Koordinator
Konstantin Rusch

Wahrscheinlichkeit & Statistik

Lösungen 9

1. Beim Abfüllen von Kartoffeln in 10–kg–Säcke variiert das Einfüllgewicht zwischen 9.750 kg und 10.750 kg. Die Einfüllgewichte sind in diesem Bereich gleichverteilt, Gewichte ausserhalb kommen nicht vor.

a) Wie gross ist approximativ die Wahrscheinlichkeit, dass beim Beladen eines Lieferwagens mit 146 Säcken das zulässige Ladegewicht von 1500 kg überschritten wird?

b) Wieviele Säcke darf man höchstens laden, damit das zulässige Ladegewicht nur mit (approximativer) Wahrscheinlichkeit 1% überschritten wird?

Hinweis: Die Gewichte der einzelnen Säcke können als unabhängig voneinander angenommen werden.

Lösung:

a) Für $i \geq 1$ sei X_i das Einfüllgewicht des i-ten Kartoffelsacks. Die X_i, $i \geq 1$, sind $i.i.d$ mit $X_i \sim \mathcal{U}(9.75, 10.75)$. Sei $S_n = \sum_{i=1}^{n} X_i$. Dann gilt:

$$\begin{aligned} E[X_i] &= 10.25, \qquad \text{also } E[S_{146}] = 146E[X_1] = 146 \cdot 10.25 = 1496.5, \\ \operatorname{Var}(X_i) &= \frac{1}{12}, \qquad \text{also } \operatorname{Var}(S_{146}) = \frac{146}{12} = 12.1\bar{6} \quad \text{und} \quad \sigma(S_{146}) \approx 3.48. \end{aligned}$$

Mit dem Zentralen Grenzwertsatz folgt

$$Z = \frac{S_{146} - E[S_{146}]}{\sqrt{\operatorname{Var}(S_{146})}} \overset{\text{approx.}}{\sim} \mathcal{N}(0,1).$$

Also ist

$$\begin{aligned} P[S_{146} > 1500] &= P\left[\frac{S_{146} - E[S_{146}]}{\sqrt{\operatorname{Var}(S_{146})}} > \frac{1500 - 1496.5}{3.48}\right] \\ &\approx P[Z > 1.01] = 1 - P[Z \leq 1.01] = 1 - \Phi(1.01) \approx 0.156. \end{aligned}$$

b) Die Gleichung $P[S_n > 1500] = 0.01$ ist nach n aufzulösen, d.h.

$$0.01 \overset{!}{=} P[S_n > 1500] = P\left[\frac{S_n - E[S_n]}{\sigma(S_n)} > \frac{1500 - E[S_n]}{\sigma(S_n)}\right] \approx 1 - \Phi\left(\frac{1500 - 10.25n}{\sqrt{n/12}}\right);$$

also möchte man approximativ

$$\frac{1500 - 10.25n}{\sqrt{n/12}} \stackrel{!}{=} \Phi^{-1}(0.99) \approx 2.326.$$

Mit $x = \sqrt{n}$ erhalten wir somit die quadratische Gleichung

$$1500 - 10.25x^2 - \frac{2.326}{\sqrt{12}}x = 0$$

mit Lösungen $x_1 \approx -12.13$ und $x_2 \approx 12.06$ (wegen $x = \sqrt{n}$ ist nur die positive Lösung relevant). Man kann also höchstens $n = 145$ Säcke laden.
Bemerkung: Obwohl man im Vergleich zu **a)** nur einen Sack weniger lädt, reduziert sich die Wahrscheinlichkeit das zulässige Ladegewicht von 1500 kg zu überschreiten von 15.6% auf unter 1%. Dies ist plausibel, denn während der Erwartungswert $E[S_{146}] = 1496.5$ nur etwa eine Standardabweichung $\sigma(S_{146}) \approx 3.48$ unter dem zulässigen Ladegewicht liegt, ist das erwartete Ladegewicht von 145 Säcken $E[S_{145}] = 1486.25$ fast vier Standardabweichungen $\sigma(S_{145}) \approx 3.48$ unterhalb von 1500.

2. **a)** Berechne mit Hilfe des zentralen Grenzwertsatzes

$$\lim_{n\to\infty} e^{-n} \sum_{k=0}^{n} \frac{n^k}{k!}.$$

Hinweis Verwende folgendes Resultat:

Wenn $X \sim \text{Pois}(\lambda)$, $\lambda > 0$, und $Y \sim \text{Pois}(\mu)$, $\mu > 0$, und X und Y sind unabhängig, dann ist $X + Y \sim \text{Pois}(\lambda + \mu)$.

b) Sei $(X_k)_k$ eine Folge von i.i.d. Zufallsvariablen mit Dichtefunktion

$$f(x) = \begin{cases} \frac{1}{6}, & \text{für } 0 \leq x < 2, \\ \frac{1}{3}, & \text{für } 2 \leq x \leq 4, \\ 0, & \text{sonst.} \end{cases}$$

Berechne den (P-f.s.) Grenzwert

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \sqrt{X_k}.$$

Lösung:

a) Betrachte $X_1, X_2, \ldots$ iid mit $X_i \sim \text{Pois}(1)$. Insbesondere gilt $\mu = E[X_1] = 1 = \text{Var}[X_1] = \sigma^2$. Aussderdem ist $S_n = X_1 + \cdots + X_n$ wieder Poisson verteilt mit Parameter n. Wegen des zentralen Grenzwertsatzes folgt

$$\begin{aligned} e^{-n} \sum_{k=0}^{n} \frac{n^k}{k!} &= \sum_{k=0}^{n} e^{-n} \frac{n^k}{k!} = \sum_{k=0}^{n} P[S_n = k] = P[S_n \leq n] \\ &= P\left[\frac{S_n - n\mu}{\sqrt{n}\sigma} \leq \frac{n - n\mu}{\sqrt{n}\sigma}\right] = P\left[\frac{S_n - n\mu}{\sqrt{n}\sigma} \leq 0\right] \stackrel{n\to\infty}{\longrightarrow} \Phi(0) = \frac{1}{2} \end{aligned}$$

b) Nach dem starken Gesetz der grossen Zahl gilt mit Wahrscheinlichkeit 1

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} Y_k = E[Y_1].$$

In unserem Fall, $Y_k = X_k^{1/2}$. D.h. mit Wahrscheinlichkeit 1 gilt

$$\begin{aligned}
\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} X_k^{1/2} &= E[X_1^{1/2}] \\
&= \int_{-\infty}^{\infty} x^{1/2} f(x)dx \\
&= \frac{1}{6}\int_0^2 x^{1/2}dx + \frac{1}{3}\int_2^4 x^{1/2}dx \\
&= \frac{1}{6}\left(\frac{2}{3}x^{3/2}\Big|_0^2\right) + \frac{1}{3}\left(\frac{2}{3}x^{3/2}\Big|_2^4\right) \\
&= \frac{16-2^{3/2}}{9}.
\end{aligned}$$

3. Der **Median** m einer Verteilung F wurde definiert durch $m = F^{-1}(1/2)$. Seien $X_1, X_2, \ldots$ i.i.d. mit Verteilungsfunktion F, Dichtefunktion f, und Median $m = 0$. Ferner soll $f(0) > 0$ sein. Ferner sei Z_n der sogenannte **Stichprobenmedian** von $X_1, \ldots, X_n$, d.h. Z_n ist die mittlere Beobachtung, oder formelmässig $Z_n = X_{(k)}$ mit $k = [\frac{n}{2}+1]$, wobei $X_{(1)} \leq \cdots \leq X_{(n)}$ die der Grösse nach geordneten Zufallsvariablen $X_1, \ldots, X_n$ bezeichnen und $[x]$ den ganzzahligen Teil von x.

a) Seien $Y_i = 1_{\{X_i \leq x\}}$ und $S_n(x) := \sum_{i=1}^n Y_i$. Berechnen Sie $\mathbb{E}[S_n(x)]$ und $\mathrm{Var}[S_n(x)]$.

b) Beschreiben Sie das Ereignis $\{Z_n \leq x\}$ mit Hilfe der Zufallsvariable $S_n(x)$.

c) Geben Sie eine Approximation für $\mathbb{P}(Z_n \leq x)$ als $n \to \infty$ und berechnen Sie die Grenze

$$\lim_{n\to\infty} \frac{1/2 - \alpha_n}{\sqrt{\alpha_n(1-\alpha_n)}/\sqrt{n}},$$

wobei $\alpha_n = F(\frac{x}{\sqrt{n}})$.

Lösung:

a) Da $S_n(x) \sim Bin(n, F(x))$ und $\mathbb{E}[S_n(x)] = nF(x)$, $\mathrm{Var}[S_n(x)] = nF(x)(1-F(x))$.

b) Da

$$Z_n \leq x \Leftrightarrow S_n(x) \geq k, \tag{1}$$

wobei $k = \left[\frac{n}{2}+1\right]$ ist.

c)

$$\mathbb{P}(Z_n \leq x) \overset{(1)}{=} \mathbb{P}\left(\frac{S_n(x)}{n} \geq \frac{k}{n}\right)$$
$$= \mathbb{P}\left(\frac{\frac{S_n(x)}{n} - F(x)}{\sqrt{F(x)(1-F(x))}/\sqrt{n}} \geq \frac{k/n - F(x)}{\sqrt{F(x)(1-F(x))}/\sqrt{n}}\right)$$
$$\approx 1 - \Phi\left(\frac{k/n - F(x)}{\sqrt{F(x)(1-F(x))}/\sqrt{n}}\right) = \Phi(-\frac{k/n - F(x)}{\sqrt{F(x)(1-F(x))}/\sqrt{n}})$$

Wir setzen $\alpha_n := \mathbb{E}[S_n(\frac{x}{\sqrt{n}})] = F\left(\frac{x}{\sqrt{n}}\right)$ und $Var[S_n(\frac{x}{\sqrt{n}})] = \alpha_n(1-\alpha_n)/n$.

$$\begin{aligned} &\lim_{n\to\infty} \frac{1/2 - \alpha_n}{\sqrt{\alpha_n(1-\alpha_n)}/\sqrt{n}} \\ &= \lim_{n\to\infty} -\frac{\alpha_n - 1/2}{x/\sqrt{n}} \frac{x}{\sqrt{\alpha_n(1-\alpha_n)}}. \end{aligned} \tag{2}$$

Beachten Sie, dass

$$\lim_{n\to\infty} \frac{\alpha_n - \frac{1}{2}}{x\sqrt{n}} = \frac{F(\frac{x}{\sqrt{n}}) - F(0)}{x/\sqrt{n}} = F'(0) = f(0).$$

und

$$\lim_{n\to\infty} F(x/\sqrt{n}) = F(0) = \frac{1}{2}.$$

Also ist,

$$\lim_{n\to\infty} \frac{1/2 - \alpha_n}{\sqrt{\alpha_n(1-\alpha_n)}/\sqrt{n}} = -2f(0)x. \tag{3}$$

4. Eine Zufallsvariable X_ν heisst χ^2-verteilt mit Freiheitsgrad $\nu \in$ (geschrieben $X_\nu \sim \chi^2_\nu$), falls

$$X_\nu = \sum_{k=1}^{\nu} Z_k^2.$$

wobei $(Z_k)_k$ i.i.d. $\mathcal{N}(0,1)$-verteilt ist. Das verallgemeinert die χ^2_1-Verteilung aus dem Skript S. 104/105.

a) Zeige, dass

$$E[X_\nu] = \nu \qquad \text{und} \qquad \mathbb{V}[X_\nu] = 2\nu$$

gilt.

b) Gib mit Hilfe der Chebyshev-Ungleichung eine untere Schranke für die Wahrscheinlichkeit

$$\mathbb{P}\left[\left|\frac{X_\nu}{\nu} - 1\right| \leq 0.75\right].$$

Berechnen Sie die Schranke für $\nu = 12$.

c) Berechnen Sie für $\nu = 12$ eine Annäherung für die obige Wahrscheinlichkeit mit Hilfe des zentralen Grenzwertsatzes.

Lösung:

a) Für ganzzahlige Werte von ν kann X_ν geschrieben werden als $X_\nu = \sum_{i=1}^{\nu} Z_i^2$ mit $Z_i \overset{iid}{\sim} \mathcal{N}(0,1)$.

Daraus folgt einerseits

$$E[X_\nu] = \sum_{i=1}^{\nu} E[Z_i^2] = \nu \cdot 1 = \nu.$$

Andererseits ist

$$E[X_\nu^2] = E[(Z_1^2 + Z_2^2 + \cdots + Z_\nu^2)^2] = \nu E[Z_i^4] + \nu(\nu-1)E[Z_i^2].$$

Mit $E[Z_i^2] = 1$ und

$$E[Z_i^4] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^4 e^{-x^2/2}\, dx = \frac{1}{\sqrt{2\pi}} \left[-x^3 e^{-x^2/2}\right]_{-\infty}^{\infty} + \frac{3}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 e^{-x^2/2}\, dx = 3$$

erhalten wir schliesslich $E[X_\nu^2] = 3\nu + \nu(\nu-1) = \nu^2 + 2\nu$, und daraus

$$\mathrm{Var}(X_\nu) = (\nu^2 + 2\nu) - \nu^2 = 2\nu.$$

b) Chebyshev-Ungleichung (Proposition 5.2): $P[|X_\nu - \nu| > c] \leq \mathrm{Var}(X_\nu)/c^2 = \frac{2\nu}{c^2}$

$$\begin{aligned} P\left[\left|\frac{X_\nu}{\nu} - 1\right| \leq 0.75\right] &= P\left[\left|\frac{X_\nu - \nu}{\nu}\right| \leq \frac{3}{4}\right] = 1 - P\left[|X_\nu - \nu| > \frac{3\nu}{4}\right] \\ &\geq 1 - \frac{2\nu}{9\nu^2/16} = 1 - \frac{32}{9\nu} \quad \left(= \frac{19}{27} \approx \mathbf{0.7037} \text{ für } \nu = 12\right) \end{aligned}$$

c) $Y_i \overset{\text{iid}}{\sim} \chi_1^2, \quad E[Y_i] = 1, \quad \mathbb{V}[Y_i] = 2, \quad i = 1, \ldots, \nu.$
Der zentrale Grenzwertsatz für $X_\nu = \sum_{i=1}^{\nu} Y_i$ lautet:

$$Z = \frac{X_\nu - \nu E[Y_i]}{\sqrt{\nu \mathbb{V}[Y_i]}} = \frac{X_\nu - \nu}{\sqrt{2\nu}} \sim N(0,1)$$

$$\begin{aligned} P\left[\left|\frac{X_\nu}{\nu} - 1\right| \leq 0.75\right] &= P\left[\left|\frac{X_\nu - \nu}{\nu}\right| \leq \frac{3}{4}\right] = P\left[\left|\frac{X_\nu - \nu}{\sqrt{2\nu}}\right| \leq \frac{3}{4}\sqrt{\frac{\nu}{2}}\right] \qquad (\nu = 12) \\ &\approx \Phi\left(\frac{3}{4}\sqrt{6}\right) - \Phi\left(-\frac{3}{4}\sqrt{6}\right) = \mathbf{0.9338}. \end{aligned}$$