Koordinator
Konstantin Rusch

Wahrscheinlichkeit & Statistik

Lösungen 8

1. Es seien $\{X_i\}_{1\leq i\leq 50}$ unabhängige und normalverteilte Zufallsvariablen mit Erwartungswert $\mu = 1$ und Standardabweichung $\sigma = 2$. Darüber hinaus sind folgende Zufallsvariablen definiert:

$$S_n = X_1 + X_2 + \cdots + X_n$$

und

$$\bar{X}_n = \frac{1}{n}(X_1 + X_2 + \cdots + X_n) = \frac{S_n}{n}$$

Dabei ist $n = 50$.

a) Bestimmen Sie die Verteilung von S_n sowie $\bar{X}_n$.

b) Berechnen sie die Wahrscheinlichkeit $P[E[X_1] - 1 \leq X_1 \leq E[X_1] + 1]$.

c) Berechnen Sie $P[E[S_n] - 1 \leq S_n \leq E[S_n] + 1]$.

d) Berechnen Sie $P[E[\bar{X}_n] - 1 \leq \bar{X}_n \leq E[\bar{X}_n] + 1]$.

Lösung:

a) Für S_n gilt

$$E[S_n] = E\left[\sum_{i=1}^n X_i\right] = \sum_{i=1}^n E[X_i] = 50 \cdot 1 = 50,$$
$$\mathrm{Var}[S_n] = \mathrm{Var}\Big[\sum_{i=1}^n X_i\Big] = \sum_{i=1}^n \mathrm{Var}[X_i] = 50 \cdot 4 = 200.$$

S_{50} ist normalverteilt mit $\mathcal{N}(\mu_{S_{50}} = 50, \sigma^2_{S_{50}} = 200)$.
Für $\bar{X}_n$ gilt

$$E[\bar{X}_n] = \frac{1}{n}E\left[\sum_{i=1}^n X_i\right] = 1 \ , \qquad \mathrm{Var}[\bar{X}_n] = \frac{1}{n^2}\mathrm{Var}\left[\sum_{i=1}^n X_i\right] = 0.08$$

$\bar{X}_{50}$ ist normalverteilt mit $\mathcal{N}(\mu_{\bar{X}_{50}} = 1, \sigma^2_{\bar{X}_{50}} = 0.08)$.

b) X_1 ist normalverteilt mit $\mathcal{N}(\mu_{X_1} = 1, \sigma^2_{X_1} = 4)$. Durch Standardisieren erhalten wir $Z \sim \mathcal{N}(0,1)$ wobei $Z = \frac{X_1 - 1}{2}$. Wir erhalten dann

$$\begin{aligned} P[E[X_1] - 1 \leq X_1 \leq E[X_1] + 1] &= P[0 \leq X_1 \leq 2] = P\Big[-\frac{1}{2} \leq Z \leq \frac{1}{2}\Big] \\ &= \Phi\Big[\frac{1}{2}\Big] - \Phi\Big[-\frac{1}{2}\Big] = \Phi\Big[\frac{1}{2}\Big] - \Big[1 - \Phi\Big[\frac{1}{2}\Big]\Big] \\ &= \Phi\Big[\frac{1}{2}\Big] + \Phi\Big[\frac{1}{2}\Big] - 1 = 2 \cdot \Phi\Big[\frac{1}{2}\Big] - 1 \\ &= 2 \cdot 0.6915 - 1 = 0.3830 \end{aligned}$$

c) Durch Standardisieren: für $n = 50$ erhalten wir $Z = \frac{S_{50} - 50}{\sqrt{200}}$.
Es folgt:

$$P[49 \leq S_n \leq 51] = P[-0.07 \leq Z \leq 0.07] = 2 \cdot \Phi[0.07] - 1 = 2 \cdot 0.528 - 1 = 0.056$$

d) Durch Standardisieren: für $n = 50$ erhalten wir $Z = \frac{\bar{X}_{50} - 1}{\sqrt{0.08}}$.
Es folgt:

$$P[0 \leq \bar{X}_{50} \leq 2] = P[-3.5 \leq Z \leq 3.5] = 2 \cdot \Phi[3.5] - 1 = 2 \cdot 0.99977 - 1 = 0.99954$$

2. Seien X_1, X_2, ..., X_n unabhängige und identisch verteilte Zufallsvariablen mit Dichtefunktion f und Verteilungsfunktion F. Definiere die zwei Zufallsvariablen

$$X_{(1)} = \min\{X_1, X_2, \ldots, X_n\} \quad \text{und} \quad X_{(n)} = \max\{X_1, X_2, \ldots, X_n\}.$$

a) Bestimme die Dichtefunktionen $f_{(1)}$ von $X_{(1)}$ und $f_{(n)}$ von $X_{(n)}$.

b) Nehme an, dass X_k exponential verteilt ist mit Parameter $\lambda > 0$. Definiere die Folgen

$$a_n := \frac{\log n}{\lambda} \quad \text{und} \quad b_n := \frac{1}{\lambda} \quad \text{für } n \in \mathbb{N}.$$

Zeige, dass

$$\lim_{n\to\infty} P\left[\frac{X_{(n)} - a_n}{b_n} \leq t\right] = e^{-e^{-t}} \quad \text{für } t > 0.$$

Lösung:

a) Wir berechnen zuerst die Dichtefunktion von $X_{(n)}$. Dazu berechnen wir die Verteilungsfunktion $F_{(n)}$ und leiten diese danach ab.

$$F_{(n)}(t) = P[X_{(n)} \leq t] = P[X_1 \leq t, \ldots, X_n \leq t] = \prod_{k=1}^{n} P[X_k \leq t] = (F(t))^n.$$

Demzufolge erhalten wir

$$f_{(n)}(t) = \frac{d}{dt} F_{(n)}(t) = nF^{n-1}(t) f(t).$$

Analog gehen wir beim Minimum vor.

$$F_{(1)}(t) = 1 - P[X_{(1)} > t] = 1 - P[X_1 > t, \ldots, X_n > t] = 1 - (1 - F(t))^n .$$

Demzufolge erhalten wir

$$f_{(1)}(t) = \frac{d}{dt} F_{(1)}(t) = n\,(1 - F(t))^{n-1}\, f(t).$$

b) Einsetzen ergibt

$$\begin{aligned} P\left[\frac{X_{(n)} - a_n}{b_n} \leq t\right] &= P\left[X_{(n)} \leq a_n + b_n t\right] \\ &= (F(a_n + b_n t))^n \\ &= \left(1 - e^{-\lambda(a_n + b_n t)}\right)^n \\ &= \left(1 - \frac{1}{n} e^{-t}\right)^n . \end{aligned}$$

Das bedeutet, dass

$$\lim_{n\to\infty} P\left[\frac{X_{(n)} - a_n}{b_n} \leq t\right] = \lim_{n\to\infty} \left(1 - \frac{1}{n} e^{-t}\right)^n = e^{-e^{-t}}.$$

3. Die Chebyshev-Ungleichung liefert eine Abschätzung von Wahrscheinlichkeiten, auch wenn die genaue Verteilungsfunktion nicht bekannt ist. Es müssen nur der Erwartungswert μ und die Varianz $\sigma^2 < \infty$ einer Zufallsvariablen X bekannt sein; dann gilt für jedes $k > 0$, dass

$$P[|X - \mu| \geq k] \leq \frac{\sigma^2}{k^2}.$$

Die Anzahl der Turbinen, die pro Woche in einer Fabrik hergestellt werden, sei eine Zufallsvariable mit Erwartungswert $\mu = 50$ und Varianz $\sigma^2 = 25$.

a) Angenommen, es ist nichts weiter über die Zufallsvariable bekannt. Was kann man dann über die Wahrscheinlichkeit aussagen, dass in dieser Woche zwischen 40 und 60 Turbinen hergestellt werden?

b) Nun nehmen wir an, dass zusätzliche Informationen über die Verteilung von X erhältlich sind (dann kann die Wahrscheinlichkeit aus der letzten Teilaufgabe genauer berechnet werden und wir sehen, wie gut die Chebyshev-Abschätzung ist). Nehme an, dass gilt $X \sim \mathcal{N}(\mu, \sigma^2)$. Wie gross ist dann die Wahrscheinlichkeit, dass in dieser Woche zwischen 40 und 60 Turbinen hergestellt werden?

Bemerkung: Die Chebyshev-Ungleichung ist bei vielen theoretischen Überlegungen zentral. Da sie aber unter so allgemeinen Bedingungen gültig ist, sind die angegebenen Schranken für die Wahrscheinlichkeit nicht sehr genau. In der Praxis ist sie daher nur dann sinnvoll, wenn wirklich keine Zusatzinformationen über die Verteilung erhältlich sind.

Lösung:

a) Mit der Chebyshev-Ungleichung gilt

$$P[|X - 50| \geq 10] \leq \frac{\sigma^2}{10^2} = 0.25.$$

Daher ist $P[|X - 50| < 10] \geq 1 - 0.25 = 0.75$.

b) Mit Annahme der Normalverteilung folgt

$$\begin{aligned} P[|X - 50| < 10] &= \\ &= P[40 < X < 60] = P[X < 60] - P[X \leq 40] = \\ &= P\left[\frac{X - 50}{5} < \frac{60 - 50}{5}\right] - P\left[\frac{X - 50}{5} \leq \frac{40 - 50}{5}\right] = \\ &= \Phi(2) - \Phi(-2) = \Phi(2) - (1 - \Phi(2)) = 2\Phi(2) - 1 = \\ &= 2 \cdot 0.97725 - 1 = 0.9545. \end{aligned}$$

Die Chebyshev-Ungleichung liefert also eine korrekte Abschätzung der Wahrscheinlichkeit (0.9545 ist grösser als 0.75), allerdings ist sie nicht sehr genau (0.75 liegt nicht sehr nahe an 0.9545). Dies war allerdings zu erwarten, weil die Chebyshev-Ungleichung für **jede** Zufallsvariable mit endlicher Varianz gilt.

4. Die Krankenkasse easyHealth bietet einen sehr günstigen Studententarif für 160 Franken pro Jahr an. Sie hat pro Kunde einen jährlichen Administrationsaufwand von 100 Franken. Zusätzlich fallen die variierenden Gesundheitskosten (Arztrechnungen etc.) an. Zur Modellierung der *Gesamtkosten* X pro Kunde und Jahr benutzt die Krankenkasse die Dichte

$$f(x) = \begin{cases} 0 & x \leq c, \\ \frac{\alpha c^\alpha}{x^{\alpha+1}} & x > c, \end{cases}$$

wobei $c > 0, \alpha > 1$ Parameter sind.

easyHealth hat zur Zeit 250 studentische Kunden. Was ist (approximativ) die Wahrscheinlichkeit, dass easyHealth mit diesem Tarif in einem Jahr Gewinn macht, wenn wir annehmen, dass $\alpha = 3$, $c = 100$?

Hinweis: Die folgenden Formeln können ohne Herleitung benutzt werden:

$$E[X] = c\frac{\alpha}{\alpha - 1}, \qquad \mathrm{Var}(X) = \frac{c^2\alpha}{(\alpha - 1)^2(\alpha - 2)}.$$

Siehe nächste Seite!

Lösung: Wir definieren die Zufallsvariable G als Differenz aus Einnahmen und Kosten: $G = 250 \cdot 160 - \sum_{i=1}^{250} X_i$, wobei X_i iid Kopien von X sind. Dann gilt mit dem zentralen Grenzwertsatz

$$P(G > 0) = P\left(\sum_{i=1}^{250} X_i < 250 \cdot 160\right) = P\left(\frac{\sum_{i=1}^{250} X_i - 250\mu}{\sqrt{250\sigma^2}} < \frac{250(160-\mu)}{\sqrt{250\sigma^2}}\right)$$
$$\approx \Phi\left(\sqrt{\frac{250}{\sigma^2}}(160-\mu)\right),$$

wobei $\mu = E[X] = 150$ und $\sigma^2 = \mathrm{Var}(X) = 7500$. Somit erhalten wir

$$P(G > 0) \approx \Phi(1.8257) = 0.96.$$