Angewandte Stochastik

Prof. Dr. Evgeny Spodarev | Vorlesungskurs |

8. Thema

Heutiges Thema

- Schätzer der Varianz

Voraussetzungen (Schätzer der Varianz)

- Seien $X_i\,, \quad i = 1, \ldots, n$ unabhängig identisch verteilt, $X_i \stackrel{d}{=} X\,, \quad \mathrm{E}_\theta\, X^2 < \infty \quad \forall \theta \in \Theta\,, \quad \theta = (\theta_1, \ldots, \theta_m)^T\,,$ $\theta_{j_0} = \sigma^2 = \mathrm{Var}_{\,\theta}\, X$ für ein $j_0 \in \{1, \ldots, m\}$.
- Die *Stichprobenvarianz*

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

ist dann ein Schätzer für σ^2.

Voraussetzungen (Schätzer der Varianz)

- Falls der Erwartungswert $\mu = \mathsf{E}_\theta X$ der Stichprobenvariablen explizit benannt ist, so kann ein Schätzer für σ^2 auch als

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

definiert werden.
- Wir werden nun die Eigenschaften von S_n^2 und $\tilde{S}_n^2$ untersuchen und sie miteinander vergleichen.

Satz

1. Die Stichprobenvarianz S_n^2 ist erwartungstreu für σ^2:

$$\mathrm{E}_\theta\, S_n^2 = \sigma^2\,, \qquad \theta \in \Theta\,.$$

2. Wenn $\mathrm{E}_\theta\, X^4 < \infty$, dann gilt

$$\mathrm{Var}_\theta\, S_n^2 = \frac{1}{n}\left(\mu_4' - \frac{n-3}{n-1}\sigma^4\right)\,,$$

wobei $\mu_4' = \mathrm{E}_\theta\,(X-\mu)^4$.

Satz

1. Der Schätzer $\tilde{S}_n^2$ für σ^2 ist erwartungstreu.
2. Es gilt $\operatorname{Var}_\theta \tilde{S}_n^2 = \frac{1}{n}(\mu_4' - \sigma^4)$.

Folgerung

Der Schätzer $\tilde{S}_n^2$ für σ^2 ist besser als S_n^2, weil beide erwartungstreu sind und

$$\operatorname{Var}_\theta \tilde{S}_n^2 = \frac{\mu_4' - \sigma^4}{n} < \frac{\mu_4' - \frac{n-3}{n-1}\sigma^4}{n} = \operatorname{Var}_\theta S_n^2 .$$

Diese Eigenschaft von $\tilde{S}_n^2$ im Vergleich zu S_n^2 ist intuitiv klar, da man in $\tilde{S}_n^2$ mehr Informationen über die Verteilung der Stichprobenvariablen X_i (nämlich den bekannten Erwartungswert μ) reingesteckt hat.

Satz

1. Die Schätzer S_n^2 bzw. $\tilde{S}_n^2$ sind stark konsistent für σ^2.
2. Sie sind asymptotisch normalverteilt, d.h.

$$\sqrt{n}\frac{S_n^2 - \sigma^2}{\sqrt{\mu_4' - \sigma^4}} \xrightarrow[n\to\infty]{d} Y \sim N(0,1)\,,$$

$$\sqrt{n}\frac{\tilde{S}_n^2 - \sigma^2}{\sqrt{\mu_4' - \sigma^4}} \xrightarrow[n\to\infty]{d} Y \sim N(0,1)\,.$$

Folgerung

Es gilt

1. $$\sqrt{n}\frac{\bar{X}_n - \mu}{S_n} \xrightarrow[n\to\infty]{d} Y \sim N(0,1)$$

   und somit

2. $$P\left(\mu \in \left[\bar{X}_n - \frac{z_{1-\frac{\alpha}{2}} S_n}{\sqrt{n}}, \bar{X}_n + \frac{z_{1-\frac{\alpha}{2}} S_n}{\sqrt{n}}\right]\right) \xrightarrow[n\to\infty]{} 1-\alpha \quad (1)$$

   für ein $\alpha \in (0,1)$, wobei z_α das α-Quantil der $N(0,1)$-Verteilung ist, d.h. $z_\alpha = \Phi^{-1}(\alpha)$ mit $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{\frac{-t^2}{2}} dt$.

Bemerkung

1. Das Intervall $I_n^\alpha = \left[\bar{X}_n - \frac{z_{1-\frac{\alpha}{2}} S_n}{\sqrt{n}}, \bar{X}_n + \frac{z_{1-\frac{\alpha}{2}} S_n}{\sqrt{n}}\right]$ heißt asymptotisches Konfidenz- bzw. Vertrauensintervall zum Konfidenzniveau $1-\alpha$. Typisch hierbei sind die Werte $\alpha \in \{0,01; 0,05; 0,001\}$.
2. Für $n \to \infty$ gilt $|I_n^\alpha| = 2z_{1-\frac{\alpha}{2}} \frac{S_n}{\sqrt{n}} \xrightarrow{f.s.} 0$.
3. Der Gebrauch von I_n^α macht ab $n \approx 100$ Sinn.

Betrachten wir weiterhin den wichtigen Spezialfall der normalverteilten Stichprobenvariablen $X_i, \quad i = 1, \ldots, n$, also $X \sim N(\mu, \sigma^2)$ und $\theta = (\mu, \sigma^2)$.

Definition

1. Seien $Y_1, \ldots, Y_n$ u.i.v. $N(0,1)$-Z.V.. Dann wird die Verteilung von $Y = \sum_{i=1}^{n} Y_i^2$ als χ_n^2-Verteilung mit n Freiheitsgraden bezeichnet.
2. Sei $Y \sim N(0,1)$ und $Z \sim \chi_n^2$ unabhängig. Dann wird die Verteilung von $T = \frac{Y}{\sqrt{\frac{Z}{n}}}$ als t_n- Verteilung (Student t-Verteilung) mit n Freiheitsgraden bezeichnet.

Satz

Falls $X_1, \ldots, X_n$ normalverteilt sind mit Parametern μ und σ^2, dann gilt

1. $$\frac{(n-1)S_n^2}{\sigma^2} \sim \chi^2_{n-1}\,,$$

2. $$\frac{n\tilde{S}_n^2}{\sigma^2} \sim \chi^2_n\,.$$

Satz

Falls $X \sim N(\mu, \sigma^2)$, $X_1, \ldots, X_n$ unabhängige identisch verteilte Zufallsvariablen sind, $X_i \overset{d}{=} X$, dann

- sind $\bar{X}_n$ und S_n^2 unabhängig, und
- es gilt

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{S_n} \sim t_{n-1}\,.$$

Folgerung

Mit Hilfe des letzten Satzes kann folgendes Konfidenzintervall für den Erwartungswert μ einer normalverteilten Stichprobe $(X_1, \ldots, X_n)$ bei unbekannter Varianz σ^2 $(X_i \sim N(\mu, \sigma^2)\,, \quad i = 1, \ldots, n)$ konstruiert werden:

$$P\left(\mu \in \left[\bar{X}_n - \frac{t_{n-1,1-\frac{\alpha}{2}}}{\sqrt{n}} S_n\,,\ \bar{X}_n + \frac{t_{n-1,1-\frac{\alpha}{2}}}{\sqrt{n}} S_n\right]\right) = 1 - \alpha$$

für $\alpha \in (0, 1)$,

Folgerung

denn

$$
\begin{aligned}
&P\left(\sqrt{n}\frac{\bar{X}_n-\mu}{S_n}\in\Big[\underbrace{t_{n-1,\frac{\alpha}{2}}}_{=-t_{n-1,1-\frac{\alpha}{2}}\text{ wg. Sym. } t\text{-Vert.}},\; t_{n-1,1-\frac{\alpha}{2}}\Big]\right)\\
&=F_{t_{n-1}}(t_{n-1,1-\frac{\alpha}{2}})-F_{t_{n-1}}(t_{n-1,\frac{\alpha}{2}})\\
&=1-\frac{\alpha}{2}-\frac{\alpha}{2}=1-\alpha\,,
\end{aligned}
\tag{2}
$$

wobei $t_{n-1,\alpha}$ das α-Quantil der t_{n-1}-Verteilung darstellt.
Der Rest folgt aus (2) durch das Auflösen bzgl. μ.