$$

Nachdem wir nun $P[V \leq v]$ nach v abgeleitet haben (siehe S. 86 im Skript) erhalten wir $f_V(v) = \lambda^2 v e^{-\lambda v} 1_{\{v>0\}}$.

c) (4 Punkte) Bestimmen Sie die gemeinsame Dichtefunktion $f_{(U,V)}$. Sind U und V unabhängig?

$$
\begin{aligned}
P[U \leq u,\, V \leq v] &= \lambda^2 \int_0^\infty e^{-\lambda x} \left(\int_0^\infty 1_{\{\frac{x}{x+y} \leq u,\, x+y \leq v\}} e^{-\lambda y} dy \right) dx \\
&= \lambda^2 \int_0^\infty e^{-\lambda x} \left(\int_0^\infty 1_{\{x(u^{-1}-1) \leq y \leq v-x,\, x \leq uv\}} e^{-\lambda y} dy \right) dx \\
&= \lambda \int_0^{uv} e^{-\lambda x} \left(\int_{x(u^{-1}-1)}^{v-x} \lambda e^{-\lambda y} dy \right) dx \\
&= \lambda \int_0^{uv} e^{-\lambda x} \left(e^{-\lambda x(u^{-1}-1)} - e^{-\lambda(v-x)} \right) dx \\
&= \lambda \int_0^{uv} \left(e^{-\lambda u^{-1} x} - e^{-\lambda v} \right) dx \\
&= u \left(1 - e^{-\lambda v} - \lambda v e^{-\lambda v}\right).
\end{aligned}
$$

Dies stimmt aber genau mit dem Produkt der Dichtefunktionnen von U und V überein, sodass wir sofort schlussfolgern, dass U und V unabhängig sind und

$$
f_{(U,V)}(u,v) = f_U(u) f_V(v) = \lambda^2 v e^{-\lambda v} 1_{\{0 \leq u \leq 1,\, v>0\}}.
$$

Siehe nächste Seite!

4. (**10 Punkte**) Ein pharmazeutisches Unternehmen behauptet, dass die mittlere Konzentration eines gewissen aktiven Wirkstoffes pro Tablette $40\,mg$ sei. Die Standardabweichung ist bekannt und beträgt $\sigma = 3\,mg$. Das Gesundheitsamt möchte überprüfen, dass die Angabe von $40\,mg$ im Mittel stimmt, und misst dazu den Inhalt dieses Wirkstoffes in zehn verschiedenen Tabletten, mit folgenden Ergebnissen:

Messung i	1	2	3	4	5	6	7	8	9	10
$mg\ x_i$	41.42	39.52	41.12	42.78	42.00	39.34	41.84	40.62	40.78	41.44

$$\sum_{i=1}^{10} x_i = 410.86$$

Nehmen Sie an, dass die gemessenen Werte unabhängig und normalverteilt sind.

a) Führen Sie einen Test zum Signifikanzniveau $\alpha = 5\%$ durch, um festzustellen, ob die mittlere Konzentration tatsächlich $40\,mg$ pro Tablette ist.

Geben Sie folgendes an:

1. (**1 Punkt**) das Modell,
   Lösung:
   $X_1, \dots, X_{10}$ unter P_ϑ unabhängig und normalverteilt sind mit unbekannte Erwartungswert μ und bekannte $\sigma = 3\,mg$.
2. (**1 Punkt**)die Nullhypothese,
   Lösung:
   H_0: $\mu = 40$.
3. (**1 Punkt**)die Alternativhypothese,
   Lösung:
   $H_A : \mu \neq 40$.
4. (**1 Punkt**)die Teststatistik,
   Lösung:
   $$T = \frac{\bar{X}_{10} - 40}{3/\sqrt{10}}$$
5. (**1 Punkt**) die Verteilung der Teststatistik unter der Nullhypothese,
   Lösung:
   Unter H_0, T ist standardnormalverteilt.
6. (**1 Punkt**) den Verwerfungsbereich,
   Lösung:
   H_0 wird verworfen falls $|T| > 1.96$. Unter H_0, $P_{H_0}[|T| > 1.96] \approx 0.05$.
7. (**1 Punkt**) den beobachteten Wert der Teststatistik, sowie
   Lösung:
   $\bar{x}_{10} = 41.086$, und $t = 1.14$

Siehe nächste Seite!

8. **(1 Punkt)** den Testentscheid.
   Lösung:
   Wir können nicht H_0 verwerfen.

Hinweis: Sie dürfen folgende Approximationen verwenden:

$$\frac{\sqrt{10}}{3} \approx 1.05 \qquad \frac{10}{3} \approx 3.33 \qquad \frac{3}{\sqrt{10}} \approx 0.95$$

b) **(2 Punkte)** Wie hoch ist das kleinste Niveau, bei welchem die Nullhypothese verworfen wird (dies ist der so genannte P-Wert)?

Lösung:

$P_{H_0}[|T| > 1.14] = 2P_{H_0}[T > 1.14] = 2(1-\Phi(1.14)) = 2(1-0.8729) \approx 0.254.$

Viel Erfolg!