Prof. Vincent Tassion
Laurin Köhler-Schindler

Wahrscheinlichkeit & Statistik

Serie 11

Abgabe bis Mittwoch (18.05.2022) um 10:15 Uhr

Diese Serie beschäftigt sich mit Konfidenzintervallen und der t-Verteilung.

Weitere Informationen und Instruktionen zur Abgabe unter
https://metaphor.ethz.ch/x/2022/fs/401-0614-00L/

Wichtiger Hinweis: Ein Konfidenzintervall für θ mit Niveau $1-\alpha$ ist ein Zufallsintervall $I = [A, B]$, sodass für alle $\theta \in \Theta$ gilt

$$\mathbb{P}_\theta[A \leq \theta \leq B] \geq 1 - \alpha.$$

In der Definition, die in der Vorlesung gegeben wurde, wurde α und $1-\alpha$ vertauscht. Im Skript wurde dies nun korrigiert.

Aufgabe 11.1 [Konfidenzintervall für $\mathcal{N}(\mu, \sigma^2)$ mit bekannter Standardabweichung]
In einem Sägewerk wird das Schnittholz einer bestimmten Sortierklasse einer Qualitätskontrolle unterzogen. An jedem Produktionstag wird ein Brett auf seine Steifigkeit getestet. $X_n, n \geq 1$, bezeichne das Ergebnis am n-ten Tag. Basierend auf Erfahrung kann angenommen werden, dass die Steifigkeit eines Brettes normalverteilt ist mit bekannter Standardabweichung $\sigma = 1430$ MPa[1]. Wir wählen also den Parameterraum $\Theta = \mathbb{R}$ und die Modellfamilie $(\mathbb{P}_\theta)_{\theta \in \Theta}$, wobei die $(X_n)_{n \geq 1}$ unter $\mathbb{P}_\theta$ unabhängig und $\mathcal{N}(\theta, \sigma^2)$-verteilt sind.

(a) Leite die Formel eines Konfidenzintervalls für θ mit Niveau 95% nach 150 Produktionstagen her.

(b) Berechne aus (a) das realisierte Konfidenzintervall für einen beobachteten Stichprobenmittelwert von $11'000$ MPa (nach 150 Produktionstagen).

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

Aufgabe 11.2 [t-Verteilung mit m Freiheitsgraden]
Seien X und Y unabhängige Zufallsvariablen mit $X \sim \mathcal{N}(0, 1)$ und $Y \sim \chi^2_m$. Ziel dieser Aufgabe ist es zu zeigen, dass der Quotient

$$Z := \frac{X}{\sqrt{\frac{1}{m} Y}}$$

t-verteilt ist mit m Freiheitsgraden.

(a) Bestimme die Dichte von $Y' := \sqrt{\frac{1}{m} Y}$.

(b) Seien $\hat{X}$ und $\hat{Y}$ zwei unabhängige Zufallsvariablen mit Dichte $f_{\hat{X}}$ resp. $f_{\hat{Y}}$. Definiere $\hat{Z} := \frac{\hat{X}}{\hat{Y}}$ und bestimme die Dichte von $\hat{Z}$ (in Abhängigkeit von $f_{\hat{X}}$ und $f_{\hat{Y}}$).

Hinweis: Benutze in (a) und (b) die Charakterisierung der Dichte aus Proposition 4.16.

(c) Zeige, dass Z t-verteilt ist mit m Freiheitsgraden.

---

[1] Das Pascal ist eine abgeleitete SI-Einheit des Drucks sowie der mechanischen Spannung. Sie ist nach Blaise Pascal benannt und folgendermassen definiert: 1 Pa = $1\text{kg}\times\text{m}^{-1}\times\text{s}^{-2} = 1\text{N}\times\text{m}^{-2}$. Ein Pascal ist also der Druck, den eine Kraft von einem Newton auf eine Fläche von einem Quadratmeter ausübt. 1 MPa = 10^6 Pa = 10 bar.

Aufgabe 11.3 [Konfidenzintervall für $\mathcal{N}(\mu, \sigma^2)$ mit unbekannter Standardabweichung]
Wir betrachten erneut die Situation aus Aufgabe 11.1 und nehmen weiterhin an, dass die Steifigkeit eines Brettes normalverteilt ist. In diesem Fall gehen wir aber davon aus, dass die Standardabweichung σ unbekannt ist. Der unbekannte Parameter $\theta = (\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}_+$ ist in diesem Fall also 2-dimensional und wir wählen die Modellfamilie $(\mathbb{P}_\theta)_{\theta \in \Theta}$, wobei die $(X_n)_{n \geq 1}$ unter $\mathbb{P}_\theta$ unabhängig und $\mathcal{N}(\mu, \sigma^2)$-verteilt sind.
Zur Erinnerung: Aus der Vorlesung kennen wir bereits zwei Schätzer

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i \quad \text{und} \quad S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2$$

für μ und σ^2.

(a) Zeige für jedes $\theta = (\mu, \sigma^2) \in \Theta$, dass die Zufallsvariable

$$\frac{\bar{X}_n - \mu}{\sqrt{S^2/n}}$$

t_{n-1}-verteilt ist unter $\mathbb{P}_\theta$.
Hinweis: Verwende $\bar{X}_n \sim \mathcal{N}(\mu, \sigma^2/n)$, $\frac{n-1}{\sigma^2} S^2 \sim \chi^2_{n-1}$, die Unabhängigkeit von $\bar{X}_n$ und S^2 sowie das Resultat aus der vorherigen Aufgabe.

(b) Leite die Formel eines Konfidenzintervalls für μ mit Niveau 95% nach 10 Produktionstagen her.

Lösung 11.1

(a) In unserem Modell sind $X_1, \dots, X_n$ unabhängig, identisch verteilt mit $X_1 \sim \mathcal{N}(\mu, \sigma^2)$ unter $\mathbb{P}_\theta$, wobei $\theta = \mu$ der unbekannte Parameter ist und die Varianz σ^2 bekannt ist. Dann ist

$$\overline{X}_n \sim \mathcal{N}(\mu, \sigma^2/n) \quad \text{unter } \mathbb{P}_\theta.$$

Also gilt

$$\mathbb{P}_\theta\left[-c_{1-\alpha/2} \leq \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \leq c_{1-\alpha/2}\right] = 1 - \alpha,$$

wobei $c_{1-\alpha/2} = \Phi^{-1}(1 - \frac{\alpha}{2})$. Durch Umformen erhalten wir

$$\mathbb{P}_\theta\left[\overline{X}_n - c_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X}_n + c_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right] = 1 - \alpha.$$

Also ist

$$\left[\overline{X}_n - c_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\, , \, \overline{X}_n + c_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

ein Konfidenzintervall für $\theta = \mu$ mit Niveau $1 - \alpha$. Für $\alpha = 0.05$ erhält man $c_{1-\alpha/2} = c_{0.975} = 1.96$. Das Vertrauensintervall nach 150 Produktionstagen ist also

$$\left[\overline{X}_n - c_{1-\alpha/2}\frac{\sigma}{\sqrt{150}}, \overline{X}_n + c_{1-\alpha/2}\frac{\sigma}{\sqrt{150}}\right] = \left[\overline{X}_n - 228.8477, \overline{X}_n + 228.8477\right].$$

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

(c) Die Breite des Vertrauensintervalls ist gegeben durch $2c_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}$. Also wollen wir

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

Das heisst, es muss $n \geq 786$ gelten.

Lösung 11.2

(a) Wir verwenden die Charakterisierung der Dichte aus Proposition 4.16. Nach Annahme hat Y die Dichte

$$f_Y(y) = \frac{1}{2^{\frac{m}{2}}\Gamma(\frac{m}{2})} y^{\frac{m}{2}-1} e^{-\frac{1}{2}y} \mathbb{1}_{y \geq 0}.$$

Für $\phi : \mathbb{R} \to \mathbb{R}$ stückweise stetig und beschränkt gilt

$$\mathbb{E}\left[\phi(Y')\right] = \mathbb{E}\left[\phi(\sqrt{Y/m})\right] = \int_0^\infty \phi(\sqrt{y/m}) \frac{1}{2^{\frac{m}{2}}\Gamma(\frac{m}{2})} y^{\frac{m}{2}-1} e^{-\frac{1}{2}y} dy.$$

Mit dem Variablenwechsel $z = \sqrt{y/m}$ erhält man

$$\begin{aligned} \mathbb{E}\left[\phi(Y')\right] &= \int_0^\infty \phi(z) \frac{1}{2^{\frac{m}{2}}\Gamma(\frac{m}{2})} (z^2 m)^{\frac{m}{2}-1} e^{-\frac{1}{2}z^2 m} (2mz) dz \\ &= \int_0^\infty \phi(z) \frac{m^{\frac{m}{2}}}{2^{\frac{m}{2}-1}\Gamma(\frac{m}{2})} z^{m-1} e^{-\frac{1}{2}z^2 m} dz. \end{aligned}$$

und somit hat Y' die Dichte $f_{Y'}(z) = \frac{m^{\frac{m}{2}}}{2^{\frac{m}{2}-1}\Gamma(\frac{m}{2})} z^{m-1} e^{-\frac{1}{2}z^2 m} \mathbb{1}_{z \geq 0}$.

(b) Wir verwenden die Charakterisierung der Dichte aus Proposition 4.16. Nach Annahme sind die Zufallsvariablen $\hat{X}$ und $\hat{Y}$ unabhängig und haben somit die gemeinsame Dichte $f(x,y) = f_{\hat{X}}(x) \cdot f_{\hat{Y}}(y)$. Für $\phi : \mathbb{R} \to \mathbb{R}$ stückweise stetig und beschränkt gilt somit

$$\begin{aligned}
\mathbb{E}\left[\phi(\hat{Z})\right] = \mathbb{E}\left[\phi\left(\frac{\hat{X}}{\hat{Y}}\right)\right] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \phi(x/y) f_{\hat{X}}(x) f_{\hat{Y}}(y)\, dx\, dy \\
&= \int_{-\infty}^{0}\left(\int_{-\infty}^{\infty} \phi(x/y) f_{\hat{X}}(x)\, dx\right) f_{\hat{Y}}(y)\, dy + \int_{0}^{\infty}\left(\int_{-\infty}^{\infty} \phi(x/y) f_{\hat{X}}(x)\, dx\right) f_{\hat{Y}}(y)\, dy
\end{aligned}$$

Mit dem Variablenwechsel $z = x/y$ (für fixiertes y) erhält man

$$\begin{aligned}
\mathbb{E}\left[\phi(\hat{Z})\right] &= \int_{-\infty}^{0}\left(-\int_{-\infty}^{\infty} \phi(z) f_{\hat{X}}(yz) y\, dz\right) f_{\hat{Y}}(y)\, dy + \int_{0}^{\infty}\left(\int_{-\infty}^{\infty} \phi(z) f_{\hat{X}}(yz) y\, dz\right) f_{\hat{Y}}(y)\, dy \\
&= \int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty} \phi(z) f_{\hat{X}}(yz)|y|\, dz\right) f_{\hat{Y}}(y)\, dy \\
&= \int_{-\infty}^{\infty} \phi(z)\left(\int_{-\infty}^{\infty} f_{\hat{X}}(yz) f_{\hat{Y}}(y)|y|\, dy\right) dz
\end{aligned}$$

und somit hat Z' die Dichte $f_{\hat{Z}}(z) = \int_{-\infty}^{\infty} f_{\hat{X}}(yz) f_{\hat{Y}}(y)|y|\, dy$.

(c) Wir kombinieren die Teilaufgaben (a) und (b), um die Dichte von

$$Z = \frac{X}{\sqrt{Y/m}} = \frac{X}{Y'}$$

zu bestimmen. Da $X \sim \mathcal{N}(0,1)$, ist die Dichte $f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$, und da $Y \sim \chi^2_m$ ist die Dichte von Y' aus (a) bekannt. Durch Einsetzen in die allgemeine Formel aus (b) erhalten wir

$$\begin{aligned}
f_Z(z) &= \int_{-\infty}^{\infty} f_X(yz) f_{Y'}(y)|y|\, dy \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-(yz)^2/2} \cdot \frac{m^{\frac{m}{2}}}{2^{\frac{m}{2}-1}\Gamma(\frac{m}{2})} y^{m-1} e^{-\frac{1}{2}y^2 m} \mathbb{1}_{y\geq 0} \cdot |y|\, dy \\
&= \frac{1}{\sqrt{m\pi}\,\Gamma(\frac{m}{2})} \int_0^{\infty} \frac{m^{\frac{m+1}{2}}}{2^{\frac{m-1}{2}}} y^m e^{-\frac{1}{2}y^2(m+z^2)}\, dy
\end{aligned}$$

Mit dem Variablenwechsel $t = \frac{1}{2}y^2(m+z^2)$ erhalten wir

$$\begin{aligned}
f_Z(z) &= \frac{1}{\sqrt{m\pi}\,\Gamma(\frac{m}{2})} \int_0^{\infty} \frac{m^{\frac{m+1}{2}}}{2^{\frac{m-1}{2}}} \left(\frac{2t}{m+z^2}\right)^{m/2} e^{-t} \frac{1}{\sqrt{2t(m+z^2)}}\, dt \\
&= \frac{1}{\sqrt{m\pi}\,\Gamma(\frac{m}{2})} \left(\frac{m}{m+z^2}\right)^{\frac{m+1}{2}} \int_0^{\infty} t^{\frac{m-1}{2}} e^{-t}\, dt \\
&= \frac{\Gamma(\frac{m+1}{2})}{\sqrt{m\pi}\,\Gamma(\frac{m}{2})} \left(1 + \frac{z^2}{m}\right)^{-\frac{m+1}{2}},
\end{aligned}$$

also ist die Zufallsvariable Z t-verteilt mit m Freiheitsgraden.

Lösung 11.3

(a) Aus dem Hinweis folgt, dass

$$\sqrt{n}\frac{\bar{X}_n - \mu}{\sigma} \sim \mathcal{N}(0,1)$$

und somit erhalten wir aus den Hinweisen und dem Resultat aus Aufgabe 11.2 (c), dass

$$\frac{\sqrt{n}\frac{\bar{X}_n-\mu}{\sigma}}{\sqrt{\frac{\frac{n-1}{\sigma^2}S^2}{n-1}}} = \frac{\bar{X}_n - \mu}{\sqrt{\frac{S^2}{n}}}$$

t-verteilt ist mit $n-1$ Freiheitsgraden.

(b) In unserem Modell sind $X_1, \dots, X_n$ unabhängig, identisch verteilt mit $X_1 \sim \mathcal{N}(\mu, \sigma^2)$ unter $\mathbb{P}_\theta$, wobei $\theta = (\mu, \sigma^2)$ ein 2-dimensionaler unbekannter Parameter ist. Aus (a) wissen wir, dass

$$\frac{\bar{X}_n - \mu}{\sqrt{\frac{S^2}{n}}} \quad \text{unter } \mathbb{P}_\theta$$

t-verteilt ist mit $n-1$ Freiheitsgraden.
Also gilt

$$\mathbb{P}_\theta\left[-c_{1-\alpha/2} \le \frac{\bar{X}_n - \mu}{\sqrt{S^2/n}} \le c_{1-\alpha/2}\right] = 1-\alpha,$$

wobei $c_{1-\alpha/2} = F_{n-1}^{-1}(1-\frac{\alpha}{2})$ und F_{n-1} die Verteilungsfunktion der t-Verteilung mit $n-1$ Freiheitsgraden ist. Durch Umformen erhalten wir

$$\mathbb{P}_\theta\left[\bar{X}_n - c_{1-\alpha/2}\sqrt{\frac{S^2}{n}} \le \mu \le \overline{X}_n + c_{1-\alpha/2}\sqrt{\frac{S^2}{n}}\right] = 1-\alpha.$$

Also ist

$$\left[\overline{X}_n - c_{1-\alpha/2}\sqrt{\frac{S^2}{n}}\,,\, \overline{X}_n + c_{1-\alpha/2}\sqrt{\frac{S^2}{n}}\right]$$

ein Konfidenzintervall für μ mit Niveau $1-\alpha$. Für $\alpha = 0.05$ erhält man $c_{1-\alpha/2} = c_{0.975} = 2.262$. Das Vertrauensintervall nach 10 Produktionstagen ist also

$$\left[\overline{X}_n - c_{1-\alpha/2}\sqrt{\frac{S^2}{10}}, \overline{X}_n + c_{1-\alpha/2}\sqrt{\frac{S^2}{10}}\right] = \left[\overline{X}_n - 0.715\sqrt{S^2}, \overline{X}_n + 0.715\sqrt{S^2}\right].$$