Technische Universität München
Theoretische Informatik
Prof. J. Esparza / S. Kiefer / M. Luttenberger

Sommersemester 2009
Übungsblatt 4

LÖSUNG

Übungen zu Diskrete Wahrscheinlichkeitstheorie

Dieses Blatt ist am 25. Mai bis 10:15 abzugeben und wird am 25./26. Mai besprochen.

Beachten Sie: Alle Ergebnisse sind zu begründen!

Aufgabe 4.1

1P+1P+1P+1P+1P

Seien X und Y zwei Zufallsvariablen, sodass X, Y, $X+Y$ und $X-Y$ alle dieselbe Verteilung haben.

(a) Zeigen Sie: $\mathbb{E}[X] = 0$.

(b) Zeigen Sie: $\mathbb{E}[XY] = 0$. *Hinweis*: Vergleichen Sie z.B. $\mathbb{E}[(X+Y)^2]$ und $\mathbb{E}[(X-Y)^2]$.

(c) Berechnen Sie $\mathrm{Var}[X]$.

(d) Berechnen Sie die Verteilung von X.

(e) Geben Sie zwei Zufallsvariablen X und Y an, sodass X, Y und $X+Y$ alle dieselbe Verteilung haben, aber $X-Y$ eine andere Verteilung hat. *Hinweis*: Es gibt z.B. eine Lösung, in der X und Y den Wertebereich $\{-1, 0, 1\}$ haben.

Lösungsvorschlag:

(a) $\mathbb{E}[X] = \mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y] = \mathbb{E}[X] + \mathbb{E}[X]$, folglich $\mathbb{E}[X] = 0$.

(b) $\mathbb{E}[(X+Y)^2] = \mathbb{E}[X^2 + Y^2 + 2XY] = \mathbb{E}[X^2] + \mathbb{E}[Y^2] + \mathbb{E}[2XY]$.
$\mathbb{E}[(X-Y)^2] = \mathbb{E}[X^2 + Y^2 - 2XY] = \mathbb{E}[X^2] + \mathbb{E}[Y^2] - \mathbb{E}[2XY]$.
Da $X+Y$ und $X-Y$ identisch verteilt sind, folgt $\mathbb{E}[2XY] = -\mathbb{E}[2XY]$, also $\mathbb{E}[XY] = 0$.

(c) Aus $\mathbb{E}[(X+Y)^2] = \mathbb{E}[X^2] + \mathbb{E}[Y^2]$ und $\mathbb{E}[X^2] = \mathbb{E}[Y^2] = \mathbb{E}[(X+Y)^2]$ folgt $\mathbb{E}[X^2] = \mathbb{E}[X^2] + \mathbb{E}[X^2]$, also $\mathbb{E}[X^2] = 0$.
Also $\mathrm{Var}[X] = \mathbb{E}[X^2] - (\mathbb{E}X)^2 = 0$.

(d) Wir haben $0 = \mathbb{E}[X^2] = \sum_{\omega \in \Omega} (X(\omega))^2 \Pr[\omega]$. Also $\Pr[X = 0] = 1$.

(e) Wie in (a) muss $\mathbb{E}[X] = 0$ gelten.

$\Omega = \{a, b, c, d, e, f\}$ mit $\Pr[\omega] = 1/6$ für alle $\omega \in \Omega$.

ω	a	b	c	d	e	f
X	-1	-1	0	0	$+1$	$+1$
Y	0	$+1$	-1	$+1$	0	-1
$X+Y$	-1	0	-1	$+1$	$+1$	0
$X-Y$	-1	-2	$+1$	-1	$+1$	$+2$

Aufgabe 4.2

2P+1P+1P+1P+1P

Hinweis: Schauen Sie sich die „wichtigen diskreten Verteilungen“ im Kapitel 7 an.

(a) Was ist wahrscheinlicher: In sechs Mal Würfeln mindestens einen Sechser zu würfeln, oder in zwölf Mal Würfeln mindestens zwei Sechser zu würfeln? Berechnen Sie diese W'keiten.

(b) Sie würfeln, bis Sie einen Sechser bekommen. Wie groß ist die W'keit, dass Sie den Sechser beim 4. Wurf bekommen?

(c) Sie würfeln, bis Sie drei Sechser (nicht unbedingt hintereinander) bekommen haben. Wie groß ist die W'keit, dass Sie den dritten Sechser beim 7. Wurf bekommen?

(d) Sie würfeln, bis Sie drei Sechser (nicht unbedingt hintereinander) bekommen haben. Wie oft müssen Sie im Schnitt würfeln, bis Sie den dritten Sechser bekommen haben?

(e) Sie würfeln, bis Sie alle sechs Augenzahlen jeweils mindestens ein Mal bekommen haben. Wie oft würfeln Sie im Schnitt?

Lösungsvorschlag:

(a) Sei X die Anzahl der Sechser in 6 Würfen. Dann ist $X \sim \text{Bin}(6, \frac{1}{6})$.

$$\Pr[X \geq 1] = 1 - \Pr[X = 0] = 1 - \binom{6}{0}\left(\frac{1}{6}\right)^0\left(\frac{5}{6}\right)^6 = 1 - \left(\frac{5}{6}\right)^6 \approx 0.665$$

Sei Y die Anzahl der Sechser in 12 Würfen. Dann ist $Y \sim \text{Bin}(12, \frac{1}{6})$.

$$\Pr[Y \geq 2] = 1 - \Pr[Y = 0] - \Pr[Y = 1] = 1 - \binom{12}{0}\left(\frac{1}{6}\right)^0\left(\frac{5}{6}\right)^{12} - \binom{12}{1}\left(\frac{1}{6}\right)^1\left(\frac{5}{6}\right)^{11}$$
$$= 1 - \left(\frac{5}{6}\right)^{12} - 12 \cdot \frac{5^{11}}{6^{12}} \approx 0.619$$

Folglich ist ein Sechser in 6 Würfen wahrscheinlicher.

(b) Sei Z die Zahl der Würfe, bis der Sechser kommt (der Wurf mit dem Sechser eingeschlossen). Dann ist $Z \sim \text{Geo}(\frac{1}{6})$.

$$\Pr[Z = 4] = \left(\frac{5}{6}\right)^3 \cdot \frac{1}{6} \approx 0.096$$

(c) Sei Z jetzt die Zahl der Würfe, bis der dritte Sechser kommt (der Wurf mit dem dritten Sechser eingeschlossen). Dann ist Z laut Folien "negativ binomialverteilt" und es gilt

$$\Pr[Z = 7] = \binom{6}{2} \cdot \left(\frac{1}{6}\right)^3 \cdot \left(\frac{5}{6}\right)^4 = 15 \cdot \frac{5^4}{6^7} \approx 0.0335$$

(d) Sei X_1 die Zahl der Würfe bis einschließlich zum 1. Sechser. Sei X_2 die Zahl der Würfe nach dem 1. Sechser bis einschließlich zum 2. Sechser. Sei X_3 die Zahl der Würfe nach dem 2. Sechser bis einschließlich zum 3. Sechser. Dann sind X_1, X_2, X_3 alle $\text{Geo}(\frac{1}{6})$-verteilt, daher gilt $\mathbb{E}X_1 = \mathbb{E}X_2 = \mathbb{E}X_3 = \frac{1}{1/6} = 6$. Bis zum dritten Sechser dauert es $X_1 + X_2 + X_3$ Würfe, im Schnitt also $\mathbb{E}[X_1 + X_2 + X_3] = \mathbb{E}X_1 + \mathbb{E}X_2 + \mathbb{E}X_3 = 18$ Würfe.

(e) Sei X_1 die Zahl der Würfe bis zur 1. Augenzahl. Offenbar ist $X_1 = 1$. Sei X_2 die Zahl der Würfe nach der 1. Augenzahl bis einschließlich zur 2. Augenzahl. Es gilt $X_2 \sim \text{Geo}(\frac{5}{6})$ und daher $\mathbb{E}X_2 = \frac{6}{5}$. Die Zufallsvariablen X_3, X_4, X_5, X_6 seien ähnlich definiert. Bis alle sechs Augenzahlen mindestens ein Mal gekommen sind, dauert es $X_1 + \cdots + X_6$ Würfe, im Schnitt $\mathbb{E}[X_1 + \cdots + X_6] = \mathbb{E}X_1 + \cdots + \mathbb{E}X_6 = 1 + \frac{6}{5} + \frac{6}{4} + \frac{6}{3} + \frac{6}{2} + 6 = 14.7$ Würfe.

Aufgabe 4.3 **1P+1P+1P+2P+1P**

Es werden 200 jeweils 1GB große Filmdateien zufällig auf 100 Server verteilt. Jeder Server hat eine Kapazität von k GB, d.h., bekommt ein Server mehr als k Filme, dann „läuft er über". Sie betreiben die Serverfarm und möchten wissen, auf welche Größe Sie die Server auslegen sollten.

(a) Sei, für $i \in \{1, \ldots, 100\}$, die Zufallsvariable X_i die Anzahl der Filme auf Server i. Geben Sie die Verteilung von X_i an. *Hinweis*: Überlauf spielt in (a)–(c) noch keine Rolle.

(b) Berechnen Sie $\Pr[X_i > 6]$.

(c) Sei $k = 6$. Benutzen Sie das Ergebnis aus (b) und die boolesche Ungleichung aus den Folien, um zu zeigen, dass die W'keit, dass mindestens ein Server überläuft, kleiner als 0.5 ist.

Sie möchten die W'keit, dass mindestens ein Server überläuft, auf eine W'keit deutlich unter 0.5 begrenzen. Dazu müssen Sie die Kapazität k erhöhen. Allerdings sind Rechnungen wie in (b) mühsam, weil Binomialkoeffizienten und Summen berechnet werden müssen.

(d) Zeigen Sie wieder mithilfe der booleschen Ungleichung:

$$\Pr[X_1 \geq j] \leq \binom{200}{j} \cdot 0.01^j$$

(e) Mit Teil (d) und der Abschätzung $\binom{n}{m} \le \left(\frac{en}{m}\right)^m$, wobei e für die Eulersche Zahl steht, erhalten Sie

$$\Pr[X_1 \ge j] \le \binom{200}{j} \cdot 0.01^j \le \left(\frac{2e}{j}\right)^j$$

Bestimmen Sie damit eine möglichst kleine Kapazität k, sodass die W'keit, dass *mindestens ein* Server überläuft, auf höchstens 0.01 begrenzt ist.

Lösungsvorschlag:

(a) $X_i \sim \mathrm{Bin}(200, 1/100)$, d.h. $\Pr[X_i = j] = \binom{200}{j} \cdot 0.01^j \cdot 0.99^{200-j}$.

(b) $\Pr[X_i > 6] = 1 - \Pr[X_i \le 6] = 1 - \sum_{j=0}^{6} \Pr[X_i = j] = 1 - \sum_{j=0}^{6} \binom{200}{j} \cdot 0.01^j \cdot 0.99^{200-j} \approx 0.00430$.

(c) $\Pr[„X_1 > 6“ \cup \cdots \cup „X_{100} > 6“] \le \Pr[„X_1 > 6“] + \cdots + \Pr[„X_{100} > 6“] \le 100 \cdot 0.0044 = 0.44$.

(d) Für jede j-elementige Teilmenge D von $\{1, \ldots, 200\}$ sei $E(D)$ das Ereignis, dass alle j Filme mit Nummern in D auf Server 1 landen. Für alle j-elementigen Teilmengen D von $\{1, \ldots, 200\}$ gilt $\Pr[E(D)] = 0.01^j$. Es gilt:

$$„X_1 \ge j“ = \bigcup_{\substack{D \subseteq \{1,\ldots,200\} \\ |D| = j}} E(D)$$

Mit der booleschen Ungleichung folgt:

$$\Pr[„X_1 \ge j“] = \Pr\left[\bigcup_{\substack{D \subseteq \{1,\ldots,200\} \\ |D| = j}} E(D)\right] \le \sum_{\substack{D \subseteq \{1,\ldots,200\} \\ |D| = j}} \Pr[E(D)] = \sum_{\substack{D \subseteq \{1,\ldots,200\} \\ |D| = j}} 0.01^j = \binom{200}{j} \cdot 0.01^j$$

(e) Mit dem Argument aus (c) muss ein k gefunden werden mit $\Pr[X_1 \ge k+1] \le 0.0001$. Es gilt

$$\left(\frac{2e}{11}\right)^{11} \approx 0.000430 \quad \text{und} \quad \left(\frac{2e}{12}\right)^{12} \approx 0.0000748$$

Folglich kann $k = 11$ gewählt werden.

Aufgabe 4.4 **1P+2P**

(a) Zeigen Sie, dass

$$\sum_{i=0}^{k} \binom{r}{i} \cdot \binom{b}{k-i} = \binom{r+b}{k}$$

gilt.

Hinweis: Nehmen Sie an, dass Sie r rote und b blaue Bälle hätten. Wie kann dann der Ausdruck auf der linken bzw. der rechten Seite interpretiert werden?

(b) Zwei Personen werfen jeweils n-mal eine Münze. Bestimmen Sie die W'keit, dass beide Personen dieselbe Anzahl von "Kopf" erhalten.

Lösungsvorschlag:

(a) Sei $R = \{\rho_1, \ldots, \rho_r\}$ und $B = \{\beta_1, \ldots, \beta_b\}$ Mengen von r bzw. b verschiedenen Bällen der Farbe Rot bzw. Blau.

Dann gibt $\binom{r+b}{k}$ die Anzahl der Teilmenge $A \subseteq R \cup B$ mit $|A| = k$ an. Jede k-elementige Menge A mit genau i roten Bällen erhalten wir aber auch, indem wir zunächst i rote Bälle aus R wählen, und anschließend $k - i$ blaue Bälle aus B wählen. Hierfür gibt es für festes i genau $\binom{r}{i} \cdot \binom{b}{k-i}$ Möglichkeiten. Insgesamt also

$$\sum_{i=0}^{k} \binom{r}{i} \cdot \binom{b}{k-i}.$$

(b) Sei X_i die Zufallsvariable, die zählt, wie häufig Person i "Kopf" erhalten hat. Es gilt $X_i \sim \mathrm{Bin}(n; 1/2)$, wobei X_1 und X_2 unabhängig sind. Dann ist gesucht

$$\Pr[X_1 = X_2] = \sum_{i=0}^{n} \Pr[X_1 = i, X_2 = i] = \sum_{i=0}^{n} \Pr[X_1 = i]^2 = \sum_{i=0}^{n} \binom{n}{i}^2 2^{-2n} = 2^{-2n} \sum_{i=0}^{n} \binom{n}{i} \cdot \binom{n}{n-i}.$$

Setzt man $r := b := k := n$ in der Gleichung aus (a), so erhält man:

$$\sum_{i=0}^{n} \binom{n}{i} \cdot \binom{n}{n-i} = \binom{2n}{n}.$$

Es folgt:

$$\Pr[X_1 = X_2] = 2^{-2n} \binom{2n}{n}.$$