Lösung

Diskrete Wahrscheinlichkeitstheorie – Hausaufgabe 5

Freiwillige Abgabe wegen der Pfingstferien bis zum 30.5. bis 8:30—wird korrigiert aber nicht gewertet!

Alle Antworten sind unter Angabe des Rechenwegs zu begründen, soweit nicht anders gefordert! Fragen gerne im infler-Forum posten :).

Aufgabe 5.1

Nachdem sein Sohn Michel vor Kurzem die Kontonummer der Familie auf 4chan gepostet hat, ist der alleinerziehende Vater Xaver in Geldnöte geraten.

Er beschließt daher sein Leben komplett umzukrempeln, kündigt seinen schlecht bezahlten Job bei der örtlichen Ganztagesbetreung für verhaltensauffällige Jugendliche und widmet seine ganze Energie seiner wahren Bestimmung: Barkeeper!

Xaver plant nun eine Gutscheinaktion, um für seine neue Bar zu werben.

Jeder eingelöste Gutschein kostet ihn selbst dabei 0.5 Euro. Insgesamt möchte er jedoch – zumindest mit hoher W'keit – nicht mehr als 500 Euro verschenken:

Xaver geht davon aus, dass ein Gutschein mit W'keit $p = 0.85$ eingelöst werden wird.

Wie viele Gutscheine n sollte Xaver höchstens verteilen, damit die W'keit, dass ihn die Gutscheinaktion mehr als 500 Euro kostet, höchstens $\frac{1}{100}$ ist?

(a) Bestimmen Sie mithilfe der Markov-Ungleichung ein möglichst großes n.

(b) Bestimmen Sie mithilfe der Chebyshev-Ungleichung ein möglichst großes n.

(c) Leiten Sie (für allgemeine Konstanten) möglichst große untere Schranken für n mithilfe der Chebyshev- und Chernoff-Ungleichungen her. Unter welchen Umständen ist die Chernoff-Schranke besser als die Schranke aus der Chebyshev-Ungleichung?

Lösung: Sei X die ZV, welche die eingelösten Gutscheine zählt. Dann ist $X \sim \mathrm{Bin}(n, p)$ mit $p = 0.85$. Wir sind am Ereignis $[X \geq 1001]$ interessiert.

Zuerst einmal ist $\mathbb{E}[X] = np$ und $\mathrm{Var}[X] = npq$ $(q = 0.15)$.

(a) Markov: $\Pr[X \geq t] \leq \frac{\mathbb{E}[X]}{t}$. Dies liefert allgemein für binomialverteiltes X: $\Pr[X \geq t] \leq \frac{np}{t}$ Wir wollen, dass die W'keit $\Pr[X \geq t]$ höchstens $1/100$ ist; das ist sicher der Fall falls $\frac{np}{t} \leq \frac{1}{100}$. Dies gilt genau dann, wenn $n \leq \frac{t}{100p}$. Wenn wir $t = 1001$ und $p = 0.85$ einsetzen erhalten wir die (völlig unbrauchbare!) Schranke $n \leq 11.77\ldots$ also würden wir $n = 11$ wählen.

(b) Wir schätzen allgemein $\Pr[X \geq t]$ für binomialverteiltes X durch die Chebyshev-Ungleichung ab. Dazu müssen wir den abzuschätzenden Term erstmal auf die richtige Form bringen um die Ungleichung anwenden zu können! (Bem: als Abkürzung setzen wir $\mu := \mathbb{E}[X] = np$)

$$\Pr[X \geq t] = \Pr[X - \mu \geq t - \mu] \leq \Pr[|X - \mu| \geq t - \mu]$$

Falls $t - \mu > 0$, also $n < \frac{t}{p}$ können wir die Chebyshev-Ungleichung anwenden:

$$\Pr[|X - \mu| \geq t - \mu] \leq \frac{\mathrm{Var}[X]}{(t - \mu)^2}$$

Um auf der sicheren Seite zu sein, wählen wir also n so, dass $\frac{\mathrm{Var}[X]}{(t-\mu)^2} \leq \frac{1}{100}$.

$$\frac{\mathrm{Var}[X]}{(t - \mu)^2} \leq \frac{1}{100} \Leftrightarrow \frac{npq}{(t - np)^2} \leq \frac{1}{100}$$

$$\Leftrightarrow p^2 \cdot n^2 - (2pt + 100pq)n + t^2 \geq 0$$

$$\Leftarrow n \leq \frac{(2pt + 100pq) - \sqrt{(2pt + 100pq)^2 - 4p^2t^2}}{2p^2}$$

Wenn wir nun $t = 1001, p = 0.85, q = 0.15$ einsetzen erhalten wir $n \leq 1042.04\ldots$ also wählen wir $n = 1042$.

Xaver kann also nach seinem Modell bis zu 1042 Gutscheine verteilen, ohne in Gefahr zu laufen (d.h. mit einer W'keit ≤ 0.01) mehr als 500 Euro auszugeben.

Bemerkung: Die quadratische Ungleichung $p^2 \cdot n^2 - (2pt + 100pq)n + t^2 \geq 0$ hat eigentlich noch die Lösung $n \geq \frac{(2pt+100pq)+\sqrt{(2pt+100pq)^2-4p^2t^2}}{2p^2}$, was $n \geq 1330.0\ldots$ liefert. Aber diese Lösung ist ein Artefakt und schon dadurch ausgeschlossen, dass die Chebyshev-Ungleichung nur anwendbar ist, falls $n < \frac{t}{p} = 1177.6\ldots$.

(c) Für Chebyshev (wie in b): Sei X binomialverteilt, also ist $\mathbb{E}[X] =: \mu = np$ und $\mathrm{Var}[X] = npq$.

$$\Pr[X \geq t] = \Pr[X - \mu \geq t - \mu] \leq \Pr[|X - \mu| \geq t - \mu] \leq \frac{\mathrm{Var}[X]}{(t-\mu)^2} \overset{!}{\leq} \varepsilon$$

$$\frac{\mathrm{Var}[X]}{(t-\mu)^2} \leq \varepsilon \Leftrightarrow \frac{(t-\mu)^2}{npq} \geq \frac{1}{\varepsilon} \Leftrightarrow p^2n^2 - (2pt + \frac{1}{\varepsilon}pq)n + t^2 \geq 0$$

$$\Leftarrow n \leq \frac{(2pt + \frac{1}{\varepsilon}pq) - \sqrt{(2pt + \frac{1}{\varepsilon}pq)^2 - 4p^2t^2}}{2p^2}$$

Für Chernoff: $\Pr[X \geq t] = \Pr[X \geq (1+\delta)\mu]$ für $(1+\delta)\mu = t$ also $\delta = \frac{t}{\mu} - 1$.

Also:

$$\Pr[X \geq t] \leq e^{-\frac{1}{3}\mu\delta^2} \overset{!}{\leq} \varepsilon$$

$$e^{-\frac{1}{3}\mu\delta^2} \leq \varepsilon \Leftrightarrow \mu\delta^2 \geq -3\log\varepsilon$$

Beide Seiten mit μ multiplizieren und Ausrechnen von δ^2 ergibt:

$$\Leftrightarrow t^2 - 2t\mu + \mu^2 \geq -3\log\varepsilon\mu$$

$$\Leftrightarrow p^2n^2 - (2tp - 3\log\varepsilon p)n + t^2 \geq 0$$

$$\Leftarrow n \leq \frac{(2pt - 3\log\varepsilon p) - \sqrt{(2pt - 3\log\varepsilon p)^2 - 4p^2t^2}}{2p^2}$$

Rechenbeispiel (Xavers Bar, wie oben für $p = 0.85, t = 1001$):

- $\varepsilon = 1/100$: Mit Chebyshev $n \leq 1042$, mit Chernoff $n \leq 1047$.
- $\varepsilon = 1/1000$: Chebyshev $n \leq 801$ (sinnlose Schranke!—selbst wenn jeder dieser 801 Gutscheine eingelöst wird bleiben Xavers Kosten unter 500 Euro), Chernoff dagegen liefert $n \leq 1019$.

Chernoff ist also (falls X eine Summe von unabhängigen, nicht notwendigerweise gleich verteilten, Bernoulli-ZV ist) die bessere Wahl, dafür ist Chebyshev für beliebige ZV anwendbar!

Aufgabe 5.2

Sei X eine diskrete Zufallsvariable mit Wertemenge $W_X = \{1, 2, \ldots\}$. Wir schreiben kurz p_i für $\Pr[X = i]$ und nehmen an, dass $0 < p_1 < 1$ gilt.

Weiterhin sei X gedächtnislos, d.h., es gelte

$$\Pr[X > k + l \mid X > k] = \Pr[X > l]$$

für all $k, l \in \mathbb{N}_0$.

Zeigen Sie, dass X geometrisch mit Parameter p_1 verteilt ist.

Lösung: Wegen der Gedächtnislosigkeit gilt:

$$\Pr[X > k+1] = \Pr[X > k+1 \mid X > k] \Pr[X > k] = \Pr[X > 1] \Pr[X > k] = (1-p_1)\Pr[X > k]\,.$$

Induktiv folgt $\Pr[X > k] = (1-p_1)^k$, da für $k = 0$ offensichtlich $\Pr[X > 0] = 1$ gilt.

Daher:

$$\Pr[X = k] = \Pr[X > k-1] - \Pr[X > k] = (1-p_1)^{k-1} - (1-p_1)^k = (1-p_1)^{k-1}(1-1+p_1) = p_1(1-p_1)^{k-1}.$$

Aufgabe 5.3

Wir betrachten folgendes Spiel: Zu Beginn befinden sich r_0 rote Kugeln und b_0 blaue Kugeln in einer Urne. In jedem Zug wird zufällig eine Kugel gezogen *und wieder zurückgelegt.* Ist sie rot, so werfen wir eine *neue* rote Kugel in die Urne, ist sie blau so werfen wir eine *neue* blaue Kugel hinzu. Das Spiel läuft n Züge lang.

Wir modellieren die Ergebnisse des Spieles als Wörter der Länge n über dem Alphabet $\{r, b\}$, also sei $\Omega = \{r, b\}^n$.

Seien im Folgenden $n = 4$, $r_0 = 3$, $b_0 = 2$.

(a) Sei $X_i : \Omega \to \{0, 1\}$ die ZV, welche angibt, ob im i-ten Zug eine rote Kugel gezogen wurde. Berechnen Sie $\Pr[X_2 = 1]$.

(b) Sei R_i die ZV welche die Anzahl der roten Bälle in der Urne nach i Zügen zählt (also $R_0(\omega) = r_0$).

Bestimmen Sie die Dichte der ZV R_2.

(c) Berechnen Sie $\mathbb{E}[R_3]$ (*Hinweis:* $R_3 = R_2 + X_3$).

(d) Bestimmen Sie für allgemeine Parameter n, r_0, b_0 die Dichte von R_k.

Lösung:

(a) Mit dem Satz der totalen W'keit:

$$\Pr[X_2 = 1] = \Pr[X_2 = 1 | X_1 = 1] \Pr[X_1 = 1] + \Pr[X_2 = 1 | X_1 = 0] \Pr[X_1 = 0] = \frac{4}{6}\frac{3}{5} + \frac{3}{6}\frac{2}{5} = \frac{3}{5}$$

(b) Der Wertebereich von R_2 ist $W_{R_2} = \{3, 4, 5\}$.

$$\begin{aligned}
\Pr[R_2 = 3] &= \Pr[X_1 = 0, X_2 = 0] = \tfrac{2}{5}\tfrac{3}{6} = \tfrac{3}{15} \\
\Pr[R_2 = 4] &= \Pr[X_1 = 0, X_2 = 1] + \Pr[X_1 = 1, X_2 = 0] = \tfrac{2}{5}\tfrac{3}{6} + \tfrac{3}{5}\tfrac{2}{6} = \tfrac{6}{15} \\
\Pr[R_2 = 5] &= \Pr[X_1 = 1, X_2 = 1] = \tfrac{3}{5}\tfrac{4}{6} = \tfrac{2}{5}
\end{aligned}$$

(c)

$$\mathbb{E}[R_3] = \mathbb{E}[R_2] + \mathbb{E}[X_3] = \mathbb{E}[R_2] + \Pr[X_3 = 1] =$$

$$\frac{9+24+30}{15} + \Pr[X_3 = 1 | R_2 = 3] \Pr[R_2 = 3] + \Pr[X_3 = 1 | R_2 = 4] \Pr[R_2 = 4] + \Pr[X_3 = 1 | R_2 = 5] \Pr[R_2 = 5] =$$

$$\frac{63}{15} + \frac{3}{7}\frac{3}{15} + \frac{4}{7}\frac{6}{15} + \frac{5}{7}\frac{2}{5} = 24/5 \;(= 4.8)$$

(d) Wir suchen also $\Pr[R_n = r_0 + k]$ für alle $k \in \{0, \ldots, n\}$.

Am besten man zeichnet sich die Zustände (r, b) des Spiels als Markov-Diagramm auf. "Startzustand" ist (r_0, b_0) und alle Zustände in Entfernung k des Startzustands enthalten $r_0 + b_0 + k$ Kugeln. Die W'keit von (r, b) nach $(r+1, b)$ zu gehen ist $\frac{r}{r+b}$ (analog die W'keit nach $(r, b+1)$).

Aus dem Markov-Diagram kann man auch schon die gesuchte W'keit praktisch "ablesen" (formal könnte man das z.B. per Induktion nach n zeigen).

Es gibt $\binom{n}{k}$ Pfade vom Zustand (r_0, b_0) zu $(r_0 + k, b_0 + (n-k)$. Jeder dieser Pfade hat die W'keit (setze zur Abkürzung $N = r_0 + b_0$)

$$\frac{r_0 \cdot (r_0+1) \cdots (r_0+k-1) \cdot b_0 \cdot (b_0+1) \cdots (b_0+(n-k-1))}{N \cdot (N+1) \cdots (N+n-1)}.$$

Daher:

$$\Pr[R_n = r_0 + k] = \frac{\binom{n}{k}(r_0+k-1)!(b_0+n-k-1)!}{(r_0-1)!(b_0-1)!} \cdot \frac{(N-1)!}{(N+n-1)!}$$

Aufgabe 5.4

Es sei S_n die Menge aller Permutationen der Menge $[n]$. Ein $x \in [n]$ heißt Fixpunkt einer Permutation $\pi \in S_n$, falls $\pi(x) = x$. Sei $\Omega = S_n$ mit $\Pr[\pi] = \frac{1}{|S_n|} = \frac{1}{n!}$ $(\forall \pi \in \Omega)$. Sei X die ZV, welche die Anzahl der Fixpunkte einer Permutation $\pi \in \Omega$ zählt.

Beispiel: Für $\pi = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 4 & 3 & 5 & 2 \end{pmatrix}$ ist $X(\pi) = 2$, da 1 und 3 Fixpunkte von π sind: $\pi(1) = 1$ und $\pi(3) = 3$.

(a) Sei n fest. Für $S \subseteq [n]$ sei A_S das Ereignis, dass zumindest alle $x \in S$ Fixpunkte sind, d.h.

$$A_S = \{\pi \in \Omega \mid \forall x \in S\colon \pi(x) = x\}.$$

Zeigen Sie mit Hilfe der Siebformel, dass $\Pr[X_n = 0] = \sum_{k=0}^{n} (-1)^{k+1} \frac{1}{k!}$.

(b) Zeigen Sie, dass $\Pr[X_n = k] = \frac{1}{k!} \cdot \Pr[X_{n-k} = 0]$ für $0 \leq k \leq n$.

(c) Gegen welche Werte konvergieren der Erwartungswert und die Varianz von X_n für $n \to \infty$?

Lösung:

(a) Es gilt:

- $\Pr[X_n > 0] = \Pr\left[\bigcup_{x\in[n]} A_{\{x\}}\right]$.
- $A_S = \bigcap_{x\in S} A_{\{x\}}$.
- $|A_S| = (n - |S|)!$.
- $\Pr[A_S] = \frac{(n-|S|)!}{n!}$.

Daher:

$$\Pr[X_n > 0] = \sum_{k=1}^{n} (-1)^{k+1} \sum_{S\subseteq[n],|S|=k} \Pr[A_S] = \sum_{k=1}^{n} (-1)^{k+1} \binom{n}{k} \frac{(n-k)!}{n!} = \sum_{k=1}^{n} \frac{(-1)^{k+1}}{k!}.$$

Also:

$$\Pr[X_n = 0] = 1 - \sum_{k=1}^{n} (-1)^{k+1} \frac{1}{k!} = \sum_{k=0}^{n} \frac{(-1)^k}{k!}.$$

(b) Man wählt zunächst genau die k Fixpunkte ($\binom{n}{k}$) und wählt dann eine beliebige fixpunktfreie Permutation der verbleibenden $n - k$ Elemente.

Somit: $|[X_n = k]| = \binom{n}{k} \cdot |[X_{n-k} = 0]|$.

Bzw: $\Pr[X_n = k] = \frac{\binom{n}{k}\cdot(n-k)!\cdot\Pr[X_{n-k}=0]}{n!} = \frac{1}{k!} \cdot \Pr[X_{n-k} = 0]$.

(c) Es gilt $e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}$.

Somit $\Pr[X_n = k] = \frac{1}{k!} \sum_{j=0}^{n-k} \frac{(-1)^j}{j!} \xrightarrow{n\to\infty} \frac{e^{-1}}{k!}$, was gerade die Dichte einer Po(1) verteilten ZV ist.

Also konvergieren $\mathbb{E}[X_n]$ und $\mathrm{Var}[X_n]$ gegen 1.