Technische Universität München
Institut für Informatik
Prof. Dr. Javier Esparza Sommersemester 2007

Diskrete Wahrscheinlichkeitstheorie
Wiederholungsklausur

LÖSUNG

***Hinweis:* Bei allen Aufgaben wird neben dem gefragten Ergebnis eine kurze Begründung erwartet.**

Aufgabe 1 — 2+2+2+2 P

(a) Geben Sie die erzeugende Funktion der Zufallsvariablen X mit $W_X = \{0, 1, 2\}$ und $\Pr[X = 0] = \frac{1}{4}$, $\Pr[X = 1] = \frac{1}{3}$ an.

(b) Für die diskrete Zufallsvariable X gelte $W_X = \mathbb{N} = \{0, 1, 2, \ldots\}$. X sei gedächtnislos mit Erwartungswert $\mathbb{E}[X] = 3$. Geben Sie $\mathrm{Var}[X]$ an.

(c) Es sei X eine Zufallsvariable mit parametrisierter Dichte $f_X(x; \theta)$. U sei ein erwartungstreuer Schätzer für θ mit $\mathbb{E}[3U + 2] = 3$. Geben Sie θ an.

(d) Es seien A, B Ereignisse im diskreten Wahrscheinlichkeitsraum $(\Omega, \Pr)$ mit $\Pr[A], \Pr[B] > 0$. $\overline{A} := \Omega \setminus A$ bezeichne das Komplementereignis zu A. $\overline{A}$ und B seien unabhängig. Weiterhin gelte $\Pr[A|B] = \frac{1}{3}$. Bestimmen Sie $\Pr[\overline{A}]$.

Antwort:

(a)

$$G_X(z) = \frac{1}{4} + \frac{1}{3}z + \frac{5}{12}z^2.$$

(b) Da X gedächtnislos mit Erwartungswert 3 ist, muss $X \sim \mathrm{Geo}(\frac{1}{3})$ gelten. Damit folgt $\mathrm{Var}[X] = \frac{q}{p^2} = \frac{2/3}{1/9} = 6$.

(c) Da U erwartungstreu ist, gilt $\mathbb{E}[U] = \theta$. Wegen $3 = \mathbb{E}[3U + 2] = 3\mathbb{E}[U] + 2 = 3\theta + 2$ folgt $\theta = \frac{1}{3}$.

(d) Da $\overline{A}$ und B unabhängig sind, sind auch A und B unabhängig. Daher folgt $\Pr[A|B] = \Pr[A]$ und somit $\Pr[A] = \frac{1}{3}$. Es ergibt sich daher $\Pr[\overline{A}] = \frac{2}{3}$.

Aufgabe 2 — 2+3+3 P

Sie sind – zumindest für diese Aufgabe – Besitzer eines Hauses in Kalifornien. Da Sie sich vor Einbrechern schützen wollen, haben Sie Ihr Haus mit einer Alarmanlage ausstatten lassen. Die Alarmanlage kann entweder durch einen Einbruch oder durch ein Erdbeben ausgelöst werden.

A bezeichne das Ereignis, dass der *Alarm* aktiviert wird. B bezeichne das Ereignis, dass die Erde *bebt*. Schließlich sei mit E das Ereignis bezeichnet, dass in das Haus *eingebrochen* wird.

Mit Wahrscheinlichkeit $\Pr[B] = 0.02$ bebt die Erde, mit Wahrscheinlichkeit $\Pr[E] = 0.01$ findet ein Einbruch statt. Wir nehmen an, dass B und E unabhängig sind.

Für die Aktivierung des Alarms gilt weiter

$$\Pr[A|B \cap E] = 0.95,\ \Pr[A|\overline{B} \cap E] = 0.94,\ \Pr[A|B \cap \overline{E}] = 0.29,\ \Pr[A|\overline{B} \cap \overline{E}] = 0.01.$$

(a) Bestimmen Sie die Wahrscheinlichkeit $\Pr[A]$, dass der Alarm ausgelöst wird.

(b) Berechnen Sie die Wahrscheinlichkeit $\Pr[E|A]$, dass ein Einbruch vorliegt, falls der Alarm ausgelöst wird.

(c) Da Sie auf Grund einer Klausur außer Haus sind, bitten Sie Ihren Nachbarn Arnold, Sie anzurufen, falls er den Alarm hört. Nun ist Ihr Nachbar Arnold auch noch in der Politik tätig, und ruft Sie daher auch gelegentlich an, um Sie über den Vorteil von mit Biodiesel betriebenen 20l-Autos aufzuklären.

Bezeichnet T das Ereignis, dass Arnold Sie per *Telefon* anruft, so gilt

$$\Pr[T|A] = 0.9.$$

Bestimmen Sie die Wahrscheinlichkeit $\Pr[E \cap A \cap T]$.

Hinweis: Da Arnold nur zwischen Alarm und kein Alarm unterscheiden kann, nehmen wir an, dass $\Pr[T|A \cap E] = \Pr[T|A \cap \overline{E}] = \Pr[T|A]$ gilt.

Antwort:

(a) Es gilt

$$\begin{aligned}
\Pr[A] &= \left(\Pr[A|B,E]\cdot\Pr[B] + \Pr[A|\overline{B},E]\cdot\Pr[\overline{B}]\right)\cdot\Pr[E] \;+ \left(\Pr[A|B,\overline{E}]\cdot\Pr[B] + \Pr[A|\overline{B},\overline{E}]\cdot\Pr[\overline{B}]\right)\cdot\Pr[\overline{E}]\\
&= 0.00019 + 0.009212 + 0.005742 + 0.009702\\
&= 0.024846.
\end{aligned}$$

(b)

$$\begin{aligned}
\Pr[E|A] &= \frac{\Pr[A\cap E]}{\Pr[A]}\\
&= \frac{\Pr[A\cap E\cap B] + \Pr[A\cap E\cap \overline{B}]}{\Pr[A]}\\
&= \frac{\Pr[A|E,B]\cdot\Pr[E]\cdot\Pr[B] + \Pr[A|E,\overline{B}]\cdot\Pr[E]\cdot\Pr[\overline{B}]}{\Pr[A]}\\
&= \left(\Pr[A|E,B]\cdot\Pr[B] + \Pr[A|E,\overline{B}]\cdot\Pr[\overline{B}]\right)\cdot\frac{\Pr[E]}{\Pr[A]}\\
&= (0.019 + 0.9212)\cdot\frac{0.01}{0.024846}\\
&= \frac{0.009402}{0.024846} \approx 0.378411.
\end{aligned}$$

(c)

$$\begin{aligned}
\Pr[E,A,T] &= \Pr[T|A,E]\cdot\Pr[A|E]\cdot\Pr[E]\\
&= \Pr[T|A]\cdot\Pr[A|E]\cdot\Pr[E]\\
&= \Pr[T|A]\cdot\Pr[E|A]\cdot\frac{\Pr[A]}{\Pr[E]}\cdot\Pr[E]\\
&= \Pr[T|A]\cdot\Pr[E|A]\cdot\Pr[A]\\
&\approx 0.9\cdot 0.378411\cdot 0.024846 = 0.008462.
\end{aligned}$$

Aufgabe 3 — 4+4 P

Die stetigen Zufallsvariablen X_1, X_2 seien unabhängig und identisch gleichverteilt über dem Intervall $[1,3]$. Die X_i besitzen somit jeweils die Dichte

$$f(t) = \begin{cases} \frac{1}{2} & \text{falls } t\in[1,3]\\ 0 & \text{sonst}\end{cases}$$

(a) Zeigen Sie, dass $Y := \max\{X_1, X_2\}$ die Dichte

$$f_Y(t) = \begin{cases} \frac{1}{2} \cdot (t-1) & \text{falls } t \in [1,3] \\ 0 & \text{sonst} \end{cases}$$

besitzt.

Bestimmen Sie hierfür zunächst die Verteilungsfunktion von Y!

(b) Bestimmen Sie den Erwartungswert und die Varianz von Y.

Antwort:

(a)

$$F_Y(t) = \Pr[Y \le t] = \Pr[X_1 \le t, X_2 \le t, \ldots, X_n \le t] = \prod_{i=1}^{n} \Pr[X_i \le t] = F(t)^n$$

Es folgt

$$f_Y(t) = F_Y'(t) = n \cdot F(t)^{n-1} \cdot f(t) \overset{\text{für } t \in [a,b]}{=} n \cdot \left(\frac{t-a}{b-a}\right)^{n-1} \cdot \frac{1}{b-a} = n \cdot \frac{(t-a)^{n-1}}{(b-a)^n} = \frac{1}{2} \cdot (t-1).$$

(b)

$$\begin{aligned}
\mathbb{E}[Y] &= \int_{t=a}^{b} t \cdot n \cdot \frac{(t-a)^{n-1}}{(b-a)^n} \cdot dt \\
&= \frac{n}{(b-a)^n} \cdot \int_{t=0}^{b-a} t^n + a \cdot t^{n-1} \cdot dt \\
&= \frac{n}{n+1}(b-a) + a = \frac{7}{3}
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}[Y^2] &= \int_{t=a}^{b} t^2 \cdot n \cdot \frac{(t-a)^{n-1}}{(b-a)^n} \cdot dt \\
&= \frac{n}{(b-a)^n} \cdot \int_{t=0}^{b-a} (t+a)^2 \cdot t^{n-1} \cdot dt \\
&= \frac{n}{(b-a)^n} \cdot \int_{t=0}^{b-a} t^{n+1} + 2a \cdot t^n + a^2 \cdot t^{n-1} \cdot dt \\
&= \frac{n}{n+2}(b-a)^2 + 2a \cdot \frac{n}{n+1}(b-a) + a^2 = 2 + \frac{8}{3} + 1 = \frac{17}{3} \\
\mathrm{Var}[Y] &= \mathbb{E}[Y^2] - \mathbb{E}[Y]^2 = \frac{17}{3} - \frac{49}{9} = \frac{51-49}{9} = \frac{2}{9}.
\end{aligned}$$

Aufgabe 4 **4+4 P**

Sie sind mit der Organisation einer Klausur beauftragt worden. Für die Klausur sind $n = 100$ Studenten angemeldet. Um einen passenden Hörsaal zu finden, wollen Sie die Zahl N der Studenten, die tatsächlich zur Prüfung erscheinen, abschätzen.

Sie modellieren daher jeden der n Studenten durch eine binomial-verteilte Zufallsvariable X_i mit $X_i \sim \mathrm{Bin}(1, 0.9)$, wobei $\Pr[X_i = 1] = 0.9$ die Wahrscheinlichkeit ist, dass Student i zur Klausur kommt. Weiterhin nehmen Sie an, dass jeder Student unabhängig von seinen Komilitonen zur Klausur erscheint.

Die Anzahl Y der Studenten, die zur Klausur erscheinen, ist dann ebenfalls binomial-verteilt. Es existiert also eine kleinste Zahl N^*, so dass mit mindestens Wahrscheinlichkeit 0.95 höchstens N^* viele Studenten zur Klausur erscheinen.

Da es Ihnen zu aufwendig ist, N^* direkt zu bestimmen, approximieren Sie die Binomialverteilung, um eine möglichst kleine obere Schranke für N^* zu finden.

(a) Der Einfachheit wegen verwenden Sie zunächst die Markov-Ungleichung, um eine möglichst kleine obere Schranke für N^* zu bestimmen. Welche Zahl erhalten Sie unter Verwendung der Markov-Ungleichung?

(b) Da die Markov-Ungleichung Ihnen nicht wirklich hilft, benutzen Sie die Standardnormalverteilung zur Approximation der Binomialverteilung, um eine möglichst kleine obere Schranke für N^* zu finden. Welche obere Schranke erhalten Sie nun?

Hinweis: $\Phi(1.645) \approx 0.95$.

Antwort:

(a) Setze $Y := \sum_{i=1}^{n} X_i$ mit $\mathbb{E}[Y] = n \cdot p = 90$ und $\mathrm{Var}[Y] = n \cdot p \cdot (1-p) = 9$.

Es soll ein möglichst kleines N_M bestimmt werden, so dass

$$\Pr[Y \leq N_M] \geq 0.95.$$

Mittels der Markov-Ungleichung

$$\Pr[Y \geq t] \leq \frac{\mathbb{E}[Y]}{t}$$

ergibt sich

$$\Pr[Y \leq N_M] = 1 - \Pr[Y > N_M] = 1 - \Pr[Y \geq N_M + 1] \geq 1 - \frac{\mathbb{E}[Y]}{N_M + 1} = 0.95.$$

Damit folgt:

$$N_M = \left\lceil \frac{\mathbb{E}[Y]}{0.05} - 1 \right\rceil = 1799.$$

(b) Nach dem ZGWS (mit Stetigkeitskorrektur) gilt:

$$\Pr[Y \leq N_S] = \Pr\left[\frac{Y - \mathbb{E}[Y]}{\sqrt{\mathrm{Var}[Y]}} \leq \frac{N_S - \mathbb{E}[Y]}{\sqrt{\mathrm{Var}[Y]}}\right] \approx \Phi\left(\frac{N_S(+0.5) - \mathbb{E}[Y]}{\sqrt{\mathrm{Var}[Y]}}\right) = \Phi\left(\frac{N_S(+0.5) - 90}{\sqrt{9}}\right)$$

Laut Hinweis ist $\Phi(1.645) \approx 0.95$, was auf

$$\frac{N_S + (0.5) - 90}{\sqrt{9}} = 1.645$$

und damit auf

$$N_S = \left\lceil \sqrt{9} \cdot 1.645 + 90(-0.5) \right\rceil = 95$$

führt.

Aufgabe 5 — 2+3+3 P

Die Firma S. produziert Fahrradreifen und verspricht, dass ihre durchschnittliche Laufleistung mindestens 8 Mm (=8000 Kilometer) beträgt. Herr M., der seit Jahren die Produkte der Firma S. benutzt, möchte diese Angabe auf Glaubwürdigkeit überprüfen. Aus seinen Aufzeichnungen weiß er, dass die letzten vier Reifenwechsel an seinem Fahrrad nach folgenden Laufleistungen (in Mm) stattgefunden haben: 7, 3, 7, 7. Herr M. bekommt den Verdacht, dass die Reifen nicht so dauerhaft sind, wie die Firma behauptet.

(a) Berechnen Sie das Stichprobenmittel und die Stichprobenvarianz für die Reifen des Herrn M.

(b) Aus seiner Stochastik-Vorlesung kann sich Herr M. an folgende statistische Tests erinnern:

- *Gaußtest*: $X_1, \ldots, X_n$ unabhängig und identisch verteilt mit $X_i \sim \mathcal{N}(\mu, \sigma^2)$, wobei σ^2 bekannt ist.
  Testgröße: $Z := \frac{\bar{X}-\mu_0}{\sigma}\sqrt{n}$.
  Ablehnungskriterium: $|Z| > z_{1-\alpha/2}$ für $H_0: \mu = \mu_0$, $Z < z_\alpha$ für $H_0\colon \mu \geq \mu_0$ bzw. $Z > z_{1-\alpha}$ für $H_0\colon \mu \leq \mu_0$.

- *t-Test*: $X_1, \ldots, X_n$ unabhängig und identisch verteilt mit $X_i \sim \mathcal{N}(\mu, \sigma^2)$.
  Testgröße: $T := \frac{\bar{X}-\mu_0}{S}\sqrt{n}$.
  Ablehnungskriterium: $|T| > t_{n-1,1-\alpha/2}$ für $H_0: \mu = \mu_0$, $T < t_{n-1,\alpha}$ für $H_0\colon \mu \geq \mu_0$ bzw. $T > t_{n-1,1-\alpha}$ für $H_0\colon \mu \leq \mu_0$.

- *Zwei-Stichproben-t-Test*: $X_1, \ldots, X_m$ und $Y_1, \ldots, Y_n$ unabhängig und jeweils identisch verteilt mit gleicher Standardabweichung, d.h. $X_i \sim \mathcal{N}(\mu_X, \sigma^2)$ und $Y_j \sim \mathcal{N}(\mu_Y, \sigma^2)$.
  Testgröße: $T := \sqrt{\frac{n+m-2}{\frac{1}{m}+\frac{1}{n}}} \cdot \frac{\bar{X}-\bar{Y}}{\sqrt{(m-1)\cdot S_X^2+(n-1)\cdot S_Y^2}}$
  Ablehnungskriterium: $|T| > t_{m+n-2,1-\alpha/2}$ für $H_0\colon \mu_X = \mu_Y$, $T < t_{m+n-2,\alpha}$ für $H_0\colon \mu_X \geq \mu_Y$ bzw. $T > t_{m+n-2,1-\alpha}$ für $H_0\colon \mu_X \leq \mu_Y$.

Formulieren Sie die Aussage der Firma als Nullhypothese und wählen Sie unter den oben angegebenen Tests einen geeigneten aus. Begründen Sie für jeden Test, warum er geeignet oder ungeeignet ist.

(c) Führen Sie den von Ihnen in (b) gewählten Test durch, d.h., berechnen Sie die Testgröße und überprüfen Sie, ob das Ablehnungskriterium erfüllt ist. Verwenden Sie ein Signifikanzniveau $\alpha = 0.05$ sowie die unten angegebenen Quantile. (Hinweis: Die zugrundeliegenden Verteilungen sind symmetrisch zu 0.)

$z_{0.95}$	$z_{0.975}$	$t_{3,0.95}$	$t_{3,0.975}$	$t_{4,0.95}$	$t_{4,0.975}$	$t_{6,0.95}$	$t_{6,0.975}$
1.645	1.960	2.353	3.182	2.132	2.776	1.943	2.447

Antwort:

(a) $\bar{X} = \frac{24}{4} = 6$, $S^2 = \frac{1}{3}(1+9+1+1) = 4$.

(b) Der Zwei-Stichproben-Test ist offensichtlich ungeeignet, weil wir nur eine Stichprobe haben. Der Gaußtest ist ungeeignet, weil die Standardabweichung nicht bekannt ist. Es verbleibt der t-Test mit Nullhypothese $H_0\colon \mu \geq 8 =: \mu_0$

(c) $T = \frac{\bar{X}-\mu_0}{S}\sqrt{n} = \frac{-2}{2} \cdot 2 = -2$

Das Ablehnungskriterium $T < t_{n-1,\alpha} = -t_{n-1,1-\alpha} = -t_{3,0.95} = -2.353$ ist nicht erfüllt. Die Aussage der Firma kann also durch das Testergebnis nicht abgelehnt werden.