Wahrscheinlichkeitstheorie und Statistik

13. Woche

Dávid Tóth (BME SZIT)

25., 28. November 2024

Eigenschaften der Schätzfunktionen

Beispiel. Sei $\mathbf{X} = (X_1, \ldots, X_n)$ eine einfache Stichprobe, wo $X_i \sim \mathrm{Exp}(\vartheta)$ gilt. Das empirische Mittel ist ein erwartungstreuer Schätzer für den Erwartungswert, also

$$\mathbb{E}_\vartheta(\overline{X}) = \mathbb{E}(X_1) = \frac{1}{\vartheta}.$$

Aber die Stichprobefunktion $1/\overline{X}$ ist nicht ein erwartungstreuer Schätzer für den Parameter ϑ!

Wir berechneten früher die Dichtefunktion der Summe $Y = X_1 + \cdots + X_n$:

$$f_Y(t) = \begin{cases} \dfrac{\vartheta^n t^{n-1} e^{-\vartheta t}}{(n-1)!}, & \text{falls } t > 0, \\ 0 & \text{sonst.} \end{cases}$$

Eigenschaften der Schätzfunktionen

Beispiel. Der Erwartungswert der Variablen

$$1/\overline{X} = \frac{n}{X_1 + \cdots + X_n} = \frac{n}{Y}$$

ergibt sich wegen der Transformationsformel folgenderweise:

$$\begin{aligned}\mathbb{E}_\vartheta\left(\frac{1}{\overline{X}}\right) &= n \int_{-\infty}^{\infty} \frac{1}{t} f_Y(t)\, dt \\ &= \frac{n}{n-1}\vartheta \int_0^\infty \frac{\vartheta^{n-1} t^{n-2} e^{-\vartheta t}}{(n-2)!}\, dt = \frac{n}{n-1}\vartheta,\end{aligned}$$

falls $n > 1$. D.h.

$$\frac{n-1}{n} \cdot \frac{1}{\overline{X}} = \frac{n-1}{X_1 + \cdots + X_n}$$

ist ein erwartungstreuer Schätzer für ϑ.

Eigenschaften der Schätzfunktionen

Definition. Sei $\mathbf{X} = (X_1, \ldots, X_n)$ eine einfache Stichprobe, wo die Verteilung der Variablen durch einen Parameter $\vartheta \in \Theta \subset \mathbb{R}^d$ bestimmt ist. Sei $\psi : \Theta \to \mathbb{R}$ eine Parameterfunktion.

Seien noch $T, T' : \mathbb{R}^n \to \mathbb{R}$ erwartungstreue Schätzer für $\psi(\vartheta)$, dann heißt T *mindestens so effizient* (oder *wirksam*) wie T', falls $\mathbb{D}^2_\vartheta(T(\mathbf{X})) \leq \mathbb{D}^2_\vartheta(T'(\mathbf{X}))$ für alle $\vartheta \in \Theta$ gilt (T legalább annyira hatásos, mint T').

Ist T mindestens so effizient wie alle (anderen) Schätzer einer Klasse von erwartungstreuen Schätzfunktionen für $\psi(\vartheta)$ (die auch T enthält), so nennt man T *effizient* in dieser Klasse (T hatásos becslés).

Eigenschaften der Schätzfunktionen

Beispiel. Sei $\mathbf{X} = (X_1, X_2, X_3)$ eine einfache Stichprobe, wo $X_i \sim U(\vartheta, \vartheta + 1)$ für eine $\vartheta \geq 0$ gilt. Für $\vartheta = 0$ bestimmten wir früher die Dichtefunktion und den Erwartungswert von X_2^* (siehe Aufgabe 15 auf dem 5-ten Übungsblatt und Augabe 3 auf dem 6-ten Übungsblatt):

$$f_{X_2^*}(t) = \begin{cases} 6t - 6t^2, & \text{falls } 0 < t < 1, \\ 0 & \text{ansonsten,} \end{cases} \qquad \mathbb{E}(X_2^*) = \frac{1}{2}.$$

Dann kann auch die Varianz berechnet werden: $\mathbb{D}^2(X_2^*) = 0{,}05$.

Falls $\vartheta > 0$ gilt, so sind die Variablen $Y_i = X_i - \vartheta$ unabhängig und uniformverteilt auf dem Intervall $(0;1)$, also

$$\mathbb{E}_\vartheta(X_2^*) = \mathbb{E}(Y_2^*) + \vartheta = \frac{1}{2} + \vartheta \quad \text{und} \quad \mathbb{D}_\vartheta^2(X_2^*) = \mathbb{D}^2(Y_2^*) = 0{,}05.$$

D.h.: X_2^* ist ein erwartungstreuer Schätzer für $\mathbb{E}_\vartheta(X_1)$.

Eigenschaften der Schätzfunktionen

Beispiel. Wir sahen auch, dass $\overline{X_3}$ ein erwartungstreuer Schätzer für $\mathbb{E}_\vartheta(X_1)$ ist, für den

$$\mathbb{D}^2_\vartheta(\overline{X_3}) = \frac{1}{3}\mathbb{D}^2_\vartheta(X_1) = \frac{1}{3 \cdot 12} = \frac{1}{36} < \frac{1}{20} = \mathbb{D}^2_\vartheta(X_2^*)$$

gilt, also ist $\overline{X_3}$ ein effizienter Schätzer für $\mathbb{E}_\vartheta(X_1)$ als X_2^*.

Eigenschaften der Schätzfunktionen

Satz. Sei $\mathbf{X} = (X_1, \ldots, X_n)$ eine einfache Stichprobe vom Umfang n, wo die Verteilung der Variablen von einem Parameter $\vartheta \in \Theta$ abhängt und $\mathbb{D}^2_\vartheta(X_1) < \infty$ für alle $\vartheta \in \Theta$ gilt. Nehmen wir an, dass T_1 und T_2 erwartungstreue und effiziente Schätzer in einer Klasse für die Parameterfunktion $\psi(\vartheta)$ sind. Falls diese Klasse auch den Schätzer $\frac{T_1+T_2}{2}$ enthält, dann gilt

$$\mathbb{P}_\vartheta(T_1(\mathbf{X}) = T_2(\mathbf{X})) = 1$$

für alle $\vartheta \in \Theta$.

Eigenschaften der Schätzfunktionen

Beweis. T_1 und T_2 sind erwartungstreu, also gilt

$$\mathbb{E}_\vartheta(T_1(\mathbf{X})) = \mathbb{E}_\vartheta(T_2(\mathbf{X})) = \psi(\vartheta)$$

und deshalb auch

$$\mathbb{E}_\vartheta\left(\frac{T_1(\mathbf{X}) + T_2(\mathbf{X})}{2}\right) = \psi(\vartheta),$$

also ist auch $\frac{T_1+T_2}{2}$ ein erwartungstreuer Schätzer für $\psi(\vartheta)$.

Eigenschaften der Schätzfunktionen

Beweis. T_1 und T_2 sind auch effizient, also

$$\mathbb{D}^2_\vartheta(T_1(\mathbf{X})) \leq \mathbb{D}^2_\vartheta(T_2(\mathbf{X})) \leq \mathbb{D}^2_\vartheta(T_1(\mathbf{X})),$$

d.h.

$$\mathbb{D}^2_\vartheta(T_1(\mathbf{X})) = \mathbb{D}^2_\vartheta(T_2(\mathbf{X})),$$

weiter

$$\begin{aligned}
\mathbb{D}^2_\vartheta(T_1(\mathbf{X})) &\leq \mathbb{D}^2_\vartheta\left(\frac{T_1(\mathbf{X}) + T_2(\mathbf{X})}{2}\right) \\
&= \frac{1}{4}[\mathbb{D}^2_\vartheta(T_1(\mathbf{X})) + \mathbb{D}^2_\vartheta(T_2(\mathbf{X})) + 2\mathrm{cov}\,(T_1(\mathbf{X}), T_2(\mathbf{X}))] \\
&= \frac{\mathbb{D}^2_\vartheta(T_1(\mathbf{X})) + \mathrm{cov}\,(T_1(\mathbf{X}), T_2(\mathbf{X}))}{2}.
\end{aligned}$$

Eigenschaften der Schätzfunktionen

Beweis. Wir bewiesen früher die obere Schranke

$$\operatorname{cov}(T_1(\mathbf{X}), T_2(\mathbf{X})) \leq \mathbb{D}_\vartheta(T_1(\mathbf{X})) \cdot \mathbb{D}_\vartheta(T_2(\mathbf{X})) = \mathbb{D}^2_\vartheta(T_1(\mathbf{X}))$$

also folgt

$$\mathbb{D}^2_\vartheta(T_1(\mathbf{X})) \leq \frac{\mathbb{D}^2_\vartheta(T_1(\mathbf{X})) + \operatorname{cov}(T_1(\mathbf{X}), T_2(\mathbf{X}))}{2} \leq \mathbb{D}^2_\vartheta(T_1(\mathbf{X})).$$

Dann muss $\mathbb{D}^2_\vartheta(T_1(\mathbf{X})) = \operatorname{cov}(T_1(\mathbf{X}), T_2(\mathbf{X}))$ gelten und deshalb

$$\begin{aligned}&\mathbb{D}^2_\vartheta(T_1(\mathbf{X}) - T_2(\mathbf{X})) = \\ &\qquad = \mathbb{D}^2_\vartheta(T_1(\mathbf{X})) + \mathbb{D}^2_\vartheta(T_2(\mathbf{X})) - 2\operatorname{cov}(T_1(\mathbf{X}), T_2(\mathbf{X})) = 0,\end{aligned}$$

also ist $T_1(\mathbf{X}) - T_2(\mathbf{X})$ mit Wahrscheinlichkeit 1 eine Konstante c.

Eigenschaften der Schätzfunktionen

Beweis. Aber

$$c = \mathbb{E}_\vartheta(c) = \mathbb{E}_\vartheta(T_1(X) - T_2(X)) = \psi(\vartheta) - \psi(\vartheta) = 0,$$

und der Satz ist bewiesen.

Eigenschaften der Schätzfunktionen

Behauptung. Sei $\mathbf{X} = (X_1, \ldots, X_n)$ eine einfache Stichprobe vom Umfang n, wo die Verteilung der Variablen von einem Parameter $\vartheta \in \Theta$ abhängt und $\mathbb{D}^2_\vartheta(X_1) < \infty$ für alle $\vartheta \in \Theta$ gilt. Nehmen wir noch an, dass $\mathbb{E}_\vartheta(X_1) \neq 0$ für mindestens einen Parameter gilt, dann ist das empirische Mittel $\overline{X}$ der effiziente Schätzung für den Erwartungswert in der Klasse der erwartungstreuen linearen Schätzer der Form $T(\mathbf{X}) = \sum_{i=1}^n c_i X_i$, wo $c_i \in \mathbb{R}$.

Beweis. Ein linearer Schätzer T ist genau dann erwartungstreu, falls

$$\mathbb{E}_\vartheta(X_1) = \mathbb{E}_\vartheta(T(\mathbf{X})) = \sum_{i=1}^n c_i \mathbb{E}_\vartheta(X_i) = \mathbb{E}_\vartheta(X_1) \sum_{i=1}^n c_i,$$

also falls

$$\sum_{i=1}^n c_i = 1.$$

Eigenschaften der Schätzfunktionen

Beweis. Wegen der Unabhängigkeit der Stichporbenvariablen ergibt sich

$$\mathbb{D}^2_\vartheta\left(\sum_{i=1}^n c_i X_i\right) = \sum_{i=1}^n c_i^2 \mathbb{D}^2_\vartheta(X_i) = \mathbb{D}^2_\vartheta(X_1)\sum_{i=1}^n c_i^2.$$

Hier können wir die Cauchy-Schwarzsche Ungleichung anwenden: für alle reelle Zahlen $a_1, \ldots, a_n, b_1, \ldots, b_n$ gilt

$$\left(\sum_{i=1}^n a_i b_i\right)^2 \leq \left(\sum_{i=1}^n a_i^2\right)\left(\sum_{i=1}^n b_i^2\right).$$

Für $a_i = c_i$ und $b_i = 1$ erhält man

$$1 = \left(\sum_{i=1}^n c_i\right)^2 \leq \left(\sum_{i=1}^n c_i^2\right)\cdot n.$$

Eigenschaften der Schätzfunktionen

Beweis. Also

$$\mathbb{D}^2_\vartheta\left(\sum_{i=1}^{n} c_i X_i\right) = \mathbb{D}^2_\vartheta(X_1)\sum_{i=1}^{n} c_i^2 \geq \frac{1}{n}\mathbb{D}^2_\vartheta(X_1) = \mathbb{D}^2_\vartheta(\overline{X}).$$

Maximum-Likelihood-Schätzung

Maximum-Likelihood-Schätzung

- Die Maximum-Likelihood-Schätzung ist eine spezielle Methode der Parameterschätzung.
- Bei dieser wird unter allen möglichen Werten eines Parameters einer Verteilung derjenige gesucht, für den die Wahrscheinlichkeit maximal wird, die beobachteten Realisierungen $x_1, \dots, x_n$ zu finden.
- Also bestimmt man den Wert $\vartheta = \vartheta_*(x_1, \dots, x_n) \in \Theta$ für jede Realisierung $x_1, \dots, x_n$ einer Stichprobe $X_1, \dots, X_n$, für den die Wahrscheinlichkeit $\mathbb{P}_\vartheta(X_1 = x_1, \dots, X_n = x_n)$ maximal ist.
- Die Funktion $\vartheta_*(x_1, \dots, x_n)$ heißt dann der *Maximum-Likelihood-Schätzer*, also wird $\vartheta_*(X_1, \dots, X_n)$ die Schätzung für den Parameter ϑ.

Maximum-Likelihood-Schätzung

Definition. Sei $\mathbf{X} = (X_1, \ldots, X_n)$ eine einfache Stichprobe vom Umfang n, wo die Verteilung der Variablen von einem Parameter $\vartheta \in \Theta$ abhängt. Nehmen wir an, dass die Verteilung der Stichprobenvariablen entweder für alle Parameter $\vartheta \in \Theta$ diskret, oder für alle Parameter ϑ absolut stetig ist. Für jede Realisierung $(x_1, \ldots, x_n)$ wird der Wert der zu ϑ gehörigen *Likelihood-Funktion* durch

$$L_\vartheta(x_1, \ldots, x_n) = \prod_{i=1}^{n} f_\vartheta(x_i)$$

angegeben, wo f_ϑ die Wahrscheinlichkeitsfunktion bzw. die Dichtefunktion der zu ϑ gehörigen Verteilung ist.

Maximum-Likelihood-Schätzung

Bemerkung.

- Falls die Verteilung der Stichprobenvariablen diskret ist, dann gilt wegen der Unabhängigkeit der Stichprobenvariablen

$$\begin{aligned} L_\vartheta(x_1, \ldots, x_n) &= \prod_{i=1}^{n} f_\vartheta(x_i) = \prod_{i=1}^{n} \mathbb{P}_\vartheta(X_i = x_i) \\ &= \mathbb{P}_\vartheta(X_1 = x_1, \ldots, X_n = x_n). \end{aligned}$$

- Für stetige Variablen ist die obige Wahtscheinlichkeit immer Null. Stattdessen betrachten wir den Wert der gemeinsamen Dichtefunktion der Stichprobenvariablen:

$$L_\vartheta(x_1, \ldots, x_n) = \prod_{i=1}^{n} f_\vartheta(x_i) = f_{\vartheta(X_1,\ldots,X_n)}(x_1, \ldots, x_n).$$

Maximum-Likelihood-Schätzung

Bemerkung. Der letzte Wert kann folgenderweise ausgdrückt werden:

$$L_\vartheta(x_1, \ldots, x_n) = \prod_{i=1}^{n} f_\vartheta(x_i) = \prod_{i=1}^{n} \lim_{\Delta x_i \downarrow 0} \frac{\mathbb{P}_\vartheta(X_i \in (x_i - \Delta x_i; x_i])}{\Delta x_i}$$

$$= \lim_{\Delta x_1 \downarrow 0, \ldots, \Delta x_n \downarrow 0} \frac{\mathbb{P}_\vartheta(X_1 \in (x_1 - \Delta x_1; x_1], \ldots, X_n \in (x_n - \Delta x_n; x_n])}{\Delta x_1 \ldots \Delta x_n},$$

falls z. B. $f_{\vartheta,(X_1,\ldots,X_n)}$ an der Stelle $(x_1, \ldots, x_n)$ stetig ist. Dann bestimmt der Wert $L_\vartheta(x_1, \ldots, x_n)$ eine approximative Wahrscheinlichkeit dafür, dass der Wert des Zufallsvektors $(X_1, \ldots, X_n)$ in der Nähe der Realisierung $(x_1, \ldots, x_n)$ liegt.

Maximum-Likelihood-Schätzung

Definition. Sei $\mathbf{X} = (X_1, \ldots, X_n)$ eine einfache Stichprobe vom Umfang n, wo die Verteilung der Variablen von einem Parameter $\vartheta \in \Theta$ abhängt und sei $(x_1, \ldots, x_n)$ eine Realisierung der Stichprobe. Der Wert des *Maximum-Likelihood-Schätzers* für diese Realisierung ist

$$\vartheta_*(x_1, \ldots, x_n) = \arg\max_{\vartheta \in \Theta} L_\vartheta(x_1, \ldots, x_n),$$

also der Parameter $\vartheta \in \Theta$, für den die Likelihood-Funktion bei dieser Realisierung ihren maximalen Wert annimmt:

$$L_\vartheta(x_1, \ldots, x_n) \geq L_{\vartheta'}(x_1, \ldots, x_n)$$

für alle $\vartheta' \in \Theta$, **wenn ein solcher Parameter ϑ eindeutig existiert**.

Dann wird $\vartheta_*(X_1, \ldots, X_n)$ die *Maximum-Likelihood-Schätzung* für den Parameter ϑ.

Maximum-Likelihood-Schätzung

Bemerkung. Zwar existiert der maximale Wert der Likelihood-Funktion in einigen Fällen nicht oder nich eindeutig, aber man hat normaleweise kein solches Problem und der Maximum-Likelihood-Schätzer wird in meisten Situationen definiert.

Maximum-Likelihood-Schätzung

Bemerkung.

- Falls die Likelihood-Funktion nach ϑ differenzierbar ist, dann können ihre Maxima durch Bestimmung der Nullstellen der ersten Ableitung dieser Funktion bestimmt werden.
- Es musst noch überprüft werden, dass die gefundenen Werte tatsächlich Maxima sind (z. B. falls die zweiten Ableitung an einer solchen Stelle negativ ist, dann hat die Funktion ein lokales Maximum hier).
- Aber die Likelihood-Funktion ist ein Produkt, und es ist oft aufwendig, ein solches Produkt von vielen Faktoren abzuleiten, hingegen eine einfache Transformation davon viel leichter zu behandeln ist:

Maximum-Likelihood-Schätzung

Definition. Sei $L_\vartheta(x_1, \dots, x_n)$ eine Likelihood-Funktion. Die *Log-Likelihood-Funktion* ist definiert als

$$l_\vartheta(x_1, \dots, x_n) = \ln L_\vartheta(x_1, \dots, x_n).$$

Bemerkung. Da die Koordinaten einer Realisierung in der Menge S_ϑ der wesentlichen Werte enthalten sind, so ist der Wert von $L_\vartheta(x_1, \dots, x_n)$ für alle Realisierungen positiv und deshalb ist die Log-Likelihood-Funktion immer definiert.

Maximum-Likelihood-Schätzung

Behauptung. Die Log-Likelihood-Funktion ist genau dann maximal, wenn die Likelihood-Funktion maximal ist.

Beweis. Dies folgt aus der Tatsache, dass der Logarithmus eine streng monoton wachsende Funktion auf dem Intervall $(0;1]$ ist.

Maximum-Likelihood-Schätzung

Die Prozedur der Maximum-Likelihood-Schätzung mit der Log-Likelihood-Funktion:

1. Die Likelihood-Funktion $L_\vartheta(x_1, \ldots, x_n)$ wird für alle Realisierungen und Parameter $\vartheta \in \Theta$ aufgestellt.

2. Anhand des natürlichen Logarithmus (Logarithmus zur Basis e) wird die Log-Likelihood-Funktion $l_\vartheta(x_1, \ldots, x_n)$ bestimmt.

3. Die Nullstellen der Ableitung der Log-Likelihood-Funktion nach ϑ werden bestimmt.

4. Es wird überlegt, ob die Nullstellen Maximalstellen sind (z. B. durch den bekannten Test mit der zweiten Ableitung). Falls Θ nicht eine offene Menge (z.B. ein offenes Intervall) ist, dann muss mand auch die Grenzpunkte betrachten, wo das Maximum auch angenommen werden kann!

5. Falls das Maximum für einen eindeutigen ϑ angenommen wird, so ist $\vartheta_*(X_1, \ldots, X_n)$ die Maximum-Likelihood-Schätzung.

Konfidenzbereiche

- Auch bei Schätzern mit günstigen Eigenschaften ist es prinzipiell immer möglich, dass die gemessenen Daten für die zugrunde liegende Verteilung "untypisch" sind, also eher extreme Ergebnisse darstellen.
- In solchen Fällen wird der aus den Daten berechnete Schätzer (trotz guter Methoden) vom echten Wert stark abweichen.
- Mithilfe von *Konfidenzbereichen* kann die Wahrscheinlichkeit von solchen Abweichungen kontrolliert werden.

Konfidenzbereiche

- Diese Bereiche, also unsere Schätzungen werden *Konfidenzintervalle*, in den der echte Wert der Parameter mit großer Wahrscheinlichkeit enthalten wird.
- Dabei wird zunächst immer eine konkrete Zahl $\varepsilon \in (0;1)$ - das sogenannte *Fehlerniveau* - festgelegt, und ein möglichst kurzes Intervall wird dann gewählt, das den Parameter mit Wahrscheinlichkeit $1 - \varepsilon$ enthält.
- Die Zahl ε ist die Wahrscheinlichkeit dafür, dass der Parameter außer dem Intervall liegt, also die Wahrschelinlichkeit des Fehlers.
- Ihr Wert wird typischerweise $\varepsilon = 0{,}05$ oder $\varepsilon = 0{,}01$ (aber andere Werte sind auch möglich).

Konfidenzbereiche

Definition. Sei $\mathbf{X} = (X_1, \ldots, X_n)$ eine einfache Stichprobe, wo die Verteilung der Variablen durch einen Parameter $\vartheta \in \Theta \subset \mathbb{R}^d$ bestimmt ist, und seien noch $\psi : \Theta \to \mathbb{R}$ eine Parameterfunktion und $\varepsilon \in (0; 1)$ eine reelle Zahl. Ein *Konfidenzintervall für* $\psi(\vartheta)$ *zum Konfidenzniveau/Sicherheitsniveau* $1 - \varepsilon$ (oder *zum Fehlerniveau/Irrtumsniveau* ε) (magyarul: $1 - \varepsilon$ (megbízhatósági) szintű konfidenciaintervallum a $\psi(\vartheta)$ paraméterfüggvényre) ist ein Intervall der Form $[T_1(\mathbf{X}); T_2(\mathbf{X})]$, wobei T_1 und T_2 Stichprobenfunktionen sind, und

$$\mathbb{P}_\vartheta(T_1(\mathbf{X}) \leq \psi(\vartheta) \leq T_2(\mathbf{X})) = 1 - \varepsilon$$

für alle $\vartheta \in \Theta$ gilt.

Konfidenzbereiche

Bemerkungen.

- Die Endpunkte des Intervalls sind Zufallsvariablen, aber für jede Realisierung erhält man schon ein konkretes Intervall.
- Die Gleichung

$$\mathbb{P}_\vartheta(T_1(\mathbf{X}) \leq \psi(\vartheta) \leq T_2(\mathbf{X})) = 1 - \varepsilon$$

  kann man im Allgemeinen nur bei stetigen Variablen erreichen, für diskrete Variablen verlangen wir

$$\mathbb{P}_\vartheta(T_1(\mathbf{X}) \leq \psi(\vartheta) \leq T_2(\mathbf{X})) \geq 1 - \varepsilon$$

  in der obigen Definition. In diesem Kurs betrachten wir nur absolut stetige Variablen und können immer mit der obigen Gleichung arbeiten.

Konfidenzbereiche

Beispiel 1: Konfidenzintervall für den Erwartungswert der Normalverteilung bei bekannter Varianz.

- Sei $\mathbf{X} = (X_1, \ldots, X_n)$ eine einfache Stichprobe, wo $X_i \sim \mathrm{N}(\mu; \sigma^2)$ gilt, und die Varianz σ^2 der Verteilung bekannt ist. Wir werden den Erwartungswert μ der Verteilung schätzen.
- Das empirische Mittel $\overline{X}$ ist eine erwartungstreue und stark konsistente Schätzung für den Erwartungswert μ, also ist es naheliegend, ein Konfidenzintervall mit Mittelpunkt $\overline{X}$ wählen.
- Also suchen wir das Intervall $[\overline{X} - r_\varepsilon; \overline{X} + r_\varepsilon]$ mit

$$\mathbb{P}_\mu(\overline{X} - r_\varepsilon \leq \mu \leq \overline{X} + r_\varepsilon) = 1 - \varepsilon.$$

Konfidenzbereiche

Wir ordnen die Ungleichungen in der letzten Wahrscheinlichkeit um:

$$\mathbb{P}_\mu(-r_\varepsilon \leq \overline{X} - \mu \leq r_\varepsilon) = 1 - \varepsilon.$$

Bemerken wir, dass

$$X_1 + \cdots + X_n \sim \mathrm{N}(n\mu; n\sigma^2) \Rightarrow \overline{X} \sim \mathrm{N}(\mu; \sigma^2/n) \Rightarrow \overline{X} - \mu \sim \mathrm{N}(0; \sigma^2/n)$$

gilt, also **hängt der Verteilung der Variablen $\overline{X} - \mu$ von μ nicht ab!!**

Wir standardisieren noch diese Variable und betrachten die Gleichung

$$\mathbb{P}_\mu\left(-\frac{r_\varepsilon\sqrt{n}}{\sigma} \leq \frac{\overline{X} - \mu}{\sigma}\sqrt{n} \leq \frac{r_\varepsilon\sqrt{n}}{\sigma}\right) = 1 - \varepsilon,$$

wo $\frac{\overline{X}-\mu}{\sigma}\sqrt{n}$ (unabhängig von μ) standardnormalverteilt ist.

Konfidenzbereiche

Also ist der Wert auf der linke Seite

$$\Phi\left(\frac{r_\varepsilon\sqrt{n}}{\sigma}\right) - \Phi\left(-\frac{r_\varepsilon\sqrt{n}}{\sigma}\right) = 2\Phi\left(\frac{r_\varepsilon\sqrt{n}}{\sigma}\right) - 1,$$

und wir erhalten die Gleichung

$$2\Phi\left(\frac{r_\varepsilon\sqrt{n}}{\sigma}\right) - 1 = 1 - \varepsilon$$

$$\Phi\left(\frac{r_\varepsilon\sqrt{n}}{\sigma}\right) = 1 - \frac{\varepsilon}{2},$$

also

$$r_\varepsilon = \frac{\sigma\Phi^{-1}\left(1 - \frac{\varepsilon}{2}\right)}{\sqrt{n}}$$

wo Φ^{-1} die Umkehrfunktion der Verteilungsfunktion der Standardnormalverteilung ist.

Konfidenzbereiche

Wir führen die folgende Bezeichnung ein: für eine $\delta \in (0;1)$ ist

$$u_\delta := \Phi^{-1}(1-\delta)$$

das $(1-\delta)$-*Quantil* ($1-\delta$ kvantilis) der Satndardnormalverteilung. Wir bewiesen den folgenden

Satz. Sei $\mathbf{X} = (X_1, \ldots, X_n)$ eine einfache Stichprobe, wo $X_i \sim \mathrm{N}(\mu;\sigma^2)$ gilt, und die Varianz σ^2 der Verteilung bekannt ist, und sei $\varepsilon \in (0;1)$ eine reelle Zahl. Dann ist das Konfidenzintervall für μ zum Konfidenzniveau $1-\varepsilon$ durch

$$\left[\overline{X} - \frac{\sigma u_{\varepsilon/2}}{\sqrt{n}}; \overline{X} + \frac{\sigma u_{\varepsilon/2}}{\sqrt{n}}\right]$$

angegeben, wo $u_{\varepsilon/2}$ ist das $(1-\varepsilon/2)$-Quantil der Standardnormalverteilung ist.

Konfidenzbereiche

Der Radius des Intervalls ist

$$r_\varepsilon = \frac{\sigma u_{\varepsilon/2}}{\sqrt{n}} = \frac{\sigma \Phi^{-1}\left(1 - \frac{\varepsilon}{2}\right)}{\sqrt{n}}.$$

Bemerkungen.

- Je größer der Umfang der Stichprobe ist, desto kleiner ist der Radius. Also wird die Bestimmtheit bei einer größeren Stichprobe größer.
- Je größer die Standardabweichung ist, desto größer ist der Radius. Bei größer Varianz ist nämlich die Dichtefunktion der Normalverteilung flacher.
- Die Funktion Φ und deshalb auch ihre Umkehrfunktion sind monoton wachsend. Also: je kleiner das Fehlerniveau ε ist, desto größer ist der Radius. Mit größerer Sicherheit können wir weniger über den Parameter sagen.

Konfidenzbereiche

Bemerkung. In unserer Argumentation war es bedeutend, dass die ursprüngliche Annahme so umformuliert wurde, dass die Verteilung der in der erhaltenen Gleichung

$$\mathbb{P}_\vartheta(T_\vartheta(\mathbf{X}) \leq t) = 1 - \varepsilon$$

auftretender Stichprobenfunktion $T_\vartheta(\mathbf{X})$ von ϑ unabhängig war. In diesem Fall gilt

$$\mathbb{P}_\vartheta(T_\vartheta(\mathbf{X}) \leq t) = F(t) = 1 - \varepsilon$$

für die Verteilungsfunktion F dieser Verteilung. Falls die Umkehrfunktion F^{-1} von F existiert, so können wir die obige Gleichung lösen.

Konfidenzbereiche

Der Wert dieser Umkehrfunktion bei der Standardnormalverteilung wurde als Quantil genannt, und wir benutzen diesen Namen auch im Allgemeinen:

Definition. Sei $F : \mathbb{R} \to [0;1]$ eine Vertelingsfunktion einer absolut stetigen Zufallsvariablen, für die die Einschränkung

$$F\mid_{(a;b)}: (a;b) \to (0;1)$$

aufs Intervall $(a;b)$ (wo beide von $a = -\infty$ und $b = \infty$ möglich sind) eine Bijektion mit Wertemenge $(0;1)$ ist. (Z.B.: Die Verteilungsfunktionen der Uniform-, Exponential-, Normal- und χ^2-Verteilungen erfüllen diese Eigenschaft.) Dann existiert die (monoton steigende) Umkehrfunktion

$$F^{-1} : (0;1) \to (a;b),$$

und für alle $y \in (0;1)$ ist der Wert $F^{-1}(y)$ das y-*Quantil* der durch F bestimmten Verteilung. Das $\frac{1}{2}$-Quantil $F^{-1}(1/2)$ heißt der *Median* der Verteilung.

Konfidenzbereiche

Beispiel 2: Konfidenzintervall für den Erwartungswert der Normalverteilung bei unbekannter Varianz.

- Sei $\mathbf{X} = (X_1, \ldots, X_n)$ eine einfache Stichprobe, wo $X_i \sim \mathrm{N}(\mu; \sigma^2)$ gilt, und beide Parameter unbekannt sind.
- Wir suchen ein Konfidenzintervall $[\overline{X} - r_\varepsilon; \overline{X} + r_\varepsilon]$ wieder mit Mittelpunkt $\overline{X}$, für die

$$\mathbb{P}_\mu(\overline{X} - r_\varepsilon \leq \mu \leq \overline{X} + r_\varepsilon) = 1 - \varepsilon$$

gilt.
- Wir können wieder die obige Ungleichungen umformen:

$$\mathbb{P}_\mu(-r_\varepsilon \leq \overline{X} - \mu \leq r_\varepsilon) = 1 - \varepsilon,$$

wo $\overline{X} - \mu \sim \mathrm{N}(0, \sigma^2/n)$, aber jetzt können wir diese Variable nicht standardisieren, weil die Standardabweichung nicht bekannt ist.

Konfidenzbereiche

Stattdessen können wir die Standardabweichung durch S_n^* schätzen und die Ungleichung

$$\mathbb{P}_\mu \left(-\frac{r_\varepsilon \sqrt{n}}{S_n^*} \leq \frac{\overline{X} - \mu}{S_n^*} \sqrt{n} \leq \frac{r_\varepsilon \sqrt{n}}{S_n^*} \right) = 1 - \varepsilon,$$

betrachten. Wir werden die Verteilung von $\dfrac{\overline{X} - \mu}{S_n^*} \sqrt{n}$ bestimmen, und wir werden sehen, dass sie von μ nicht abhängt.

Konfidenzbereiche

Erst beschreiben wir die Verteliung der empirischen Varianz $(S_n^*)^2$.

Erinnerung. Seien $X_1, \ldots, X_n \sim N(0; 1)$ unabhängige standardnormalverteilte Zufallsvariablen. Die Verteilung der Quadratsumme $X_1^2 + \cdots + X_n^2$ heißt die χ^2-*Verteilung mit n Freiheitsgraden*.

Bezeichnung: $X_1^2 + \cdots + X_n^2 \sim \chi^2(n)$

Konfidenzbereiche

Behauptung. Die Dichtefunktion einer $\chi^2(n)$-verteilten Variablen Y ist

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

wo

$$\Gamma(a) = \int_0^\infty x^{a-1} e^{-x}\, dx \qquad (a > 0)$$

die Gammafunktion ist. Weiter gelten $\mathbb{E}(Y) = n$ und $\mathbb{D}^2(Y) = 2n$.

Die Werte der y-Quantile der $\chi^2(n)$-Verteilung kann man in einer Verteilungstabelle aussuchen. Eine solche Tabelle befindet sich auf dem letzten Übungsblatt und auf der Webseite.

Konfidenzbereiche

Satz. (Lukács Jenő) Seien $X_1, \ldots, X_n \sim \mathrm{N}(\mu; \sigma^2)$ unabhängige Variablen, dann ist die Variable

$$\frac{(n-1)(S_n^*)^2}{\sigma^2}$$

χ^2-verteilt mit $n-1$ Freiheitsgraden. Weiter, die Variablen $(S_n^*)^2$ und $\overline{X}$ sind unabhängig.

Behauptung. Seien $X \sim \mathrm{N}(0; 1)$ und $Y_n \sim \chi^2(n)$ unabhängige Variablen, dann ist die Dichtefunktion von $Z_n = \dfrac{X}{\sqrt{Y_n}}\sqrt{n}$

$$f_{Z_n}(t) = \frac{1}{\sqrt{\pi n}} \cdot \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \cdot \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}}.$$

Konfidenzbereiche

Definition. Die Verteilung einer absolut stetigen Zufallsvariablen Z_n mit Dichtefunktion

$$f_{Z_n}(t) = \frac{1}{\sqrt{\pi n}} \cdot \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \cdot \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}}$$

heißt *(Student-)t-verteilt mit n Freiheitsgraden*.
Bezeichnung: $Z_n \sim t(n)$.

Bemerkungen.

- $f_{Z_n}(t)$ is eine gerade Funktion für alle n.
- Für $n = 1$ wird die obige Dichtefunktion

  $$f(t) = \frac{1}{\pi} \cdot \frac{1}{1 + t^2}.$$

  Diese Verteilung heißt auch *Cauchy-Verteilung* (siehe auch Übungsblatt 6, Aufgaben 1/c und 16).
- Die Werte der y-Quantile der t-Verteilungen kann man in Verteilungstabellen aussuchen.

Konfidenzbereiche

Korollar. Seien $X_1, \ldots, X_n \sim \mathrm{N}(\mu; \sigma^2)$ unabhängige Zufallsvariablen, dann ist die Zufallsvariable

$$\frac{\overline{X} - \mu}{S_n^*} \sqrt{n}$$

t-verteilt mit $n-1$ Freiheitsgraden.

Beweis. Wir sahen schon, dass $\frac{\overline{X}-\mu}{\sigma}\sqrt{n}$ standardnormalverteilt ist, und nach dem Satz von Lukács gilt $\frac{(n-1)(S_n^*)^2}{\sigma^2} \sim \chi^2(n-1)$, weiter sind diese Variablen unabhängig, weil $\overline{X}$ und $\frac{(n-1)(S_n^*)^2}{\sigma^2}$ unabhängig sind. Wegen der letzen Behauptung folgt

$$\frac{\frac{\overline{X}-\mu}{\sigma}\sqrt{n}}{\frac{\sqrt{n-1}S_n^*}{\sigma}} \sqrt{n-1} = \frac{\overline{X} - \mu}{S_n^*} \sqrt{n} \sim t(n-1).$$

Konfidenzbereiche

Zurück zum Konfidenzintervall:

Wir möchten den Radius r_ε so wählen, dass

$$\mathbb{P}_\mu\left(-\frac{r_\varepsilon\sqrt{n}}{S_n^*} \leq \frac{\overline{X}-\mu}{S_n^*}\sqrt{n} \leq \frac{r_\varepsilon\sqrt{n}}{S_n^*}\right) = 1-\varepsilon$$

gilt. Hier ist die Verteilung von $\frac{\overline{X}-\mu}{S_n^*}\sqrt{n}$ unabhängig von μ (und σ^2), also gilt

$$\begin{aligned}\mathbb{P}_\mu\left(-\frac{r_\varepsilon\sqrt{n}}{S_n^*} \leq \frac{\overline{X}-\mu}{S_n^*}\sqrt{n} \leq \frac{r_\varepsilon\sqrt{n}}{S_n^*}\right) &= \\ = F_{n-1}\left(\frac{r_\varepsilon\sqrt{n}}{S_n^*}\right) - F_{n-1}\left(-\frac{r_\varepsilon\sqrt{n}}{S_n^*}\right),\end{aligned}$$

wo F_n die Verteilungsfunktion der t-Verteilung mit n Freiheitsgraden ist.

Konfidenzbereiche

Die Dichtefunktion der t-Verteilung ist gerade, und nach dem Beweis der entsprechenden Formel für die Funktion Φ sahen wir, dass die Formel

$$F_n(-t) = 1 - F_n(t)$$

in diesem Fall gilt.

Also ergibt sich

$$\mathbb{P}_\mu \left(-\frac{r_\varepsilon \sqrt{n}}{S_n^*} \leq \frac{\overline{X} - \mu}{S_n^*} \sqrt{n} \leq \frac{r_\varepsilon \sqrt{n}}{S_n^*} \right) = 2F_{n-1} \left(\frac{r_\varepsilon \sqrt{n}}{S_n^*} \right) - 1 = 1 - \varepsilon,$$

d.h.

$$F_{n-1} \left(\frac{r_\varepsilon \sqrt{n}}{S_n^*} \right) = 1 - \frac{\varepsilon}{2} \iff r_\varepsilon = \frac{S_n^* t_{n-1,\varepsilon/2}}{\sqrt{n}},$$

wo $t_{n,\delta} := F_n^{-1}(1-\delta)$ das $(1-\delta)$-Quantil der t-Verteilung mit n Freiheitsgraden ist.

Konfidenzbereiche

Wir erhalten den folgenden

Satz. Sei $\mathbf{X} = (X_1, \ldots, X_n)$ eine einfache Stichprobe, wo $X_i \sim \mathrm{N}(\mu; \sigma^2)$ gilt, und die Parameter der Verteilung unbekannt sind, und sei $\varepsilon \in (0; 1)$ eine reelle Zahl. Dann ist das Konfidenzintervall für μ zum Konfidenzniveau $1 - \varepsilon$ durch

$$\left[\overline{X} - \frac{S_n^* t_{n-1,\varepsilon/2}}{\sqrt{n}}; \overline{X} + \frac{S_n^* t_{n-1,\varepsilon/2}}{\sqrt{n}}\right]$$

angegeben, wo $t_{n-1,\varepsilon/2}$ ist das $(1 - \varepsilon/2)$-Quantil der t-Verteilung mit $n - 1$ Freiheitsgraden ist.

Konfidenzbereiche

Beispiel 3: Konfidenzintervall für die Varianz der Normalverteilung bei unbekanntem Erwartungswert

- Sei $\mathbf{X} = (X_1, \ldots, X_n)$ eine einfache Stichprobe, wo $X_i \sim \mathrm{N}(\mu; \sigma^2)$ gilt, und die Parameter der Verteilung unbekannt sind.
- Im Gegensatz zum Fall des Erwartungswertes sind die möglichen Werte der Varianz nicht-negativ, wir suchen also ein Intervall $[a_\varepsilon; b_\varepsilon]$, wo $0 < a_\varepsilon < b_\varepsilon$.
- Ohne Beweis benutzen wir, dass das Konfidenzintervall genau dann möglichst kurz wird, wenn

$$\mathbb{P}_{\sigma^2}(0 < \sigma^2 < a_\varepsilon) = \mathbb{P}_{\sigma^2}(\sigma^2 > b_\varepsilon) = \frac{\varepsilon}{2}$$

gilt.

Konfidenzbereiche

Wir formulieren diese Ungleichungen folgenderweise um:

$$\mathbb{P}_{\sigma^2}\left(\frac{(n-1)(S_n^*)^2}{\sigma^2} > \frac{(n-1)(S_n^*)^2}{a_\varepsilon}\right) = \\ = \mathbb{P}_{\sigma^2}\left(\frac{(n-1)(S_n^*)^2}{\sigma^2} < \frac{(n-1)(S_n^*)^2}{b_\varepsilon}\right) = \frac{\varepsilon}{2},$$

hier gilt $\frac{(n-1)(S_n^*)^2}{\sigma^2} \sim \chi^2(n-1)$, und diese Verteilung ist von σ^2 unabhängig.

Konfidenzbereiche

Für eine reelle Zahl $\delta \in (0;1)$ sei $\chi^2_{n,\delta}$ das $(1-\delta)$-Quantil der χ^2-Verteilung mit n Freiheitsgraden. Dann

$$\mathbb{P}_{\sigma^2}\left(\frac{(n-1)(S_n^*)^2}{\sigma^2} < \frac{(n-1)(S_n^*)^2}{a_\varepsilon}\right) = 1 - \frac{\varepsilon}{2},$$

also

$$\frac{(n-1)(S_n^*)^2}{a_\varepsilon} = \chi^2_{n-1,\varepsilon/2},$$

und

$$\mathbb{P}_{\sigma^2}\left(\frac{(n-1)(S_n^*)^2}{\sigma^2} < \frac{(n-1)(S_n^*)^2}{b_\varepsilon}\right) = \frac{\varepsilon}{2}$$

deshalb

$$\frac{(n-1)(S_n^*)^2}{b_\varepsilon} = \chi^2_{n-1,1-\varepsilon/2}.$$

Konfidenzbereiche

Wir erhalten den folgenden

Satz. Sei $\mathbf{X} = (X_1, \ldots, X_n)$ eine einfache Stichprobe, wo $X_i \sim \mathrm{N}(\mu; \sigma^2)$ gilt, und die Parameter der Verteilung unbekannt sind, und sei $\varepsilon \in (0; 1)$ eine reelle Zahl. Dann ist das Konfidenzintervall für σ^2 zum Konfidenzniveau $1 - \varepsilon$ durch

$$\left[\frac{(n-1)(S_n^*)^2}{\chi^2_{n-1,\varepsilon/2}}; \frac{(n-1)(S_n^*)^2}{\chi^2_{n-1,1-\varepsilon/2}} \right]$$

angegeben, wo $\chi^2_{n-1,\delta}$ ist das $(1-\delta)$-Quantil der χ^2-Verteilung mit $n-1$ Freiheitsgraden ist.

Konfidenzbereiche

Aufgabe. Konstruieren wir das Konfidenzintervall für die Varianz der Normalverteilung bei bekanntem Erwartungswert.