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

Sommersemester 2007

Diskrete Wahrscheinlichkeitstheorie
Endklausur
LÖSUNG

Hinweis: Bei allen Aufgaben wird neben dem gefragten Ergebnis eine kurze Begründung erwartet, z.B. durch eine entsprechend ausführliche Rechnung.

Aufgabe 1 — je 3 P = 12 P

Für folgende Aufgaben reicht eine kurze Begründung Ihrer Ergebnisse von maximal zwei Zeilen Länge.

1. Es seien A, B, C Ereignisse in dem diskreten Wahrscheinlichkeitsraum $(\Omega, \Pr)$. A sei disjunkt von B und von C. Weiterhin seien B und C unabhängig. Schließlich gelte $\Pr[A] = \frac{1}{3}$, $\Pr[B] = \frac{1}{4}$ und $\Pr[C] = \frac{1}{5}$. Bestimmen Sie die Wahrscheinlichkeit $\Pr[A \cup B \cup C]$.
2. Es seien X, Y unabhängige, normalverteilte Zufallsvariablen mit $\mathbb{E}[X] = 5$, $\text{Var}[X] = \frac{1}{2}$, $\mathbb{E}[Y] = -1$, $\text{Var}[Y] = \frac{1}{3}$. Geben Sie den Namen der Verteilung, den Erwartungswert und die Varianz der Zufallsvariablen $Z := 2X - 3Y - 1$ an.
3. $X_1, X_2, \ldots, X_n$ seien unabhängig und identisch $\text{Bin}(1, \frac{1}{3})$-verteilt, und es sei $Y := \sum_{i=1}^{n} X_i$. Geben Sie den *numerischen* Wert an, den Sie durch den zentralen Grenzwertsatz für $\Pr[Y/n > \frac{1}{3}]$ für große n erhalten. *Hinweis:* Sie brauchen hierfür keine Tabelle.
4. X_1, X_2 seien unabhängig und identisch verteilt mit $X_i \sim \text{Bin}(1, p)$. Es soll die Nullhypothese $H_0 : p = \frac{1}{2}$ überprüft werden mittels der Testgröße $T = X_1 + X_2$ und dem Ablehnungsbereich $\widetilde{K} = \{0, 2\}$. Bestimmen Sie die Wahrscheinlichkeit dafür, dass dieser Test H_0 ablehnt, obwohl H_0 erfüllt ist.

Antwort:

1. Es gilt $\Pr[A \cup B \cup C] = \Pr[A] + \Pr[B \cup C] = \Pr[A] + \Pr[B] + \Pr[C] - \Pr[B \cap C] = \Pr[A] + \Pr[B] + \Pr[C] - \Pr[B] \cdot \Pr[C] = \frac{1}{3} + \frac{1}{4} + \frac{1}{5} - \frac{1}{20} = \frac{20+15+12-3}{60} = \frac{11}{15}$.
2. $Z \sim \mathcal{N}(2 \cdot 5 - 3 \cdot (-1) - 1, 4 \cdot \frac{1}{2} + 9 \cdot \frac{1}{3}) = \mathcal{N}(12, 5)$.
3. Sei $p := 1/3$. Es gilt $\mathbb{E}[Y] = np$ und $\text{Var}[Y] = np(1-p)$.
$$\Pr[Y/n > p] = \Pr[\frac{Y - n \cdot p}{\sqrt{n \cdot p \cdot (1-p)}} > 0] \approx 1 - \Phi(0) = 0.5.$$
4. Gesucht ist
$$\max_{p=\frac{1}{2}} \Pr_p[X_1 + X_2 \in \{0, 2\}] = \max_{p=\frac{1}{2}} (1 - \Pr_p[X_1 + X_2 = 1]) = 1 - 2 \cdot \frac{1}{2} \cdot (1 - \frac{1}{2}) = \frac{1}{2}.$$

Aufgabe 2 — je 2 P = 12 P

In einer Urne befinden sich zu Beginn nur ein roter, ein grüner und ein blauer Ball. Wir betrachten folgendes mehrstufiges Experiment:

In jedem Schritt des Experiments wird ein Ball aus der Urne zufällig herausgenommen, wobei jeder Ball in der Urne mit der gleichen Wahrscheinlichkeit gewählt wird. Es wird dann die Farbe des gezogenen Balls notiert. Sollte die notierte Farbe noch nicht bei vorangegangenen Ziehungen aufgetreten sein, so werden zwei Bälle dieser Farbe zurück in die Urne gelegt. Ansonsten stoppt das Experiment.

Die möglichen Experimentverläufe bilden dann die Elementarereignisse.

Beispiel: Ein mögliches Elementarereignis ist (r, g, g). Das heißt, es werden nacheinander eine rote, eine grüne und noch eine grüne Kugel gezogen. Die Wahrscheinlichkeit ist hierfür $\Pr[(r, g, g)] = \frac{1}{3} \cdot \frac{1}{4} \cdot \frac{2}{5} = \frac{1}{30}$.

1. Die Zufallsvariable Z zählt die Ziehungen bis zum Stoppen des Experiments. Geben Sie alle Werte $k \in \{0, 1, 2, \ldots\}$ an, für welche $\Pr[Z = k] > 0$ gilt.
2. Bestimmen Sie die Dichte und den Erwartungswert von Z.
3. Es sei $\overline{B}$ das Ereignis, dass im *gesamten* Experiment *keine* blaue Kugel gezogen wird.

   Listen Sie die Elementarereignisse von $\overline{B}$ auf.
4. Geben Sie $\Pr[\overline{B}]$ an.
5. Bestimmen Sie $\Pr[Z = 2|\overline{B}]$.
6. Berechnen Sie $\mathbb{E}[Z|\overline{B}]$.

Antwort:

1. Da es nur drei Farben gibt, muss spätestens im vierten Versuch eine Farbe das zweite Mal auftreten. Weiterhin kann frühestens im zweiten Versuch eine Farbe mehrfach auftreten. Damit folgt $W_Z = \{2, 3, 4\}$.
2. $Z = 2$ gilt genau dann, wenn im ersten und im zweiten Versuch dieselbe Farbe gezogen wird. Dies geschieht mit Wahrscheinlichkeit $\frac{1}{2}$.

   Für $Z = 3$ müssen im ersten und im zweiten Versuch verschiedene Farben auftreten, während im dritten Versuch eine Kugel mit der Farbe der ersten oder zweiten Kugel gezogen werden muss. Die Wahrscheinlichkeit, dass die zweite gezogene Farbe von der ersten abweicht, ist $\frac{1}{2}$. Die Wahrscheinlichkeit, dass die dritte gezogene Farbe mit der ersten oder der zweiten übereinstimmt, ist $\frac{4}{5}$. Man erhält also $\Pr[Z = 3] = \frac{2}{5}$.

   Schließlich folgt $\Pr[Z = 4] = 1 - \Pr[Z = 2] - \Pr[Z = 3] = 1 - \frac{1}{2} - \frac{2}{5} = \frac{1}{10}$.

   $$\mathbb{E}[Z] = 2 \cdot \frac{1}{2} + 3 \cdot \frac{2}{5} + 4 \cdot \frac{1}{10} = 1 + 1.2 + 0.4 = 2.6.$$

3. Das entsprechende Ereignis ist $\overline{B} = \{(r, r), (r, g, r), (g, r, r), \} \cup \{(g, g), (g, r, g), (r, g, g)\}$.
4. Damit folgt

   $$\Pr[\overline{B}] = 2 \cdot \left(\frac{1}{3} \cdot \frac{1}{2} + 2 \cdot \frac{1}{3} \cdot \frac{1}{4} \cdot \frac{2}{5}\right) = \frac{1}{3} + \frac{2}{15} = \frac{7}{15}.$$

5. Da Z eingeschränkt auf $\overline{B}$ nur die Werte $\{2, 3\}$ annehmen kann, folgt

   $$\Pr[Z = 2|\overline{B}] = \frac{\Pr[\{(r, r), (g, g)\} \cap \overline{B}]}{\Pr[\overline{B}]} = \frac{\Pr[\{(r, r), (g, g)\}]}{\frac{7}{15}} = \frac{\frac{1}{3}}{\frac{7}{15}} = \frac{5}{7}, \Pr[Z = 3|\overline{B}] = 1 - \Pr[Z = 2|\overline{B}] = \frac{2}{7}.$$

6. Damit folgt

   $$\mathbb{E}[Z|\overline{B}] = 2 \cdot \Pr[Z = 2|\overline{B}] + 3 \cdot \Pr[Z = 3|\overline{B}] = \frac{10 + 6}{7} = \frac{16}{7}.$$

Aufgabe 3 6 P

Bekanntlich läuft heute der Simpson-Film im Kino an. Rund 80% aller Informatikstudenten der Universität MTU planen, sich den Film anzuschauen. Von den restlichen Studenten der MTU möchten nur 40% den Film sehen. Weiterhin ist bekannt, dass es an der MTU 10 000 Studenten gibt, von denen 500 Informatik studieren.

I sei das Ereignis, dass ein MTU-Student Informatik studiert. Mit F sei das Ereignis bezeichnet, dass sich ein MTU-Student den Film anschauen möchte.

Bestimmen Sie die Wahrscheinlichkeit, dass ein Student der MTU, der sich den Film anschauen möchte, Informatik studiert.

Anwort:

Nach Aufgabenstellung gilt $\Pr[I] = 0.05$, $\Pr[F|I] = 0.8$ und $\Pr[F|\overline{I}] = 0.4$.

$$\begin{aligned}&\Pr[I|F] = \Pr[F|I] \cdot \tfrac{\Pr[I]}{\Pr[F]} = \Pr[F|I] \cdot \tfrac{\Pr[I]}{\Pr[F|I]\cdot\Pr[I]+\Pr[F|\overline{I}]\cdot\Pr[\overline{I}]}\\ &= 0.8 \cdot \tfrac{0.05}{0.8\cdot0.05+0.4\cdot0.95} = \tfrac{0.04}{0.04+0.38} = \tfrac{4}{42} = \tfrac{2}{21}.\end{aligned}$$

Aufgabe 4 2 P + 4 P = 6 P

Eine Spedition zahlt ihren Kunden eine Entschädigung, falls eine Lieferung zu lange benötigt. Ist $t \in \mathbb{R}_0^+$ die Zeit in Tagen, die benötigt wird, um die Lieferung zuzustellen, so beträgt die Entschädigung in Euro:

$$g(t) = \begin{cases} 0 & \text{falls } t \in [0,1] \\ t-1 & \text{falls } t \in [1,5] \\ 4 & \text{falls } t > 5. \end{cases}$$

Es sei T die kontinuierliche Zufallsvariable, die die Lieferzeit in Tagen beschreibt. T sei dabei exponentialverteilt mit $\mathbb{E}[T] = 1$.

1. Stellen Sie $\mathbb{E}[g(T)]$ zunächst als Summe von Integralen dar, ohne die Integrale selbst auszuwerten.
2. Bestimmen Sie nun die mittlere Entschädigung, die die Spedition pro Lieferung zahlen muss.

Antwort:

Es gilt:

$$\mathbb{E}[g(T)] = \int_1^5 (t-1)\cdot e^{-t}\cdot dt + \int_5^\infty 4\cdot e^{-t}\cdot dt.$$

Mit

$$\int_a^b \lambda\cdot e^{-\lambda\cdot t}\cdot dt = e^{-\lambda\cdot a} - e^{-\lambda\cdot b}$$

und

$$\int_a^b t\cdot\lambda\cdot e^{-\lambda\cdot t}\cdot dt = a\cdot e^{-\lambda\cdot a} - b\cdot e^{-\lambda\cdot b} + \frac{1}{\lambda}(e^{-\lambda\cdot a} - e^{-\lambda\cdot b})$$

folgt

$$\mathbb{E}[g(T)] = e^{-1} - 5\cdot e^{-5} + (e^{-1} - e^{-5}) - e^{-1} + e^{-5} + 4\cdot e^{-5} = e^{-1} - e^{-5}.$$

Aufgabe 5 **je 3 P = 9 P**

Die Zufallsvariablen $X_1, X_2, \ldots, X_n$ seien unabhängig und identisch Poisson-verteilt mit Erwartungswert λ. Weiterhin sei ein Stichprobenvektor $\vec{k}$

$$\vec{k} := (k_1, k_2, \ldots, k_n) = (X_1, X_2, \ldots, X_n)$$

gegeben.

1. Geben Sie die Likelihood-Funktion $L(\vec{k}; \lambda)$ zu obiger Stichprobe $\vec{k}$ an.
2. Zeigen Sie, dass sich der Maximum-Likelihood-Schätzwert (ML-Schätzwert) $\widehat{\lambda}$ für obige Stichprobe $\vec{k}$ zu

   $$\widehat{\lambda} = \frac{1}{n} \sum_{i=1}^{n} k_i$$

   ergibt. *Hinweis:* Verwenden Sie $\ln L(\vec{k}; \lambda)$.
3. Aus dem ML-Schätzwert $\widehat{\lambda}$ ergibt sich die Schätzvariable

   $$T := \frac{1}{n} \sum_{i=1}^{n} X_i.$$

   Ist T erwartungstreu?

Antwort:

1. Es gilt

   $$L(\vec{k}; \lambda) = \Pr[X_1 = k_1, X_2 = k_2, \ldots, X_n = k_n] = \prod_{i=1}^{n} \Pr[X_i = k_i] = \prod_{i=1}^{n} \left(e^{-\lambda} \cdot \frac{\lambda^{k_i}}{k_i!} \right) = e^{-n \cdot \lambda} \cdot \prod_{i=1}^{n} \frac{\lambda^{k_i}}{k_i!}.$$

   Damit folgt:

   $$\widetilde{L}(\vec{k}; \lambda) = \ln L(\vec{k}; \lambda) = -n \cdot \lambda + \sum_{i=1}^{n} \ln \left(\frac{\lambda^{k_i}}{k_i!} \right) = -n \cdot \lambda + \left(\sum_{i=1}^{n} k_i \right) \cdot \ln \lambda - \sum_{i=1}^{n} \ln (k_i!)\,.$$

2. Zu maximieren ist die Funktion $L(\vec{k}; \lambda)$ nach dem Parameter λ. Da e^x streng monton wächst, reicht es auch aus, die Funktion $\widetilde{L}(\vec{k}; \lambda)$ zu maximieren.

   Ableiten führt auf

   $$\widetilde{L}'(\vec{k}; \widehat{\lambda}) = -n + \left(\sum_{i=1}^{n} k_i \right) \cdot \frac{1}{\widehat{\lambda}} \stackrel{!}{=} 0 \text{ bzw. } \widehat{\lambda} = \frac{\sum_{i=1}^{n} k_i}{n}.$$

3. Wegen $\mathbb{E}[T] = \frac{1}{n} \cdot n \cdot \mathbb{E}[X_1] = \mathbb{E}[X_1] = \lambda$ ist T erwartungstreu.