Koordinator
Konstantin Rusch

Wahrscheinlichkeit & Statistik

Lösungen 10

1. Ein grosses Kaufhaus möchte herausfinden, welcher Anteil seiner Kunden schon mindestens ein Mal Ladendiebstahl begangen hat. Dazu werden n Kunden anonym befragt. Da anzunehmen ist, dass trotz Anonymität viele Menschen dies nicht gerne zugeben, wird der Test folgendermassen modifiziert. Jeder der Befragten wirft (unabhängig von den anderen Befragten) geheim eine (faire) Münze und anwortet nur dann wahrheitsgetreu, falls die Münze Kopf zeigt. Wenn die Münze Zahl zeigt, antwortet der Befragte in jedem Fall mit "Nein". Sei p die Wahrscheinlichkeit, dass eine typische Person mindestens einen Ladendiebstahl begangen hat. Ferner sei S_n die Anzahl der Nein-Antworten unter den n befragten Personen.

a) Wie gross ist die Wahrscheinlichkeit, dass eine unter den Befragten beliebig ausgewählte Person mit "Nein" antwortet?

b) Wie gross ist die mittlere Anzahl der Nein-Antworten bei $n = 1000$ Befragten?

c) Wir nehmen an, dass $p = 20\%$. Benutze ein Resultat aus der Vorlesung über Chernoff-Schranken, um eine obere Schranke für die Wahrscheinlichkeit zu finden, dass bei 1000 befragten Kunden mindestens 909 mit "Nein" antworten.

d) Berechne die Wahrscheinlichkeit aus **c)** approximativ.

Lösung:

a) Für $i = 1, \dots, n$ sei K_i das Ereignis, dass die Münze des i-ten Befragten Kopf zeigt und D_i das Ereignis, dass der i-te Befragte schon mindestens einmal einen Ladendiebstahl begangen hat. Wir definieren

$$X_i := I_{K_i^c} + I_{D_i^c \cap K_i}\,;$$

d.h. $X_i = 1$ genau dann, wenn der i-te Befragte mit "Nein" antwortet und $X_i = 0$ genau dann, wenn der i-te Befragte mit "Ja" anwortet. Da die Ereignisse $(D_i, K_i)_{i=1,\dots,n}$ unabhängig sind, sind auch die Zufallsvariablen $X_1, \dots, X_n$ unabhängig. Ferner ist jedes X_i $Be(r)$-verteilt mit

$$r = P[X_i = 1] = 1 - P[X_i = 0] = 1 - P[D_i \cap K_i] = 1 - P[D_i]P[K_i] = 1 - \frac{p}{2}.$$

Die X_i, $i = 1, \ldots, n$, sind also *i.i.d.* $\sim Be(r)$. Insbesondere ist die Wahrscheinlichkeit, dass eine beliebig ausgewählte Person mit "Nein" anwortet gleich $P[X_i = 1] = 1 - \frac{p}{2}$.

b) Gesucht ist $E[S_{1000}]$. Die Linearität des Erwartungswertes liefert

$$E[S_{1000}] = E\left[\sum_{i=1}^{1000} X_i\right] = \sum_{i=1}^{1000} E[X_i] = 1000\left(1 - \frac{1}{2}p\right).$$

c) Gesucht ist eine obere Schranke für $P[S_{1000} \geq 909]$. Da $X_1, \ldots, X_{1000}$ unabhängige Bernoulli-Zufallsvariablen sind (vgl. **a)**), können wir Satz 5.7 aus der Vorlesung anwenden. Zunächst müssen wir $\delta > 0$ so bestimmen, dass $(1+\delta)E[S_{1000}] = 909$. Mit **b)** und $p = 0.2$ erhält man aus dieser Gleichung $\delta = 0.01$. Satz 5.7 gibt dann die Chernoff-Schranke

$$P[S_{1000} \geq 909] = P[S_{1000} \geq (1+\delta)E[S_{1000}]] \leq \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right)^{E[S_{1000}]} \approx 0.956.$$

Die Wahrscheinlichkeit, dass mindestens 909 Personen mit "Nein" antworten ist also höchstens 95.6%.

d) Gesucht ist eine Approximation für $P[S_{1000} \geq 909]$. Wegen $S_{1000} \sim \mathrm{Bin}(1000, r)$ gilt $S_{1000} \overset{\text{approx}}{\sim} \mathcal{N}(1000\,r, 1000(1-r)r)$ und mit der Kontinuitätskorrektur (siehe Skript Seite 117) gilt

$$\begin{aligned} P[S_{1000} \geq 909] &= P\left[\frac{S_{1000} - 1000r}{\sqrt{1000r(1-r)}} \geq \frac{909 - 1000r}{\sqrt{1000r(1-r)}}\right] \\ &\approx 1 - \Phi\left(\frac{909 - \frac{1}{2} - 1000r}{\sqrt{1000r(1-r)}}\right). \end{aligned}$$

Für $r = 0.9$ folgt somit, dass

$$P[S_{1000} \geq 909] \approx 1 - \Phi\left(\frac{909 - \frac{1}{2} - 900}{\sqrt{90}}\right) = 1 - \Phi(0.8959787) \approx 0.185.$$

2. Die erwartete Lebensdauer μ eines Batterietyps ist unbekannt und soll durch das arithmetische Mittel der Lebensdauern von n unabhängigen Testbatterien dieses Typs geschätzt werden. Erfahrungsgemäss ist die Standardabweichung der Lebensdauer ungefähr 5 Stunden. Wie gross muss n ungefähr sein, damit der Absolutbetrag der Differenz zwischen arithmetischem Mittel und μ mit einer Wahrscheinlichkeit von mindestens 99% höchstens 1 Stunde beträgt?

Tipp: Zentraler Grenzwertsatz.

Lösung: Sei L_i die Lebensdauer der i-ten Batterie mit erwarteter Lebensdauer $E[L_i] = \mu$ und Streuung $\sigma_{L_i} = 5$. Für n Batterien ist das arithmetische Mittel der Lebensdauern $L_1, \ldots, L_n$ definitionsgemäss die Zufallsvariable

$$\overline{L}_n := \frac{1}{n}\sum_{i=1}^{n} L_i \qquad \text{mit} \qquad E[\overline{L}_n] = \mu \qquad \text{und} \qquad \sigma_{\overline{L}_n} = \frac{5}{\sqrt{n}}.$$

 Siehe nächste Seite!

Gesucht ist das kleinste n für welches

$$P[|\overline{L}_n - \mu| \leq 1] \geq 0.99 \tag{1}$$

gilt. Wir rechnen zunächst die linke Seite aus, d.h.

$$\begin{aligned} P[|\overline{L}_n - \mu| \leq 1] &= P[-1 \leq \overline{L}_n - \mu \leq 1] = P[\mu - 1 \leq \overline{L}_n \leq \mu + 1] \\ &= F_{\overline{L}_n}[\mu + 1] - F_{\overline{L}_n}[\mu - 1]. \end{aligned}$$

Wegen des zentralen Grenzwertsatzes nehmen wir nun an, dass $\overline{L}_n$ normalverteilt ist (mit Mittelwert μ und Streuung $\frac{5}{\sqrt{n}}$, wie oben ausgerechnet) und erhalten

$$\begin{aligned} &P[|\overline{L}_n - \mu| \leq 1] = F_{\overline{L}_n}[\mu + 1] - F_{\overline{L}_n}[\mu - 1] \\ &\approx \Phi\left[\frac{\mu + 1 - \mu}{\frac{5}{\sqrt{n}}}\right] - \Phi\left[\frac{\mu - 1 - \mu}{\frac{5}{\sqrt{n}}}\right] = \Phi\left[\frac{\sqrt{n}}{5}\right] - \Phi\left[-\frac{\sqrt{n}}{5}\right] \\ &= \Phi\left[\frac{\sqrt{n}}{5}\right] - \left[1 - \Phi\left[\frac{\sqrt{n}}{5}\right]\right] = 2\Phi\left[\frac{\sqrt{n}}{5}\right] - 1. \end{aligned}$$

Bedingung (1) lautet nun

$$\Phi\left[\frac{\sqrt{n}}{5}\right] \approx 0.995\ ,$$

aus der Tabelle der Standardnormalverteilung lesen wir das 0.995-Quantil ab und erhalten $\frac{\sqrt{n}}{5} \approx \Phi^{-1}[0.995] = 2.58$ und daraus $n \approx [2.58 \cdot 5]^2 = 166.41$. Somit muss $n \geq 167$ sein.

3. Eine Brauerei verkauft einem Getränkeverteiler einen grossen Stock Bierfässer. Der Inhalt I jedes Fasses sei gegeben durch $I = 25X$ Liter, wobei die Zufallsvariable $X \sim \text{Beta}(a = 100,\ b = 2)$ verteilt ist. Die Dichte von X ist demzufolge gegeben durch

$$f(x) = cx^{a-1}(1 - x)^{b-1}, \qquad 0 \leq x \leq 1,$$

mit Parametern $a = 100$, $b = 2$ und Normierungskonstante $c = 10100$. (Beachte, dass X nur Werte im Bereich $[0, 1]$ annehmen kann.) Weiter sind die Inhalte der Fässer unabhängig voneinander.

a) Wieviele Liter Bier sind in 120 Fässern mit einer Wahrscheinlichkeit von 95% mindestens enthalten? Rechne mit einer geeigneten Approximation.

Ein Teil der Fässer ist nicht gut verschlossen, und während der Lagerung wird das enthaltene Bier schlecht. Die Wahrscheinlichkeit, dass ein Fass unabhängig von den anderen und unabhängig vom Inhalt nicht gut verschlossen ist, sei 2%.

b) Wieviele Liter gutes Bier kann der Getränkeverteiler erwarten, wenn er 120 Fässer kauft? Mit welcher Varianz und Standardabweichung?

c) Wieviele Fässer soll der Getränkeverteiler kaufen, um mit einer Wahrscheinlichkeit von 95% mindestens 3000 Liter gutes Bier zu haben? Es genügt, die Ungleichung für die Anzahl Fässer anzugeben.

Lösung:

a) Gesucht ist x, so dass

$$P\left[\sum_{i=1}^{120} I_i > x\right] = 0.95\,.$$

Aus dem zentralen Grenzwertsatz folgt

$$\begin{aligned} 0.05 &= 1 - P\left[\sum_{i=1}^{120} I_i > x\right] = P\left[\sum_{i=1}^{120} I_i \le x\right] \\ &= P\left[\frac{\sum_{i=1}^{120} I_i - 120\mathbb{E}[I]}{\sqrt{120\mathrm{var}[I]}} \le \frac{x - 120\mathbb{E}[I]}{\sqrt{120\mathrm{var}[I]}}\right] \overset{ZGS}{\approx} \Phi\left(\frac{x-120\mathbb{E}[I]}{\sqrt{120\mathrm{var}[I]}}\right) \end{aligned}$$

Somit gilt

$$\frac{x - 120\mathbb{E}[I]}{\sqrt{120\mathrm{var}[I]}} \approx -1.645\,,$$

wobei $-1.645 =$ 5%-Quantil der Standard Normalverteilung. Folglich erhalten wir $x \approx 2935\,\ell$.

b) Seien G, G_i, $i = 1, \ldots, 120$, unabhängige Bernoulli-verteilte Zufallsvariablen mit Erfolgsparameter $p = 0.98$. Es gilt $\mathbb{E}[G] = \mathbb{E}[G^2] = p = 0.98$.

Die Anzahl der Liter gutes Bier ist gegeben durch $\sum_{i=1}^{120} G_i I_i$.

Für den Erwartungswert gilt

$$\begin{aligned} \mathbb{E}\left[\sum_{i=1}^{120} G_i I_i\right] &= \sum_{i=1}^{120} \mathbb{E}[G_i I_i] \overset{G_i, I_i\ unab.}{=} \sum_{i=1}^{120} \mathbb{E}[G_i]\mathbb{E}[I_i] \\ &= 120\mathbb{E}[G]\mathbb{E}[I] = 2882.35\,\ell\,. \end{aligned}$$

Für die Varianz hat man

$$\mathrm{var}\left[\sum_{i=1}^{120} G_i I_i\right] \overset{G_i I_i\ unab.}{=} \sum_{i=1}^{120} \mathrm{var}[G_i I_i] = 120\mathrm{var}[GI]\,.$$

Nun ist

$$\begin{aligned} \mathrm{var}[GI] &= \mathbb{E}[G^2 I^2] - \mathbb{E}[GI]^2 \\ &\overset{G,I\ unab.}{=} \mathbb{E}[G^2]\mathbb{E}[I^2] - \mathbb{E}[G]^2\mathbb{E}[I]^2 \\ &= \mathbb{E}[G^2]\left(\mathrm{var}[I] + \mathbb{E}[I]^2\right) - \mathbb{E}[G]^2\mathbb{E}[I]^2 = 11.89\,\ell^2\,. \end{aligned}$$

Somit wird

$$\mathrm{var}\left[\sum_{i=1}^{120} G_i I_i\right] = 120 \cdot 11.89 = 1426.8\,\ell^2 \approx (37.77\,\ell)^2,$$

d.h. die Standardabweichung ist ungefähr $37.77\,\ell$.

c) Gesucht ist n, so dass

$$P\left[\sum_{i=1}^{n} G_i I_i > 3000\right] = 0.95\,.$$

Aus dem zentralen Grenzwertsatz folgt

$$\begin{aligned}
0.05 &= 1 - P\left[\sum_{i=1}^{n} G_i I_i > 3000\right] = P\left[\sum_{i=1}^{n} G_i I_i \leq 3000\right] \\
&= P\left[\frac{\sum_{i=1}^{n} G_i I_i - n\mathbb{E}[GI]}{\sqrt{n\mathrm{var}[GI]}} \leq \frac{3000 - n\mathbb{E}[GI]}{\sqrt{n\mathrm{var}[GI]}}\right] \overset{ZGS}{\approx} \Phi\left(\frac{3000 - n\mathbb{E}[GI]}{\sqrt{n\mathrm{var}[GI]}}\right)\,.
\end{aligned}$$

Somit gilt approximativ

$$\frac{3000 - n\mathbb{E}[GI]}{\sqrt{n\mathrm{var}[GI]}} \approx -1.645\,,$$

und

$$24.02\,n - 5.67\sqrt{n} - 3000 \geq 0\,.$$

Daraus folgt $n \geq 128$.

4. Welche der untenstehenden Aussagen sind **richtig**?

a) Sei X eine Zufallsvariable mit Dichte

$$f(x) = \begin{cases} 6x^{-7} & \text{für } x \geq 1, \\ 0 & \text{sonst.} \end{cases}$$

Ferner sei $Y = \log X$. Dann gilt:

1. Y ist exponentialverteilt mit Parameter 7.
2. Y ist exponentialverteilt mit Parameter 6.
3. Weder 1. noch 2. trifft zu.

b) Sei $(X_i)_{i\in\mathbb{N}}$ eine Folge von Zufallsvariablen, die alle den gleichen Erwartungswert $E[X_i] = \mu$ und die gleiche Varianz $\mathrm{Var}[X_i] = \sigma^2$ haben. Sei $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. Dann gilt:

1. $\bar{X}_n$ konvergiert für $n \to \infty$ in Wahrscheinlichkeit gegen $\mu = E[X_i]$.
2. $\bar{X}_n$ konvergiert für $n \to \infty$ P-fastsicher gegen $\mu = E[X_i]$.
3. Im Allgemeinen gilt weder 1. noch 2.

c) Sei $(X_i)_{i\in\mathbb{N}}$ eine Folge von unabhängigen identisch verteilten Zufallsvariablen mit gemeinsamer Verteilungsfunktion F. Dann ist die Verteilungsfunktion F_X von $X := \max(X_1, \ldots, X_n)$

1. $F_X(x) = (1 - F(x))^n$.

2. $F_X(x) = 1 - F(x)^n$.
3. $F_X(x) = F(x)^n$.

d) Sei $(X_i)_{i\in\mathbb{N}}$ eine Folge von unabhängigen identisch verteilten Zufallsvariablen mit Erwartungswert $E[X_i] = \mu$ und Varianz $\mathrm{Var}[X_i] = \sigma^2$. Sei $S_n := \sum_{i=1}^n X_i$; dann gilt für grosse n

1. $S_n \overset{\text{approx.}}{\sim} \mathcal{N}(\mu, \frac{\sigma^2}{n})$.
2. $S_n \overset{\text{approx.}}{\sim} \mathcal{N}(n\mu, n\sigma^2)$.
3. $S_n \overset{\text{approx.}}{\sim} \mathcal{N}(n\mu, \sqrt{n}\sigma^2)$.

e) Sei Z eine Zufallsvariable mit endlicher Varianz. Für jedes $b > 0$ gilt dann $P\big[(Z - E[Z])^2 \geq\ b\big] \leq \frac{\mathrm{var}[Z]}{b}$.

1. Die Aussage ist richtig.
2. Die Aussage ist falsch.
3. Die Aussage ist nur dann richtig, wenn zusätzlich $E[Z] = 0$ gilt.

Lösung:

a) 2. ist richtig, denn

$$F_Y(t) = P[\log X \leq t] = P[X \leq e^t] = \int_1^{e^t} 6x^{-7}dx = -x^{-6}\Big|_1^{e^t} = 1 - e^{-6t}.$$

Alternativ folgt 2. auch aus

$$F_Y(t) = P[\log X \leq t] = P[X \leq e^t] = F_X(e^t),$$

und

$$f_Y(t) = \frac{d}{dt}F_X(e^t) = f_x(e^t)e^t = 6e^{-7t}e^t = 6e^{-6t}.$$

b) 3. ist die richtige Anwort, denn damit $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ für $n \to \infty$ in Wahrscheinlichkeit oder sogar P-fastsicher gegen $\mu = E[X_i]$ konvergiert, braucht man im Allgemeinen zusätzliche Bedingungen an die X_i (z.B. Unabhängigkeit). Sind beispielsweise alle X_i identisch (d.h. $X_i(\omega) = X_j(\omega)$ für alle $\omega \in \Omega$ und alle i, j), so ist $\bar{X}_n \equiv X_1$, und das konvergiert nicht gegen $E[X_1]$, sofern nicht X_1 eine Konstante ist.

c) 3. ist die richtige Antwort, weil wegen der Unabhängigkeit gilt

$$\begin{aligned} F_X(x) &= P[\max(X_1, \ldots, X_n) \leq x] = P[X_1 \leq x, \ldots, X_n \leq x] \\ &= P[X_1 \leq x] \cdots P[X_n \leq x] = F(x)^n. \end{aligned}$$

d) 2. ist richtig. Der Zentrale Grenzwertsatz besagt, dass für n gross $\frac{S_n - n\mu}{\sigma\sqrt{n}} \overset{\text{approx.}}{\sim} \mathcal{N}(0,1)$, und daraus folgt $S_n \overset{\text{approx.}}{\sim} \mathcal{N}(n\mu, n\sigma^2)$.

e) 1. ist richtig, denn das ist gerade die Chebyshev-Ungleichung:

$$P\big[(Z - E[Z])^2 \geq\ b\big] = P\big[|Z - E[Z]| \geq\ \sqrt{b}\big] \leq \frac{\mathrm{var}[Z]}{b}.$$