Technische Universität München
Fakultät für Informatik
Lehrstuhl für Effiziente Algorithmen
Prof. Dr. Ernst W. Mayr
Dr. Werner Meixner

Sommersemester 2010
Lösungsblatt 7
18. Juni 2010

Diskrete Wahrscheinlichkeitstheorie

Hausaufgabe 1 (5 Punkte)

Wir machen n unabhängige Würfe mit je 3 fairen Würfeln. Y sei die Summe aller gewürfelten Augen.

1. Berechnen Sie Erwartungswert und Varianz von Y!
2. Zeigen Sie, dass Y mit 99%-iger Wahrscheinlichkeit im Intervall $\frac{21}{2}n \pm 5\sqrt{35n}$ liegt.

Lösungsvorschlag

1. Sei X die Augenzahl, die bei einem Wurf mit einem einzigen Würfel geworfen wird. Der Erwartungswert der Zufallsvariablen X ist
$$\mathbb{E}[X] = \left(\sum_{1\leq i\leq 6} \frac{i}{6}\right) = \frac{1}{6}\cdot\frac{6\cdot 7}{2} = \frac{7}{2}\,.$$
Die Varianz der Augenzahl X bei einem Wurf beträgt
$$\mathrm{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \left(\sum_{1\leq i\leq 6}\frac{i^2}{6}\right) - \left(\frac{7}{2}\right)^2 = \frac{91}{6} - \frac{49}{4} = \frac{35}{12}\,.$$
Aufgrund der Unabhängigkeit der Würfe ergibt sich für die Varianz der Summe Y bei n Würfen von je zwei Würfeln
$$\mathrm{Var}(Y) = 3n\mathrm{Var}(X) = \frac{35}{4}n\,.$$
Um den Erwartungswert der Summe Y zu erhalten, muss die Unabhängigkeit der Würfe nicht vorausgesetzt werden. Es gilt
$$\mathbb{E}[Y] = 3n\mathbb{E}[X] = \frac{21}{2}n\,.$$
2. Mit der Standardabweichung $\sigma = \sqrt{\mathrm{Var}[Y]}$ lautet die Chebyshevsche Ungleichung für alle $k > 0$
$$\Pr[|Y - \mathbb{E}(Y)| \geq k\sigma] \leq \frac{1}{k^2}\,,$$
und wegen $\Pr[\,|Y-\mathbb{E}(Y)| \geq k\sigma\,] + \Pr[\,|Y-\mathbb{E}(Y)| < k\sigma\,] = 1$ ist diese Ungleichung gleichbedeutend mit
$$1 - \frac{1}{k^2} \;\leq\; \Pr[\,|Y-\mathbb{E}(Y)| < k\sigma\,] \;\leq\; 1\,.$$
Wir substituieren $\mathbb{E}[Y]$ und σ und erhalten mit $k^2 = 100$
$$1 - \frac{1}{100} \;\leq\; \Pr[\,|Y - \frac{21}{2}n| < 5\sqrt{35n}\,] \;\leq\; 1\,.$$

Hausaufgabe 2 (5 Punkte)

Es sei M_1 eine Maschine, die bei Aufruf zufällig mit gleicher Wahrscheinlichkeit eine der Zahlen 1, 2, 3 oder 4 ausgibt. Wir bezeichnen die entsprechende Zufallsvariable mit N. Eine Maschine M_2 werfe bei Aufruf eine faire Münze, die entweder „Kopf" oder „Wappen" zeigt.

Wir betrachten einen Algorithmus A, dessen Ausführung in 2 Schritten ein Ergebnis erzeugt. Im ersten Schritt wird M_1 veranlasst, eine Zahl k auszugeben. Im zweiten Schritt wird M_2 k mal aufgerufen. Das Ergebnis einer Ausführung von A definieren wir als diejenige Zahl, die angibt, wie oft im zweiten Schritt „Kopf" geworfen wurde.
Es sei Y die Zufallsvariable, die die Ergebnisse des Algorithmus A beschreibt.

1. Geben Sie die wahrscheinlichkeitserzeugende Funktion $G_N(z)$ für N an.
2. Berechnen Sie den Erwartungswert $\mathbb{E}[Y]$.
3. Berechnen Sie die wahrscheinlichkeitserzeugende Funktion $G_Y(z)$ für Y.

Lösungsvorschlag

1. Ausgehend von der Gleichverteilung für 4 Ereignisse erhalten wir

$$G_N(z) = \sum_{k=1}^{4} \frac{1}{4} z^k \,.$$

2. Die Indikatorvariable X gebe mit Wert 1 an, dass Kopf geworfen wurde. Die erzeugende Funktion für X ist

$$G_X(z) = \frac{1}{2} + \frac{1}{2} z.$$

Es gilt $G'_N(z) = \frac{1}{4} + \frac{2}{4}z + \frac{3}{4}z^2 + \frac{4}{4}z^3$

und $G'_X(z) = \frac{1}{2}\,.$

Mit $\mathbb{E}[X] = G'_X(1) = \frac{1}{2}$

und $\mathbb{E}[N] = G'_N(1) = \frac{1}{4}(1+2+3+4) = \frac{5}{2}$

erhalten wir

$$\mathbb{E}[Y] = \mathbb{E}[N] \cdot \mathbb{E}[X] = \frac{5}{2} \cdot \frac{1}{2} = \frac{5}{4}\,.$$

3.

$$\begin{aligned} G_Y(z) &= G_N(G_X(z)) \\ &= \frac{1}{4}\left(\frac{1+z}{2} + \left(\frac{1+z}{2}\right)^2 + \left(\frac{1+z}{2}\right)^3 + \left(\frac{1+z}{2}\right)^4\right) \\ &= \frac{1}{64}(15 + 26z + 16z^2 + 6z^3 + z^4)\,. \end{aligned}$$

Hausaufgabe 3 (5 Punkte)

Sei $F(z)$ bzw. $G(z)$ die wahrscheinlichkeitserzeugende Funktion für eine Zufallsvariable X_F bzw. X_G. Sei $H(z) = F(G(z))$.

Berechnen Sie $H(1)$, $H'(1)$ und $H''(1)$ in Abhängigkeit ggf. von $\mathbb{E}[X_F]$, $\mathrm{Var}[X_F]$, $\mathbb{E}[X_G]$ und $\mathrm{Var}[X_G]$.

Lösungsvorschlag

Es gilt $G(1) = 1$ und $F(1) = 1$, mithin

$$H(1) = F(G(1)) = F(1) = 1\,.$$

Mit der Kettenregel erhalten wir

$$H'(z) = F'(G(z)) \cdot G'(z)$$

und

$$H''(z) = F'(G(z)) \cdot G''(z) + F''(G(z)) \cdot G'(z)^2\,,$$

und damit

$$H'(1) = F'(G(1)) \cdot G'(1) = F'(1) \cdot G'(1) = \mathbb{E}[X_F] \cdot \mathbb{E}[X_G]$$

und

$$H''(1) = F'(1) \cdot G''(1) + F''(1) \cdot G'(1)^2 = \mathbb{E}[X_F] \cdot G''(1) + F''(1) \cdot \mathbb{E}[X_G]^2\,.$$

Mit $G''(1) = \mathrm{Var}[X_G] - \mathbb{E}[X_G] + \mathbb{E}[X_G]^2$ und $F''(1) = \mathrm{Var}[X_F] - \mathbb{E}[X_F] + \mathbb{E}[X_F]^2$ (siehe Vorlesung) erhalten wir

$$H''(1) = \mathbb{E}[X_F] \cdot \big(\mathrm{Var}[X_G] - \mathbb{E}[X_G] + \mathbb{E}[X_G]^2\big) + \big(\mathrm{Var}[X_F] - \mathbb{E}[X_F] + \mathbb{E}[X_F]^2\big) \cdot \mathbb{E}[X_G]^2\,.$$

Hausaufgabe 4 (5 Punkte)

Seien A, B und C jeweils (unabhängig) gleichverteilt über $[0,1]$. Wie groß ist die Wahrscheinlichkeit, dass alle Lösungen der Gleichung $Ax^2 + Bx + C = 0$ reellwertig sind? (Hinweis: Berechnen Sie zunächst die Dichtefunktion von $A \cdot C$ und verwenden Sie für die sich anschließende Rechnung $\int x^2 \ln x \,\mathrm{d}x = \frac{x^3 \ln x}{3} - \frac{x^3}{9}$.)

Lösungsvorschlag

Alle Lösungen sind genau dann reell, wenn $B^2 \geq 4AC$ gilt.
Wir definieren $V := \{(A,B,C)\,;\, B^2 \geq 4AC \wedge 0 \leq A,B,C \leq 1\}$. Aus Gründen der Unabhängigkeit und Gleichverteilung der A,B,C ist $\Pr[B^2 \geq 4AC]$ durch das Verhältnis des Volumens von V zum Volumen des Einheitswürfels $[0,1]^3$ bestimmt. Das Volumen des Einheitswürfels ist 1.

Wir definieren noch $F(B) := \{(A,C)\,;\, B^2 \geq 4AC \wedge 0 \leq A, C \leq 1\}$ und erhalten

$$\begin{aligned}
\Pr[B^2 \geq 4AC] &= \int_V \mathrm{d}A\,\mathrm{d}B\,\mathrm{d}C \\
&= \int_0^1 \left(\int_{F(B)} \mathrm{d}A\,\mathrm{d}C\right)\,\mathrm{d}B \\
&= \int_0^1 \left(\int_0^{B^2/4} \mathrm{d}A + \int_{B^2/4}^1 \frac{B^2/4}{A}\,\mathrm{d}A\right)\mathrm{d}B \\
&= \int_0^1 \frac{B^2}{4} - \frac{B^2}{4}\ln\frac{B^2}{4}\,\mathrm{d}B \\
&= \frac{\ln 2}{6} + \frac{5}{36} \approx 25.4\%\,.
\end{aligned}$$

Hinweis: *Auf den Übungsblättern in diesem Semester wird es grundsätzlich die drei Aufgabentypen Vorbereitungsaufgabe, Tutoraufgabe und Hausaufgabe geben. Die als Vorbereitung bezeichneten Aufgaben dienen der häuslichen Vorbereitung der Tutoraufgaben. Tutoraufgaben werden in den Übungsgruppen bearbeitet. Dabei wird die Lösung der Vorbereitungsaufgaben vorausgesetzt. Die Vorbereitungsaufgaben werden in der Zentralübung unterstützt.*

Vorbereitung 1

1. Sei $(H_n)_{n\geq 1}$ eine rekurrente Ereignisfolge. Die Zufallsvariable Z mit $W_Z = \mathbb{N} \cup \{\infty\}$ messe für $k \in \mathbb{N}$ die Wartezeit $Z = k$ bis zum Eintreten des ersten Ereignisses H_k der Ereignisfolge.

   Zeigen Sie die Ungleichung
   $$\sum_{k\in\mathbb{N}} \Pr[Z = k] \leq 1\,.$$

2. Sei $(X_n)_{n\geq 1}$ eine Folge unabhängiger Indikatorvariablen mit gleicher Bernoulli-Verteilung.

   Zeigen Sie, dass die Folge $(H_n)_{n\geq 1}$ der Ereignisse $H_n = (X_n = 1)$ rekurrent ist.

Lösungsvorschlag

1. *Bemerkung*: Bei oberflächlicher Betrachtung erscheint die Gültigkeit der Ungleichung als eine triviale Folge der Eigenschaft von Z, eine „Zufallsvariable" zu sein. Die Tatsache, dass $\Pr[Z = k]$ für alle $k \in \mathbb{N} \cup \{\infty\}$ definiert wurde, heißt aber noch nicht, dass Z bezüglich dieser Definition eine Zufallsvariable ist. Insbesondere haben wir bisher nur numerische Zufallsvariable mit $Z : \Omega \to \mathbb{R}$ betrachtet. Die Aufgabe liefert also erstmals den Nachweis, dass wir über Z von einer Zufallsvariablen sinnvollerweise sprechen dürfen.

   Zum Beweis genügt es im Prinzip die paarweise Disjunktheit aller Ereignisse $Z = k$ für alle $k \in \mathbb{N}$ zu beweisen. Wir beweisen etwas ausführlicher ausgehend von der Gleichung
   $$\Pr[H_1] + \Pr[\bar{H}_1] = 1\,.$$

Wegen $\Pr[Z=1] = \Pr[H_1]$ folgt $\sum_{i=1}^{1} \Pr[Z=i] + \Pr[\bar{H}_1] = 1$. Dies ist der Induktionsanfang zum Beweis der folgenden Gleichung für alle $n \geq 1$.

$$\sum_{i=1}^{n} \Pr[Z=i] + \Pr\left[\bar{H}_1 \cap \bar{H}_2 \cap \ldots \cap \bar{H}_n\right] = 1\,.$$

Den Induktionsschritt von n auf $n+1$ beweist man wie folgt.

$$\begin{aligned}
1 &= \sum_{i=1}^{n} \Pr[Z=i] + \Pr\left[\bar{H}_1 \cap \bar{H}_2 \cap \ldots \cap \bar{H}_n\right] \\
&= \sum_{i=1}^{n} \Pr[Z=i] + \Pr\left[\bar{H}_1 \cap \bar{H}_2 \cap \ldots \cap \bar{H}_n \cap H_{n+1}\right] \\
&\qquad\qquad + \Pr\left[\bar{H}_1 \cap \bar{H}_2 \cap \ldots \cap \bar{H}_n \cap \bar{H}_{n+1}\right] \\
&= \sum_{i=1}^{n} \Pr[Z=i] + \Pr[Z=n+1] + \Pr\left[\bar{H}_1 \cap \bar{H}_2 \cap \ldots \cap \bar{H}_{n+1}\right] \\
&= \sum_{i=1}^{n+1} \Pr[Z=i] + \Pr\left[\bar{H}_1 \cap \bar{H}_2 \cap \ldots \cap \bar{H}_{n+1}\right]\,.
\end{aligned}$$

Daraus folgt unmittelbar die Behauptung.

2. Wir zeigen für alle $i, j \in \mathbb{N}$ mit $i > j$

$$\Pr\left[H_i \,|\, \bar{H}_1 \cap \ldots \cap \bar{H}_{j-1} \cap H_j\right] = \Pr[H_{i-j}]\,.$$

Sei p die Erfolgswahrscheinlichkeit für die Variablen X_n. Für alle $i \in \mathbb{N}$ gilt $\Pr[H_i] = p$. Da die X_i unabhängig sind, folgt

$$\begin{aligned}
\Pr\left[H_i \,|\, \bar{H}_1 \cap \ldots \cap \bar{H}_{j-1} \cap H_j\right] &= \frac{\Pr\left[H_i \cap \bar{H}_1 \cap \ldots \cap \bar{H}_{j-1} \cap H_j\right]}{\Pr[\left[\bar{H}_1 \cap \ldots \cap \bar{H}_{j-1} \cap H_j\right]} \\
&= \Pr[H_i] \\
&= p \\
&= \Pr[H_{i-j}]\,.
\end{aligned}$$

Tutoraufgabe 1

Peter und Paul spielen ein Spiel, bei dem jeder von ihnen mit Wahrscheinlichkeit 1/2 gewinnt. In jeder Runde setzt jeder von ihnen € 10 ein (der Gewinner erhält dann € 20).

1. Wie groß ist die Wahrscheinlichkeit, dass Peter € 100 Gewinn gemacht hat, bevor er € 50 Verlust macht?

2. Wenn Peter erst mal seinen Gewinn hat, wie groß ist dann die Wahrscheinlichkeit, dass er insgesamt auf € 50 Verlust kommt?

Lösungsvorschlag

Diese Aufgabe betrachtet bis auf Skalierung (Faktor 10) das Beispiel des eindimensionalen Random Walk und gehört zum Thema rekurrente Ereignisse.

Die wichtigste Beobachtung ist zunächst, dass bei diesem Spiel sowohl Peter als auch Paul irgendwann jeden Gewinnbetrag und jeden Verlustbetrag mindestens einmal überschreiten werden. A.a., die Wahrscheinlichkeit Pr[(nie Gewinn von X) und (nie Verlust von Y)] ist stets gleich Null,

$$\Pr[(\text{nie Gewinn von } X) \text{ und } (\text{nie Verlust von } Y)] = 0\,.$$

Man beweist dies ähnlich, wie man beweisen kann, dass man mit Wahrscheinlichkeit 1 irgendwann beim Random Walk zum Nullpunkt zurückkehrt. Gewinn– bzw. Verlustüberschreitungen werden stets irgenwann das erste Mal stattfinden, und das gilt sogar nach jeder Spielsituation beliebig oft.

1. Wir definieren die abkürzende Schreibweise $\Pr[X|-Y]$ bzw. $\Pr[-X|Y]$ für die Wahrscheinlichkeit, einen

   - erstmaligen Gewinn von X zu machen, ohne vorher einen Verlust von Y gemacht zu haben, bzw. einen
   - erstmaligen Verlust von X zu machen, ohne vorher einen Gewinn von Y gemacht zu haben.

   Nun gilt

   $$\Pr[100|-50] + \Pr[-50|100] + \Pr[\text{nie Gewinn von } 100 \text{ und nie Verlust von } 50] = 1$$

   und damit

   $$\Pr[100|-50] + \Pr[-50|100] = 1\,.$$

   Aus Symmetriegründen bezüglich den Spielern Peter und Paul gilt

   $$\Pr[-50|100] = \Pr[50|-100]\,.$$

   Wenn wir abkürzen

   $$p_1 := \Pr[100|-50] \quad \text{und} \quad p_2 := \Pr[50|-100]$$

   erhalten wir

   $$p_1 + p_2 = 1\,.$$

   Jedem Gewinn von 100 ohne Verlust von 50 geht ein Gewinn von 50 ohne Verlust von 50 voraus. Der zusätzliche Gewinn von 50 hin zum Gewinn von 100 geschieht ohne vorherigem Verlust von 100. Mit der Abkürzung

   $$p := \Pr[50|-50]$$

   gilt

   $$p_1 = p \cdot p_2\,.$$

Es gilt aber auch $p = \frac{1}{2}$, weil

$$\Pr[50|-50) + \Pr[-50|50] = 1 \quad \text{und} \quad \Pr[50|-50] = \Pr[-50|50]\,.$$

Damit haben wir

$$p_2 = 2 \cdot p_1, \quad \text{und} \quad p_1 + p_2 = 1\,,$$

mithin

$$\Pr[100|-50] = p_1 = \frac{1}{3}\,.$$

2. Die Wahrscheinlichkeit ist gleich 1, den Gewinn von 100 (ohne vorausgehenden Verlust von 50) wieder zu verspielen bis hin zu einem Gesamtverlust (bezüglich Spielanfang) von 50.

Tutoraufgabe 2

Bei einem Einwahlserver für $n = 10^3$ Teilnehmer nehmen wir an, dass zu einem festen Zeitpunkt jeder Teilnehmer mit Wahrscheinlichkeit $p = 0,05$ Zugriff auf den Server wünscht.

Berechnen Sie eine Näherung der Wahrscheinlichkeit, mit der gleichzeitig mehr als 55 Verbindungswünsche auftreten? Approximieren Sie dabei die Binomialverteilung durch die entsprechende Normalverteilung und benutzen Sie ggf. geeignete Tabellen für die Werte der Standardnormalverteilung.

Lösungsvorschlag

Sei X die Anzahl der Verbindungswünsche. X ist näherungsweise normalverteilt. Wir nehmen also an $X \sim \mathcal{N}(\mu, \sigma^2)$ mit $\mu = np = 50$ und $\sigma^2 = np(1-p) = 47,5$. Damit folgt

$$\begin{aligned} \Pr[X > 55] &= 1 - \Pr[X \le 55] \\ &= 1 - \Pr\left[\frac{X-\mu}{\sigma} \le \frac{55-50}{\sqrt{47,5}}\right] \\ &= 1 - \Phi\left(\frac{55-50}{\sqrt{47,5}}\right) \\ &\approx 0,2343\,. \end{aligned}$$

Tutoraufgabe 3

Wir benutzen die Funktion $h(t) = 0.027 + 0.0025 \cdot (t-40)^2$ für $t \in \mathbb{R}$, um die „Sterblichkeitsrate“ durch Lungenkrebs von Kettenraucherinnen abzuschätzen, die mindestens $t \ge 40$ Jahre alt sind. Ihre Lebensdauer sei X und es gelte

$$\Pr[X > t \,|\, X > 40] = \exp\left(-\int_{40}^{t} h(s)\mathrm{d}s\right)\,.$$

Wie groß ist die Wahrscheinlichkeit, dass eine 45-jährige Kettenraucherin mindestens 50 Jahre alt wird?

Lösungsvorschlag

Zu berechnen ist offenbar $\Pr[X \geq 50 \mid X \geq 45]$.
Bei stetigen Zufallsvariablen gilt $\Pr[X \geq 50 \mid X \geq 45] = \Pr[X > 50 \mid X > 45]$.
Für $t \geq s \geq 40$ berechnen wir allgemein die folgende bedingte Wahrscheinlichkeit.

$$\begin{aligned}
\Pr[X > t \mid X > s] &= \frac{\Pr[(X > t) \cap (X > s)]}{\Pr[X > s]} \\
&= \frac{\Pr[X > t]}{\Pr[X > s]} \\
&= \frac{\Pr[X > t \mid X > 40]}{\Pr[X > s \mid X > 40]} \\
&= \exp\left(- \int_{40}^{t} h(x)\ \mathrm{d}x + \int_{40}^{s} h(x)\ \mathrm{d}x \right) \\
&= \exp\left(- \int_{s}^{t} h(x)\ \mathrm{d}x \right) .
\end{aligned}$$

Daher gilt

$$\Pr[X > 50 \mid X > 45] \;=\; \exp\left(- \int_{45}^{50} 0.027 + 0.0025(t-40)^2\ \mathrm{d}t \right) ,$$

woraus wegen

$$\int_{45}^{50} 0.027 + 0.0025(t-40)^2\ \mathrm{d}t \;=\; 0.135 + 2.1875/3 \;=\; 0.86417\bar{6}$$

folgt, dass gilt

$$\Pr[X \geq 50 \mid X \geq 45] \;=\; \exp(-0.86417\bar{6}) \;\approx\; 42.1402575\ldots\% .$$