ETH Zürich HS 2019
RW, D-MATL, D-MAVT
Dr. Christoph Czichowsky

Koordinator
Florian Krach

Stochastik

Interaktive Aufgabe für die Übung 9

1. In einem Sägewerk wird das Schnittholz einer bestimmten Sortierklasse einer Qualitätskontrolle unterzogen. Pro Produktionstag wird eine Stichprobe mit zehn Brettern entnommen und jedes Brett auf seine Steifigkeit getestet. Basierend auf Erfahrung kann angenommen werden, dass die Steifigkeit eines Brettes normalverteilt ist mit bekannter Standardabweichung $\sigma = 1430$ MPa. Qualitätsschwankungen äussern sich nur in Form von Schwankungen des Mittelwertes.

a) Leite die Formel des 95% Vertrauensintervalls für μ nach 15 Produktionstagen her.

b) Berechne aus a) das Vertrauensintervall für einen beobachteten Stichprobenmitterwert von $\overline{x} = 11000$ MPa (nach 15 Produktionstagen).

c) Wieviele Stichproben wären nötig, damit die Breite des Vertrauensintervalls kleiner als 200 ist?

Lösung:

a) Wir wissen dass

$$\overline{X} \sim \mathcal{N}(\mu, \sigma^2/n) \ .$$

Also gilt

$$P\left[-z_{1-\alpha/2} < \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} < z_{1-\alpha/2}\right] = 1-\alpha,$$

wobei $z_{1-\alpha/2} = \Phi^{-1}(1-\frac{\alpha}{2})$ das $(1-\frac{\alpha}{2}) \times 100\%$ Quantil der Standardnormalverteilung ist. Durch Umformen erhalten wir

$$P\left[\overline{X} - z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right] = 1-\alpha.$$

Also hat das Vertrauensintervall die Form

$$\overline{X} \pm z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

mit $\alpha = 0.05$. Das Vertrauensintervall nach 15 Produktionstagen (mit Stichproben von je 10 Brettern) ist

$$\overline{X} \pm z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{150}}.$$

Bitte wenden!

b) Aus a) wissen wir, dass das Vertrauensintervall die Form $\overline{X} \pm z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$ hat. Durch einsetzen erhalten wir $[10771.15, 11228.85]$, wobei wir verwendet haben, dass $\Phi^{-1}(1 - \frac{\alpha}{2}) = 1.96$ für $\alpha = 0.05$.

c) Die Breite des Vertrauensintervalls ist gegeben durch $2 \cdot z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$. Also folgt dass

$$\begin{aligned} 2 \cdot z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} &\leq 200 \\ \frac{z_{1-\alpha/2} \cdot \sigma}{100} &\leq \sqrt{n} \\ 785.57 &\leq n. \end{aligned}$$

Das heisst, es muss $n \geq 786$ gelten. Dies bedeutet, dass wir mindestens 79 Tage lang Stichproben sammeln müssten.

2. Approximationen durch den ZGWS

a) Sei $X \sim \text{Poisson}(\lambda)$ eine Zufallsvariable.

i) Verwende den ZGWS um X durch eine Normalverteilung zu approximieren.

ii) Begründe warum trotz des Resultates aus **a)** und dem Faktum, dass im Limes $n \to \infty$ der ZGWS exakt ist, X nicht perfekt durch eine Normalverteilung beschrieben wird. Für welche Werte von λ ist die Approximation plausibel?

b) Sei $n \in \mathbb{N}$, $\varepsilon > 0$ und seien $X_i > \varepsilon$, $1 \leq i \leq n$ i.i.d. Zufallsvariablen mit $\mathbb{E}[X_1] \in \mathbb{R}$ und $\text{Var}(X_1) < \infty$.

i) Verwende den ZGWS um die Verteilung von $S_n := \prod_{i=1}^n X_i$ zu approximieren.

ii) Zeige an Hand eines Beispiels, dass die Annahme $X_i > \varepsilon$ nicht durch $X_i > 0$ ersetzt werden kann.

Lösung:

a) i) Fixiere $n \in \mathbb{N}$ und definiere unabhängige Zufallsvariablen $Y_i \sim \text{Poisson}(\frac{\lambda}{n})$, $1 \leq i \leq n$. Dann gilt $S_n := Y_1 + \cdots Y_n \sim \text{Poisson}(\lambda)$. Außerdem gilt $\mathbb{E}[Y_1] = \text{Var}(Y_1) = \frac{\lambda}{n}$. Mit dem ZGWS können wir also die Verteilung von S_n (und somit also Poisson(λ)) wie folgt approximieren $S_n \approx N(n\frac{\lambda}{n}, n\frac{\lambda}{n}) = N(\lambda, \lambda)$. Insbesondere gilt also $P(X \leq x) \approx \Phi\left(\frac{x-\lambda}{\sqrt{\lambda}}\right)$.

ii) Man könnte nach obiger Herleitung fälschlicherweise vermuten, dass man für $n \to \infty$ erhält, dass die Approximation in Wirklichkeit exakt ist. Dass das nicht der Fall sein kann ist jedoch offensichtlich, allein schon wenn man die Wertebereiche der beiden Verteilungen betrachtet (diskret vs. stetig, einseitig beschränkt vs. unbeschränkt). Man kann wie folgt erklären warum die Approximation nicht exakt ist. Mit wachsendem n verändert sich die Verteilung der Y_i und damit wächst auch die Anzahl der benötigten i.i.d. Summanden, s.d. die Summe in etwa normalverteilt ist.

   Angenommen $\lambda \in \mathbb{N}$, dann können wir $n = \lambda$ wählen. Somit haben wir eine fixe Verteilung der Y_i und können (empirisch) beobachten, wie viele unabhängige Summanden von Poisson(1) nötig sind, s.d. die Summe gut durch eine Normalverteilung approximiert wird. Das ist der Fall für $n = \lambda \geq 9$.

b) i) Es gilt $\log(S_n) = \sum_{i=1}^{n} \log(X_i)$, wobei die Zufallsvariablen $Z_i := \log(X_i)$ i.i.d. sind. Man kann zeigen, dass $\mu := \mathbb{E}[Z_i] \in \mathbb{R}$ und $\sigma^2 := \mathrm{Var}(Z_i) < \infty$. Mit dem ZGWS folgt $\log(S_n) \approx N(n\mu, n\sigma^2)$ und somit ist S_n lognormalverteilt mit den entsprechenden Parametern.

ii) Sei $V \sim$ Poisson(1) und definieren $X_1 := \frac{1}{\exp(\exp(V^2))} > 0$. Dann ist zwar $\mathbb{E}[X_1] \in [0,1]$ aber $\mathbb{E}[\log(X_1)] = \mathbb{E}[-\exp(V^2)] = -\infty$ [1]. Also kann der ZGWS nicht verwendet werden kann.

[1] Um das zu sehen, beachte dass für $k \in \mathbb{N}$ gilt $exp(k^2) = \sum_j \frac{(k^2)^j}{j!} \geq \frac{k^{2k}}{k!}$ und $k! \leq k^{k-1}$. Somit folgt $\mathbb{E}[\exp(V^2)] = \sum_{k=0}^{\infty} \exp(k^2) e^{-1}/(k!) \geq e^{-1} \sum_{k \in \mathbb{N}} k^2 = \infty$