Wahrscheinlichkeit und Statistik

Lösungen Serie 10

Lösung 10-1. (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:

$$\begin{aligned} E[X_i] &= 10.25, \qquad \text{also } E[S_{146}] = 146E[X_1] = 146 \cdot 10.25 = 1496.5, \\ \mathrm{Var}(X_i) &= \frac{1}{12}, \qquad \text{also } \mathrm{Var}(S_{146}) = \frac{146}{12} = 12.1\bar{6} \text{ und } \sigma(S_{146}) \approx 3.46. \end{aligned}$$

Mit dem Zentralen Grenzwertsatz folgt

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

Also ist

$$\begin{aligned} P[S_{146} > 1500] &= P\left[\frac{S_{146} - E[S_{146}]}{\sqrt{\mathrm{Var}(S_{146})}} > \frac{1500 - 1496.5}{3.49}\right] \\ &\approx P[Z > 1.01] = 1 - P[Z \leq 1.01] = 1 - \Phi(1.01) \approx 0.156. \end{aligned}$$

(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

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

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 $\sigma(S_{146}) \approx 3.46$ unter dem zulässigen Ladegewicht liegt, ist das erwartete Ladegewicht von 145 Säcken $E[S_{145}] = 1486.25$ fast vier Standardabweichungen $\sigma(S_{145}) \approx 3.46$ unterhalb von 1500.

Lösung 10-2. (a) Betrachte $X_1, X_2, \dots$ iid mit $X_i \sim \mathrm{Pois}(1)$. Insbesondere gilt $\mu = E[X_1] = 1 = \mathrm{Var}[X_1] = \sigma^2$. Aussderdem ist $S_n = X_1 + \cdots + X_n$ wieder Poisson verteilt mit Parameter n. Wegen des zentralen Grenzwertsatzes folgt

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

Siehe nächste Seite!

(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, $Y_k = \sqrt{X_k}$. D.h. mit Wahrscheinlichkeit 1 gilt

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

Lösung 10-3. (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): $P[|X_\nu - \nu| > c] \le \mathrm{Var}(X_\nu)/c^2 = \frac{2\nu}{c^2}$

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

Siehe nächstes Blatt!

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

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

$$P\left[\left|\frac{X_\nu}{\nu} - 1\right| \le 0.75\right] = P\left[\left|\frac{X_\nu - \nu}{\nu}\right| \le \frac{3}{4}\right] = P\left[\left|\frac{X_\nu - \nu}{\sqrt{2\nu}}\right| \le \frac{3}{4}\sqrt{\frac{\nu}{2}}\right] \qquad (\nu = 12)$$

$$\approx \Phi\left(\frac{3}{4}\sqrt{6}\right) - \Phi\left(-\frac{3}{4}\sqrt{6}\right) = \mathbf{0.9338}.$$

Lösung Challenge Serie 10. Wir führen zunächst ein wenig Notation ein:

p_A, p_B bezeichne die W'keit, dass Peter/Paula einen Fehler endecken.

n bezeichne die (unbekannte) Anzahl Fehler im Text.

n_A, n_B bezeichne die Anzahl Fehler, die von Peter/Paula tatsächlich gefunden wurden.

n_{AB} bezeichne die Anzahl Fehler, die von beiden gefunden wurden.

Wir präsentieren zwei Lösungswege.

(a) Die erwartete Anzahl Fehler, die Peter/Paula findet ist dann np_A bzw. np_B also ist $\widehat{p}_A = \frac{n_A}{n}$ bzw. $\widehat{p}_B = \frac{n_B}{n}$ ein Schätzer für die Wahrscheinlichkeit p_A bzw. p_B. Die erwartete Anzahl Fehler, die von beiden gefunden wurden, ist dann np_Ap_B. Also erhalten wir mit den Schätzern $\widehat{p}_A$ und $\widehat{p}_B$ die Formel

$$n_{AB} = n\frac{n_A}{n}\frac{n_B}{n} = \frac{n_A n_B}{n}$$

und damit einen Schätzer für n:

$$\widehat{n} = \frac{n_A n_B}{n_{AB}}.$$

Nach Annahme haben Peter und Paula zusammen $n_A + n_B - n_{AB}$ Fehler gefunden. Insgesamt erhalten wir den Schätzer für die *nicht gefundenen Fehler*

$$\widehat{n} - (n_A + n_B - n_{AB}) = \frac{(n_A - n_{AB})(n_B - n_{AB})}{n_{AB}}.$$

Setzen wir die Daten aus der Aufgabe in die obigen Formeln ein, so erhalten wir

$$\widehat{n} = \frac{20 \cdot 15}{10} = 30$$

Fehler insgesamt und 5 die nicht gefunden wurden.

(b) Alternativ kann mit der ML–Methode argumentieren. Dazu nehmen wir an, dass Peter und Paula den Text nacheinander lesen. Peter hat n_A Fehler gefunden und $n - n_A$ Fehler übersehen. Paula findet n_B Fehler. Jeder dieser Fehler kann nun zu den Fehlern gehören, die Peter schon gefunden hat (also zu den n_A) oder zu denen, die er nicht gefunden hat (also zu den $n - n_A$). Die Wahrscheinlichkeit, dass unter den n_B Fehlern, genau n_A in n_{AB} sind und $n - n_A$ in $n_B - n_{AB}$ ist durch dir Hypergeometrische Verteilung gegeben

$$g(n; n_A, n_B, n_{AB}) = \frac{\binom{n_A}{n_{AB}} \cdot \binom{n-n_A}{n_B - n_{AB}}}{\binom{n}{n_B}} \qquad \text{für } n \ge n_A + n_B - n_{AB}.$$

Siehe nächste Seite!

Für gegebene Werte n_A, n_B und n_{AB} können wir die obige Wahrscheinlichkeit in n maximieren. Bildet man den Quotienten $\frac{g(n+1)}{g(n)}$, so erhählt man zwei Maxima, bei $\widehat{n} = 29$ und $\widehat{n} = 30$. Einsetzen liefert dann, dass 4 oder 5 Fehler nicht gefunden wurden.