18. Vorlesung: Konfidenzbereiche II: t- und χ^2-Score

Nikolas Tapia

20. Juni 2024, Stochastik für Informatik(er)

Konfidenzbereiche

Definition 18.1

Sei $\alpha \in (0,1)$ ein **Fehlerniveau** vorgegeben.
Ein **Konfidenzbereich für θ zum Fehlerniveau** α ist ein *zufälliges* Intervall $J \subseteq \Theta$, sodass

$$\mathbb{P}_\theta(\theta \in J) \geq 1 - \alpha \quad \forall \theta \in \Theta$$

gilt. übl. $\alpha = 0.05$ oder $\alpha = 0.01$.

Definition 18.2

Eine **Pivot-Statistik** ist eine Zufallsvariable $T(X_1, \dots, X_n; \theta)$, die eine Verteilung unter $\mathbb{P}_\theta$ hat, die nicht von θ abhängt und *explizit* angegeben werden kann, d.h.

$$\mathbb{P}_\theta(T(X_1, \dots, X_n; \theta) \leq t) = F(t) \quad \forall \theta \in \Theta$$

$X \sim N(\mu, 1)$

$T = X - \mu \sim N(0,1)$

$$T(X_1, \dots X_n; \mu) = \frac{\frac{1}{n}\sum_{i=1}^{n} X_i - \mu}{\sigma/\sqrt{n}}$$

$$T(X_1,\dots,X_n;\mu) = \frac{\overline{\mu}_n - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

$$1-\alpha \doteq \mathbb{P}\Big(\underset{\shortparallel}{q_{\alpha/2}} \le T \le \underset{\shortparallel}{q_{1-\alpha/2}} \Big)$$

$$= \mathbb{P}\big(-z_{1-\alpha/2} \le T \le z_{1-\alpha/2} \big)$$

z-Score

$$= \mathbb{P}\left(-z_{1-\alpha/2} \le \frac{\overline{\mu}_n - \mu}{\sigma/\sqrt{n}} \le z_{1-\alpha/2} \right)$$

$$= \mathbb{P}\left(\overline{\mu}_n - \frac{\sigma}{\sqrt{n}} z_{1-\alpha/2} \le \mu \le \overline{\mu}_n + \frac{\sigma}{\sqrt{n}} z_{1-\alpha/2} \right)$$

Unbekannter μ, bekannte σ^2

Theorem 18.1

Sei $X_1, \dots, X_n$ eine Folge von u.i.v. Zufallsvariablen mit bekannte Varianz $\sigma^2 > 0$. Ist n hinreichend groß, so ist

$$T(X_1, \dots, X_n; \mu) = \frac{\bar{\mu}_n(X_1, \dots, X_n) - \mu}{\sigma / \sqrt{n}} \sim \mathcal{N}(0,1) \to \text{Symmetrisch}$$

$$\Rightarrow z_{\alpha/2} = -z_{1-\alpha/2}$$

eine Pivot-Statistik für μ.

Quant'l

$$\Rightarrow J = \left[\bar{\mu}_n - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{\mu}_n + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

Im Allg.:

$$T \text{ Pivot} \Rightarrow \mathbb{P}\left(q_{\alpha/2} \leq T \leq q_{1-\alpha/2}\right) = 1-\alpha$$

Bekannter μ, unbekannte σ^2

Aussage 18.1

Sei $X \sim \mathcal{N}(0,1)$. Dann hat die Zufallsvariable $Y = X^2$ die Dichte

$$f_Y(y) = \begin{cases} \frac{1}{\cancel{\sqrt{\pi y}}\ \sqrt{2\pi y}} e^{-y/2} & y > 0, \\ 0 & \text{sonst} \end{cases}.$$

In diesem Fall sagt man, dass Y **chi-Quadrat-verteilt** ist mit einem Freiheitsgrad: $Y \sim \chi^2_1$

$y \geq 0. \quad Y(\Omega) = [0, \infty).$

Beweis

$$\mathbb{P}(Y \leq y) = \mathbb{P}(X^2 \leq y) = \mathbb{P}(|X| \leq \sqrt{y})$$

$$= \mathbb{P}(-\sqrt{y} \leq X \leq \sqrt{y})$$

$X \sim \mathcal{N}(0,1). \longrightarrow$

$$= \int_{-\sqrt{y}}^{\sqrt{y}} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} \, dt$$

$u = t^2$

$\rightarrow du = 2t\,dt$

$\rightarrow dt = \frac{1}{2\sqrt{u}}$

$$\rightarrow = 2 \int_0^{\sqrt{y}} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} \, dt$$

$$= 2 \int_0^{y} \frac{1}{\sqrt{2\pi}} e^{-u/2} \cdot \frac{1}{2\sqrt{u}} \, du$$

Leibniz Gemeinschaft

Beweis

$$F_Y(y) = \mathbb{P}(Y \leq y) = \frac{1}{\sqrt{2\pi}} \int_0^y \frac{1}{\sqrt{u}} e^{-u/2} \, du$$

$$f_Y(y) = F_Y'(y) = \begin{cases} \sqrt{\frac{1}{2\pi u}}\, e^{-u/2} & u \geq 0. \\ 0 & \sim \end{cases}$$

chi-Quadrat-Verteilung n Freiheitsgrad

Definition 18.3

Seien $X_1, \ldots, X_n$ u.i.v. $\mathcal{N}(0,1)$-verteilte Zufallsvariablen. Dann ist

$$Y = \sum_{i=1}^{n} X_i^2 \qquad \left(\overline{\sigma_n}^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \mu)^2 \right)$$

eine Zufallsvariable, die **chi-Quadrat-verteilt** ist mit n Freiheitsgrad: $Y \sim \chi_n^2$. Ihre Dichte ist durch

$$f_Y(y) = \begin{cases} \frac{1}{2^{n/2}\Gamma\left(\frac{n}{2}\right)} y^{n/2-1} e^{-y/2} & y > 0, \\ 0 & \text{sonst} \end{cases}$$

$$\Gamma\left(\frac{n}{2}\right) := 2^{n/2} \int_0^\infty y^{n/2-1} e^{-y/2} \, dy$$

gegeben.

$$\Rightarrow \int_0^\infty f_Y(y)\, dy = 1$$

Konfidenzintervall für σ^2 bei bekannten μ

Theorem 18.2

Sei $X_1, \dots, X_n$ eine Folge von u.i.v. $\mathcal{N}(\mu, \sigma^2)$-verteilten Zufallsvariablen, mit bekannten $\mu \in \mathbb{R}$. Dann ist

$$T(X_1, \dots, X_n; \sigma^2) = \frac{(n-1) S_n^2(X_1, \dots, X_n)}{\sigma^2} \sim \chi_n^2,$$

wobei

$$S_n^2(X_1, \dots, X_n) = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \mu)^2$$

Bew:

$$\frac{(n-1)S_n^2}{\sigma^2} = \sum_{i=1}^{n} \underbrace{\left(\frac{X_i - \mu}{\sigma}\right)}_{Z_i \sim \mathcal{N}(0,1)}^2 = \sum_{i=1}^{n} Z_i^2 \sim \chi_n^2$$

Theorem 18.3

Sei $x_1, \ldots, x_n$ eine Stichprobe aus einer $\mathcal{N}(\mu, \sigma^2)$-verteilten Grundgesamtheit und $\alpha \in (0,1)$, wobei $\mu \in \mathbb{R}$ *bekannt* ist. Dann ist

$$J = \left[\frac{(n-1)S_n^2}{\chi^2_{n,1-\alpha/2}}, \frac{(n-1)S_n^2}{\chi^2_{n,\alpha/2}} \right]$$

ein Konfidenzintervall für σ^2 zum Fehlerniveau α.

$$1-\alpha = \mathbb{P}\left(q_{\alpha/2} \leq T \leq q_{1-\alpha/2}\right) = \mathbb{P}\Big(q_{\alpha/2} \leq \underbrace{\frac{(n-1)S_n^2}{\sigma^2}}_{\chi_n^2} \leq q_{1-\alpha/2}\Big)$$

$$= \mathbb{P}\left(\frac{1}{\chi^2_{n,1-\alpha/2}} \leq \frac{\sigma^2}{(n-1)S_n^2} \leq \frac{1}{\chi^2_{n,\alpha/2}}\right)$$

$$= \mathbb{P}\left(\frac{(n-1)S_n^2}{\chi^2_{n,1-\alpha/2}} \leq \sigma^2 \leq \frac{(n-1)S_n^2}{\chi^2_{n,\alpha/2}}\right)$$

Konfidenzintervall für σ^2 bei unbekanntem μ

Theorem 18.4

Sei $X_1, \ldots, X_n$ eine Folge von u.i.v. $\mathcal{N}(\mu, \sigma^2)$-verteilten Zufallsvariablen, und $\alpha \in (0, 1)$. Dann ist

$$T(X_1, \ldots, X_n; \sigma^2) = \frac{(n-1)\bar{\sigma}_n^2(X_1, \ldots, X_n)}{\sigma^2} \sim \chi^2_{n-1}.$$

$$\bar{\sigma}_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{\mu}_n)^2$$

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \mu)^2$$

Somit ist

$$J = \left[\frac{(n-1)\bar{\sigma}_n^2}{\chi^2_{n-1,1-\alpha/2}}, \frac{(n-1)\bar{\sigma}_n^2}{\chi^2_{n-1,\alpha/2}}\right]$$

ein Konfidenzintervall für σ^2 zum Fehlerniveau α.

Konfidenzintervall für μ bei unbekannter σ^2

Definition 18.4

Seien $Z \sim \mathcal{N}(0,1)$, $V \sim \chi^2_m$ unabhängig. Die Verteilung der Zufallsvariable

$$T = \frac{Z}{\sqrt{V/m}} \qquad (\text{Pearson}).$$

heißt ***t*-Verteilung** mit m Freiheitsgraden: $T \sim t_m$. Deren Dichte ist durch

$$f_T(t) = \frac{\Gamma\left(\frac{m+1}{2}\right)}{\sqrt{m\pi}\,\Gamma\left(\frac{m}{2}\right)} \left(1 + \frac{t^2}{m}\right)^{-(m+1)/2}$$

gegeben.

Konfidenzintervall für μ bei unbekannter σ^2

Theorem 18.5

Sei $(x_1, \ldots, x_n) \in \mathbb{R}^n$ eine Stichprobe aus einer $\mathcal{N}(\mu, \sigma^2)$-verteilten Grundgesamtheit. Dann ist

$$T(x_1, \ldots, x_n; \mu) = \frac{\bar{\mu}_n - \mu}{\sqrt{\bar{\sigma}_n^2 / n}} \sim t_{n-1}$$

$\sim \chi^2_{n-1}$

eine Pivot-Statistik für μ. Deshalb ist

t-Score t-score

$$J = \left[\bar{\mu}_n - t_{n-1,1-\alpha/2}\sqrt{\bar{\sigma}_n^2/n}, \bar{\mu}_n + t_{n-1,1-\alpha/2}\sqrt{\bar{\sigma}_n^2/n}\right]$$

ein Konfidenzintervall für μ zum Fehlerniveau α.

$$Z = \frac{\bar{\mu}_n - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1), \quad V = \frac{(n-1)\bar{\sigma}_n^2}{\sigma^2} \sim \chi^2_{n-1} \Rightarrow T = \frac{Z}{\sqrt{\frac{V}{n-1}}} = \frac{\bar{\mu}_n - \mu}{\sqrt{\bar{\sigma}_n^2/n}} \sim t_{n-1}$$