Universität Karlsruhe (TH)
Institut für Stochastik
Dr. Bernhard Klar
Dipl.-Math. oec. Volker Baumstark

Name

Vorname

Matr.-Nr.:

Diplomvorprüfung Stochastik

09. März 2007

Diese Klausur hat bestanden, wer mindestens 16 Punkte erreicht.

Als Hilfsmittel sind nur zwei selbst erstellte DIN A4 Seiten sowie ein nicht programmierbarer Taschenrechner zugelassen!

Aufgabe	1 (7)	2 (7)	3 (6)	4 (6)	5 (7)	6 (7)	$\sum$ (40)
Punkte							
Korrektor							

Gesamtpunktzahl	Note

Aufgabe 1 (7 Punkte)

Ein Zufallszahlengenerator wählt in unabhängiger Folge jede der Ziffern $0, 1, 2, 3$ mit gleicher Wahrscheinlichkeit $1/4$ aus. Die Zufallsvariable Z_k bezeichne die beim k–ten Mal ausgewählte Ziffer ($k = 1, 2, \ldots$). Die Zufallsvariable $X_{n,j}$ gebe an, wie häufig die Ziffer j unter den ersten n erzeugten Zahlen auftritt, es sei also

$$X_{n,j} := \sum_{k=1}^{n} \mathbf{1}\{Z_k = j\} \qquad (j = 0, 1, 2, 3).$$

a) Bestimmen Sie $E(Z_1)$, $V(Z_1)$ und $\rho(Z_1 - 2Z_2, 2Z_2 - Z_3)$.

b) Welche Verteilung besitzt der Zufallsvektor $(X_{n,0}, X_{n,1}, X_{n,2}, X_{n,3})$?

Lösung: $\boxed{}$ (keine Begründung erforderlich)

c) Welche Verteilung besitzt $X_{n,3}$?

Lösung: $\boxed{}$ (keine Begründung erforderlich)

d) Welche Verteilung besitzt $X_{n,0} + X_{n,3}$?

Lösung: $\boxed{}$ (keine Begründung erforderlich)

e) Berechnen Sie die Wahrscheinlichkeit, dass unter 5 erzeugten Zahlen 3 Nullen und 2 Einsen vorkommen.

f) Berechnen Sie mit Hilfe des zentralen Grenzwertsatzes approximativ die Wahrscheinlichkeit, dass unter 800 erzeugten Zahlen mehr als 230 Einsen vorkommen.

Einige Werte der Verteilungsfunktion Φ der Normalverteilung $N(0, 1)$:

t	0	1	1.73	2	2.45	3
$\Phi(t)$	0.5	0.84	0.96	0.98	0.993	0.999

Lösung zu Aufgabe 1

a) Es gelten

$$\begin{aligned} EZ_1 &= \frac{1}{4}(1+2+3) = \frac{3}{2}, \qquad E(Z_1^2) = \frac{1}{4}(1+4+9) = \frac{7}{2}, \\ V(Z_1) &= \frac{14}{4} - \frac{9}{4} = \frac{5}{4} \end{aligned}$$

2.5 P

und

$$\begin{aligned} \rho(Z_1 - 2Z_2, 2Z_2 - Z_3) &= \frac{2C(Z_1, Z_2) - C(Z_1, Z_3) - 4C(Z_2, Z_2) + 2C(Z_2, Z_3)}{\sqrt{V(Z_1) + 4V(Z_2)) \cdot (4V(Z_2) + V(Z_3))}} \\ &= \frac{-4V(Z_2)}{5V(Z_1)} = -\frac{4}{5}. \end{aligned}$$

b)

$$\mathrm{Mult}\left(n; \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right).$$

$\boxed{0.5 \text{ P}}$

c)

$$\mathrm{Bin}\left(n, \frac{1}{4}\right).$$

$\boxed{0.5 \text{ P}}$

d)

$$\mathrm{Bin}\left(n, \frac{1}{2}\right).$$

$\boxed{0.5 \text{ P}}$

e)

$$P(X_{5,0} = 3,\, X_{5,1} = 2,\, X_{5,2} = X_{5,3} = 0) = \frac{5!}{3! \cdot 2! \cdot 0! \cdot 0!} \left(\frac{1}{4}\right)^5 = \frac{10}{4^5} \; (\approx 0.0098).$$

$\boxed{1 \text{ P}}$

f) Nach dem ZGWS von de Moivre-Laplace kann man hier die $\mathrm{Bin}(800, \frac{1}{4})$-Verteilung durch die Normalverteilung mit Parameter $\mu = np = 200$ und $\sigma^2 = np(1-p) = 150$ approximieren. Somit gilt $\boxed{2 \text{ P}}$

$$\begin{aligned} P(X_{800,1} > 230) &= 1 - P(X_{800,1} \leq 230) \approx 1 - \Phi\left(\frac{230 - 200}{5\sqrt{6}}\right) \\ &= 1 - \Phi(\sqrt{6}) = 1 - 0.993 = 0.007. \end{aligned}$$

Aufgabe 2 (7 Punkte)
Tom will im Karlsruher Zoo zur Eisbärenanlage. Er kann hierfür nach rechts (richtiger Weg) oder nach links (falscher Weg) gehen. Fragt er einen Besucher des Zoos mit Tageskarte nach dem Weg dorthin, so erhält er mit Wahrscheinlichkeit 2/3 die richtige Antwort und mit Wahrscheinlichkeit 1/3 eine falsche Antwort. Fragt er einen Dauerkartenbesitzer nach dem Weg dorthin, so erhält er stets die richtige Antwort. Antworten und Eintrittskarten von verschiedenen Personen sind unabhängig. Die Wahrscheinlichkeit, dass eine zufällig angesprochene Person eine Dauerkarte besitzt, sei 1/10.

a) Zeichnen Sie das zu dem 2-stufigen Experiment gehörende Baumdiagramm, und tragen Sie die Start- und Übergangswahrscheinlichkeiten ein.

b) Tom fragt einen Besucher $B1$ nach dem Weg zur Eisbärenanlage. Mit welcher Wahrscheinlichkeit erhält er eine richtige Antwort?

c) Wie groß ist die bedingte Wahrscheinlichkeit, dass der Besucher $B1$ eine Dauerkarte besitzt, wenn er die richtige Antwort gegeben hat?

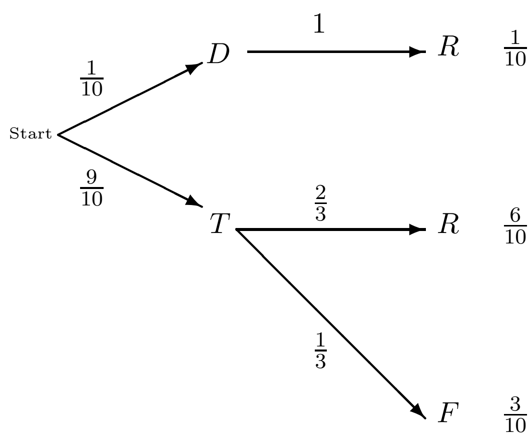
d) Tom fragt einen weiteren Besucher $B2$ nach dem Weg zur Eisbärenanlage. Mit welcher Wahrscheinlichkeit geben $B1$ und $B2$ dieselbe Antwort?

e) Wie groß ist die bedingte Wahrscheinlichkeit, dass der Besucher $B1$ eine Tageskarte besitzt, wenn $B1$ und $B2$ dieselbe Antwort geben?

Lösung zu Aufgabe 2

a) $D :=$ Dauerkarte, $T :=$ Tageskarte, $R :=$ Antwort richtig, $F :=$ Antwort falsch,

1 P



b) Der Index 1 beziehe sich auf $B1$. Aus obigem Diagramm lesen wir ab

$$P(R_1) = P(D_1 \cap R_1) + P(T_1 \cap R_1) = \frac{1}{10} + \frac{6}{10} = \frac{7}{10}.$$ 1 P

c) Unser Diagramm und Teil a) liefern

1 P

$$P(D_1|R_1) = \frac{P(D_1 \cap R_1)}{P(R_1)} = \frac{1/10}{7/10} = \frac{1}{7}.$$

d) Der Index 2 beziehe sich auf $B2$. Es sei C das Ereignis, dass $B1$ und $B2$ dieselbe Antwort geben. Analog zu a) gilt

$$P(R_2) = \frac{7}{10}.$$

1.5 P

Die gesuchte Wahrscheinlichkeit berechnet sich wegen der Unabhängigkeit von R_1 und R_2 bzw. F_1 und F_2 zu

$$\begin{aligned} P(C) &= P(R_1 \cap R_2) + P(F_1 \cap F_2) \\ &= P(R_1)P(R_2) + P(F_1)P(F_2) = \left(\frac{7}{10}\right)^2 + \left(\frac{3}{10}\right)^2 = \frac{49+9}{100} = \frac{58}{100}. \end{aligned}$$

e) Es gelten unter Beachtung des Teils a)

$$P(T_1 \cap R_1 \cap R_2) = P(T_1 \cap R_1)P(R_2) = \frac{6}{10} \cdot \frac{7}{10} = \frac{42}{100}$$

2.5 P

und

$$P(T_1 \cap F_1 \cap F_2) = P(T_1 \cap F_1)P(F_2) = \frac{3}{10} \cdot \frac{3}{10} = \frac{9}{100}.$$

Folglich

$$P(T_1 \cap C) = P(T_1 \cap R_1 \cap R_2) + P(T_1 \cap F_1 \cap F_2) = \frac{42}{100} + \frac{9}{100} = \frac{51}{100}.$$

Die gesuchte Wahrscheinlichkeit berechnet sich mit c) zu

$$P(T_1|C) = \frac{P(T_1 \cap C)}{P(C)} = \frac{51}{58}.$$

Aufgabe 3 (6 Punkte)
Sei X exponentialverteilt mit Parameter $\lambda > 0$, d.h. X besitze die Dichte

$$f(x) = \lambda \mathrm{e}^{-\lambda x}, \qquad x > 0.$$

a) Bestimmen Sie den Median von X.

b) Berechnen Sie die Verteilungsfunktion und Dichte der Zufallsvariablen $Y := e^{\alpha X}$, $\alpha > 0$.

c) Bestimmen Sie alle $\alpha > 0$, für die der Erwartungswert von Y existiert. Berechnen Sie EY für alle derartigen α.

Lösung zu Aufgabe 3

a) Die Verteilungsfunktion F von X ist gegeben durch

$$F(x) = 1 - \mathrm{e}^{-\lambda x}, \qquad x > 0,$$ 1.5 P

und $F(x) = 0$ für $x \leq 0$. Wegen

$$F(x) = \frac{1}{2} \quad \Leftrightarrow \quad \mathrm{e}^{-\lambda x} = \frac{1}{2} \quad \Leftrightarrow \quad \lambda x = -\log\frac{1}{2} \quad \Leftrightarrow \quad x = \frac{1}{\lambda} \cdot \log 2$$

berechnet sich der Median von X zu $\dfrac{\log 2}{\lambda}$.

b) Wegen $X > 0$ ist $Y = e^{\alpha X} > 1$. Für $t > 1$ gilt

$$\begin{aligned} P(Y \leq t) &= P(\alpha X \leq \log t) = P\left(X \leq \frac{\log t}{\alpha}\right) \\ &= 1 - \exp\left(-\lambda \cdot \frac{\log t}{\alpha}\right) \\ &= 1 - t^{-\lambda/\alpha}. \end{aligned}$$ 2.5 P

Eine Dichte für Y ist folglich durch

$$f(t) := \frac{dP(Y \leq t)}{dt} \cdot \mathbf{1}\{t > 1\} = \frac{\lambda}{\alpha} \cdot t^{-\lambda/\alpha - 1} \cdot \mathbf{1}\{t > 1\}$$

gegeben.

c) Nach der Transformationsformel für Erwartungswerte gilt

$$EY = \int_0^\infty e^{\alpha x} \lambda e^{-\lambda x} dx \ = \int_0^\infty \lambda e^{-(\lambda - \alpha)x} dx \ = \begin{cases} \infty & \text{für } \alpha \geq \lambda \\ \frac{\lambda}{\lambda - \alpha} & \text{für } \alpha < \lambda. \end{cases}$$ 2 P

Der Erwartungswert existiert also für $\alpha < \lambda$ und berechnet sich in diesem Fall zu $\lambda/(\lambda - \alpha)$.

Aufgabe 4 (6 Punkte)
Es sei (U, V) ein zweidimensionaler Zufallsvektor mit Dichte

$$f(u,v) = \frac{u}{v^2}\mathrm{e}^{-u/v}\,\mathbf{1}_{\{u>0,0<v<1\}}.$$

a) Zeigen Sie, dass $U \sim \mathrm{Exp}(1)$ und $V \sim \mathrm{U}(0,1)$ gilt.

b) Berechnen Sie die Kovarianz $\mathrm{C}(U,V)$ zwischen U und V.

Hinweis: Sie können ohne Beweis $\int_0^\infty x^2\mathrm{e}^{-x}dx = 2$ verwenden.

c) Sind U und V unabhängig?

Lösung zu Aufgabe 4

a) Die Dichte f_V von V ergibt sich durch

$$f_V(v) = \int_0^\infty \frac{u}{v^2}\mathrm{e}^{-\frac{u}{v}}\mathrm{d}u = \int_0^\infty x\mathrm{e}^{-x}\mathrm{d}x = 1, \qquad v \in (0,1),$$ 2 P

und $f_V(v) = 0$ sonst. Somit ist V auf $(0,1)$ gleichverteilt.

Ferner gilt für $u > 0$

$$f_U(u) = \int_0^1 \frac{u}{v^2}\mathrm{e}^{-\frac{u}{v}}\mathrm{d}v = \int_u^\infty \mathrm{e}^{-x}\mathrm{d}x = \mathrm{e}^{-u}.$$

Somit ist U exponentialverteilt mit Parameter 1.

b) Für $U \sim \mathrm{Exp}(1)$ bzw. $W \sim \mathrm{U}(0,1)$ gilt

$$EU = 1, \qquad EV = \frac{1}{2}.$$ 3 P

Ferner ist

$$\begin{aligned} E(UV) &= \int_0^1\int_0^\infty uvf(u,v)\mathrm{d}u\mathrm{d}v = \int_0^1\int_0^\infty \frac{u^2}{v}\mathrm{e}^{-u/v}\mathrm{d}u\mathrm{d}v \\ &= \int_0^1 v^2\int_0^\infty x^2\mathrm{e}^{-x}\mathrm{d}x\mathrm{d}v = 2\int_0^1 v^2\mathrm{d}v = \frac{2}{3}. \end{aligned}$$

Somit

$$\mathrm{C}(U,V) = EUV - EUEV = \frac{2}{3} - \frac{1}{2} = \frac{1}{6}.$$

c) Aus $\mathrm{C}(U,V) \neq 0$ folgt, dass U und V nicht unabhängig sind.

1 P

Aufgabe 5 (7 Punkte)
Die Zufallsvariablen $X_1, \ldots, X_n$ seien unabhängig und identisch verteilt mit

$$P(X_1 = k) = \binom{k+r-1}{k} \left(\frac{r}{r+\mu}\right)^r \left(\frac{\mu}{r+\mu}\right)^k, \qquad k \in \mathbb{N}_0.$$

Dabei ist $r \in \mathbb{N}$ bekannt und $\mu > 0$ unbekannt.

a) Zeigen Sie, dass

$$\hat{\mu}_n = \frac{1}{n}\sum_{i=1}^n X_i$$

der Maximum-Likelihood-Schätzer für μ ist, falls $\sum_{i=1}^n X_i > 0$ ist.

b) Ist $\hat{\mu}_n$ erwartungstreu für μ?

c) Berechnen Sie die Varianz von $\hat{\mu}_n$ und die mittlere quadratische Abweichung

$$MQA_{\hat{\mu}_n}(\mu) = E_\mu(\hat{\mu}_n - \mu)^2.$$

d) Ist die Folge $(\hat{\mu}_n)_{n\in\mathbb{N}}$ konsistent für μ?

Hinweis: Sie können ohne Beweis $E_\mu X_1 = \mu$ und $V_\mu(X_1) = \dfrac{\mu(r+\mu)}{r}$ benutzen.

Lösung zu Aufgabe 5

a) Für $x = (x_1, \ldots, x_n)$, wobei $x_1, \ldots, x_n \in \mathbb{N}_0$, gilt:

$$\begin{aligned} L_x(\mu) &= \prod_{i=1}^n \binom{x_i+r-1}{x_i} \left(\frac{r}{r+\mu}\right)^r \left(\frac{\mu}{r+\mu}\right)^{x_i} \\ &= \left(\frac{r}{r+\mu}\right)^{nr} \left(\frac{\mu}{r+\mu}\right)^{\sum_{i=1}^n x_i} \underbrace{\prod_{i=1}^n \binom{x_i+r-1}{x_i}}_{=:c}. \end{aligned}$$ $\boxed{4\text{ P}}$

Folglich

$$\log L_x(\mu) = nr(\log r - \log(\mu+r)) + \sum_{i=1}^n x_i(\log\mu - \log(r+\mu)) + \log c$$

und

$$\begin{aligned} \frac{d}{d\mu}\log L_x(\mu) &= -\frac{nr}{\mu+r} + \sum_{i=1}^n \overbrace{\left(\frac{1}{\mu} - \frac{1}{\mu+r}\right)}^{=\frac{r}{\mu(\mu+r)}} \cdot x_i \\ &= \frac{nr}{\mu+r}\left(-1 + \bar{x}_n \cdot \frac{1}{\mu}\right) = 0 \quad \Leftrightarrow \quad \mu = \bar{x}_n, \end{aligned}$$

sowie

$$\frac{d^2}{d^2\mu}\log L_x(\bar{x}_n) = \frac{nr}{\bar{x}_n + r} \cdot \left(-\frac{1}{\bar{x}_n}\right) < 0,$$

falls $\bar{x}_n > 0$. Folglich ist

$$\hat{\mu}_n := \bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

der gesuchte Maximum-Likelihood-Schätzer, falls $\sum_{i=1}^{n} X_i > 0$ ist.

b) Wegen $E_\mu \bar{X}_n = \mu$ ist $\hat{\mu}_n$ erwartungstreu.

c) Die mittlere quadratische Abweichung stimmt wegen der Erwartungstreue von $\hat{\mu}_n$ mit $V_\mu(\hat{\mu}_n)$ übereinn und es gilt $\boxed{1\text{ P}}$

$$\mathrm{MAQ}_{\hat{\mu}_n}(\mu) = V_\mu(\hat{\mu}_n) = \frac{V_\mu(X_1)}{n} = \frac{\mu(r+\mu)}{nr}. \qquad \boxed{1\text{ P}}$$

d) Die Folge $(\hat{\mu}_n)_{n\in\mathbb{N}}$ ist konsistent: Für $\varepsilon > 0$ folgt aus der Tschebyscheff-Ungleichung

$$P_\mu(|\hat{\mu}_n - \mu| \geq \varepsilon) \leq \frac{V_\mu(\hat{\mu}_n)}{\varepsilon^2} = \frac{\mu(r+\mu)}{nr\varepsilon^2} \to 0 \qquad \boxed{1\text{ P}}$$

für $n \to \infty$.

Alternativ folgt die Konsistenz direkt aus dem Schwachen Gesetz großer Zahlen.

Aufgabe 6 (7 Punkte)

In einer Studie soll gezeigt werden, dass bei trainierten Personen der mittlere systolische Blutdruck niedriger als der Normalwert von 80mmHG ist. Dazu wird bei n Sportlern der Blutdruck gemessen, wobei wir annehmen, dass die Blutdruckwerte als Realisierungen unabhängiger Zufallsvariablen mit gleicher Normalverteilung $N(\mu, 16)$ modelliert werden können.

a) Formulieren Sie eine geeignete Hypothese und Alternative und geben Sie ein Testverfahren basierend auf der Testgröße $T_n = \sqrt{n}(\bar{X}_n - 80)/4$ an, das geeignet ist, die obige Behauptung „statistisch nachzuweisen“.

Wie muss dabei der kritische Wert gewählt werden, wenn der Stichprobenumfang $n = 30$ und das Signifikanzniveau $\alpha = 1\%$ betragen soll?

b) Was ist das Testergebnis, wenn bei den 30 Sportlern $\bar{x}_{30}$ =78mmHG gemessen wurde.

c) Bestimmen Sie die Gütefunktion des Tests.

d) Wie groß ist die Wahrscheinlichkeit des Fehlers 2. Art, wenn im Falle $n = 30$ der mittlere systolische Blutdruck in Wirklichkeit 78mmHG beträgt?

e) Was könnte der Leiter der Studie tun, um die Fehlerwahrscheinlichkeit in d) zu verringern?

Einige Werte der Verteilungsfunktion Φ der Normalverteilung $N(0, 1)$:

t	0	0.41	1.28	1.65	1.96	2.33	2.58
$\Phi(t)$	0.5	0.66	0.9	0.95	0.975	0.99	0.995

Lösung zu Aufgabe 6

a) Man muss als Hypothese $H_0 : \vartheta \geq 80$ und als Alternative $H_1 : \vartheta < 80$ wählen. Nur dann ist bei Ablehnung von H_0 die Aussage “bei trainierten Personen ist der mittlere systolische Blutdruck niedriger als 80mmHG“ statistisch gesichert. $\boxed{2\text{ P}}$

Da kleine Werte von T_n gegen H_0 sprechen, und da T_n unter H_0 standardnormalverteilt ist, muss H_0 zum Niveau $\alpha = 0.01$ verworfen werden, wenn $T_n \leq z_{0.01} = \Phi^{-1}(0.01) = -2.33$ ist. Andernfalls besteht kein Einwand gegen H_0.

b) Wegen $T_{30} = \sqrt{30}(\bar{X}_{30} - 80)/4 = -2.74 < z_{0.01}$ wird H_0 wird auf dem 1%–Niveau abgelehnt.

c) Die Gütefunktion des Tests ist $\boxed{1\text{ P}}$

$$\begin{aligned} g(\mu) &= P_\mu(\text{Test verwirft } H_0) = P_\mu\left(\frac{\sqrt{n}(\bar{X}_n - 80)}{4} \leq z_{0.01}\right) \\ &= P_\mu\left(\frac{\sqrt{n}(\bar{X}_n - \mu)}{4} \leq z_{0.01} + \frac{\sqrt{n}(80 - \mu)}{4}\right) \\ &= \Phi\left(z_{0.01} + \frac{\sqrt{n}(80 - \mu)}{4}\right). \end{aligned}$$ $\boxed{2\text{ P}}$

d) Für $\mu = 78$ ist die Wahrscheinlichkeit des Fehlers 2. Art

$$\beta = 1 - g(78) = 1 - \Phi(-2.33 + 2.74) = 1 - \Phi(0.41) = 0.34.$$ $\boxed{1\text{ P}}$

e) Der Studienleiter könnte den Stichprobenumfang erhöhen, denn dann nimmt die Güte des Tests zu, es verringert sich also die Wahrscheinlichkeit, dass bei Vorliegen der Alternative die Hypothese H_0 fälschlicherweise nicht abgelehnt wird. $\boxed{1\text{ P}}$

Eine weitere Möglichkeit wäre, den Fehler 1. Art anzuheben (etwa auf $\alpha = 5\%$).