Schätzer

Unterscheide (konzeptionell):

- Die Abbildung $\widehat{\vartheta}_n = \widehat{\vartheta}_n(x)$,

$$x = (x_1 \ldots, x_n) \mapsto \widehat{\vartheta}_n(x),$$

  die jeder <u>Realisation</u> x des Stichprobenraums $\mathcal{X}$ einen Schätzwert zuordnet; gedanklich nach Durchführung des Zufallsexperiments.

- Die Abbildung $\widehat{\Theta}_n = \widehat{\vartheta}_n(X)$,

$$X = (X_1, \ldots, X_n) \mapsto \widehat{\vartheta}_n(X),$$

  die jedem <u>(zufälligen) Vektor</u> X die Zufallsgröße $\widehat{\vartheta}_n(X)$ zuordnet; gedanklich vor Durchführung des Zufallsexperiments).

Es ist üblich, in beiden Fällen $\widehat{\vartheta}_n$ zu schreiben und von einem 'Schätzer' zu sprechen. Ob die Statistik (als Zufallsvariable bzw. Zufallsvektor) oder eine Realisation derselben gemeint ist, muss aus dem Kontext erschlossen werden.

Gütekriterien: Erwartungstreue

- Sei $\widehat{\vartheta}_n = \widehat{\vartheta}_n(X)$ ein Schätzer für ϑ.
- Da $\widehat{\vartheta}_n = T_n(X_1, \dots, X_n)$ von den Zufallsvariablen $X_1, \dots, X_n$ abhängt, ist $\widehat{\vartheta}_n$ zufällig, streut also.
- Frage: Um welchen Wert streut der Schätzer?
- Berechne den Erwartungswert:

$$E(\widehat{\vartheta}_n) = E(T_n(X_1, \dots, X_n)) = ...?$$

- Das Ergebnis der Berechnung hängt von der Verteilung der $X_i \sim F_\vartheta$ ab! Um diese Abhängigkeit zum Ausdruck zu bringen schreibt man mitunter $E_\vartheta(\cdots)$ statt $E(\cdots)$.
- Im Allgemeinen ist $E(\widehat{\vartheta}_n)$ daher eine Funktion des Parameters ϑ!

Erwartungstreue

Ein Schätzer $\widehat{\vartheta}_n$ heißt **erwartungstreu für** ϑ, wenn für alle $\vartheta \in \Theta$ gilt:

$$E(\widehat{\vartheta}_n) = \vartheta$$

$g(\widehat{\vartheta}_n)$ heißt **erwartungstreu für** $g(\vartheta)$, wenn für alle $\vartheta \in \Theta$ gilt:

$$E(g(\widehat{\vartheta}_n)) = g(\vartheta)$$

Sinngemäß gelten diese Definitionen auch für nichtparametrische Modelle: T_n heißt erwartungstreu für eine Kenngröße $g(F)$, wenn $E(T_n) = E_F(T_n) = g(F)$ für alle Verteilungsfunktionen F der betrachteten Verteilungsklasse. Hierbei deutet $E_F(\cdot)$ an, dass der EW unter der Annahme $X_i \sim F$ berechnet wird.

Erwartungstreue

Beispiele: a) $X_1, \ldots, X_n$ seinen unabhängig und identisch verteilt mit Erwartungswert $\mu \in \mathbb{R}$. $\overline{X}_n$ ist erwartungstreu für μ.

b) Parameter: $\vartheta = g(\mu) = \mu^2$.
$g(\overline{X}_n) = (\overline{X}_n)^2$ ist nicht erwartungstreu für $\vartheta = g(\mu) = \mu^2$.

c) $X_1, \ldots, X_n \sim U(0, \vartheta)$ mit $\vartheta > 0$ unbekannt.

Der ML-Schätzer $\widehat{\vartheta}_n = \max_{i=1,\ldots,n} X_i$ für ϑ ist nicht erwartungstreu, aber der Schätzer

$$\widehat{\vartheta}_n^* = \frac{n+1}{n}\widehat{\vartheta}_n$$

Erwartungstreue: Anschauung

Anschauung:

- Wende erwartungstreuen Schätzer N Mal auf Stichproben vom Umfang n an.
- N Schätzungen: $\widehat{\vartheta}_n(1), \dots, \widehat{\vartheta}_n(N)$.
- Wende Gesetz der großen Zahlen an!

$$\frac{1}{N} \sum_{i=1}^{N} \widehat{\vartheta}(i) \to E(\widehat{\vartheta}_n(i)) = E(\widehat{\vartheta}_n)$$

- $\widehat{\vartheta}_n$ erwartungstreu: rechte Seite ist ϑ unabhängig von $\vartheta \in \Theta$.
- sonst: rechte Seite $\neq \vartheta$.

Werden Schätzungen aus einer täglichen Stichprobe vom Umfang n über einen langen Zeitraum gemittelt, so schwankt dieses Mittel um $E(\widehat{\vartheta}_n)$. Bei einer erwartungstreuen Schätzfunktion also um den wahren Wert ϑ.

Bias (Verzerrung)

Verzerrung (Bias)

Die **Verzerrung** (engl.: *bias*) wird gemessen durch

$$\mathrm{Bias}(\widehat{\vartheta}_n; \vartheta) = E_\vartheta(\widehat{\vartheta}) - \vartheta.$$

Beispiele:
$X_1, \ldots, X_n$ seinen unabhängig und identisch verteilt mit EW μ und Varianz $\sigma^2 > 0$.

Der Bias von $(\overline{X}_n)^2$ bzgl. des Parameters μ^2 ist:

$$\text{Bias}((\overline{X}_n)^2; \mu^2) = \frac{\sigma^2}{n}$$

Gütekriterien: Asymptotische Erwartungstreue

(Asymptotische) Erwartungstreue, Unverfälschtheit

Ein Schätzer $\widehat{\vartheta}_n$ für einen Parameter ϑ heißt **asymptotisch erwartungstreu für** ϑ, wenn für alle ϑ

$$E_\vartheta(\widehat{\vartheta}_n) \to \vartheta,$$

gilt.

Gütekriterien: Konsistenz

Das Gütekriterium der **Konsistenz** fragt danach, ob bei wachsendem Stichprobenumfang n die Wahrscheinlichkeit gegen 1 strebt, dass der Unterschied zwischen Schätzer $\widehat{\vartheta}_n$ und wahrem Wert ϑ kleiner als eine beliebig vorgegebene Toleranz $\delta > 0$ ist:
Für beliebiges $\delta > 0$ gilt:

$$P(|\widehat{\vartheta}_n - \vartheta| \leq \delta) \to 1, \qquad n \to \infty$$

oder gleichbedeutend:

$$P(|\widehat{\vartheta}_n - \vartheta| > \delta) \to 0, \qquad n \to \infty$$

Diese Eigenschaft entspricht der **stochastischen Konvergenz**:

$$\widehat{\vartheta}_n \xrightarrow{P} \vartheta, \qquad n \to \infty.$$

Gütekriterien: Konsistenz

Konsistenz

Ein Schätzer $\widehat{\vartheta}_n = T(X_1, \dots, X_n)$ basierend auf einer Stichprobe vom Umfang n heißt **(schwach) konsistent für** ϑ, falls

$$\widehat{\vartheta}_n \stackrel{P}{\to} \vartheta, \qquad n \to \infty,$$

Gilt sogar fast sichere Konvergenz, dann heißt $\widehat{\vartheta}_n$ **stark konsistent für** ϑ.

1. Ist $\widehat{\vartheta}_n$ konsistent für ϑ und ist g stetig, dann ist $g(\widehat{\vartheta}_n)$ konsistent für den abgeleiteten Parameter $g(\vartheta)$.
2. Die obige Aussage gilt auch für vektorwertige Parameter und ihre Schätzer. Insbesondere folgt aus der Konsistenz von $\widehat{\vartheta}_n$ für ϑ und $\widehat{\xi}_n$ für ξ die Konsistenz von $\widehat{\vartheta}_n \pm \widehat{\xi}_n$ für $\vartheta \pm \xi$.

Konsistenz: Beispiele

1. $X_1, \dots, X_n$ i.i.d. mit $\mu = E(X_1)$. Dann ist $\widehat{\mu}_n = \overline{X}_n$ konsistent für μ.
2. $g(\overline{X}_n) = (\overline{X}_n)^2$ ist konsistent für den abgeleiteten Parameter $g(\mu) = \mu^2$.
3. Gilt $E(X_1^2) < \infty$, dann folgt (starkes Gesetz der großen Zahlen):

$$\widehat{m}_{2,n} = \frac{1}{n} \sum_{i=1}^{n} X_i^2$$

ist (stark) konsistent für das zweite Moment $m_2 = E(X_1^2)$.
Dann ist auch die Stichprobenvarianz

$$\widehat{\sigma}_n^2 = \widehat{m}_{2,n} - \widehat{\mu}_n^2 = \frac{1}{n} \sum_{i=1}^{n} X_i^2 - (\overline{X}_n)^2$$

konsistent für $\sigma^2 = E(X_1^2) - (E(X_1))^2 = \mathrm{Var}(X_1)$.

Konsistenz: Beispiele

Schätzung der Varianz σ^2: (s. Basiswissen, S. 192)
$X_1, \dots, X_n$ einfache Zufallsstichprobe mit $\mu = E(X_1)$, $\sigma^2 = \mathrm{Var}(X_1) < \infty$.
Stichprobenvarianz:

$$\widehat{\sigma}_n^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X}_n)^2$$

Dieser Schätzer ist konsistent aber nicht erwartungstreu:

$$E(\widehat{\sigma}_n^2) = \frac{n-1}{n} \sigma^2 = \sigma^2 - \frac{\sigma^2}{n}$$

Im Mittel wird σ^2 unterschätzt. Man verwendet daher

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

Dieser Schätzer ist konsistent und erwartungstreu für σ^2.

Effizienz und mittlerer quadratischer Fehler

- Mitunter stehen mehrere Schätzfunktionen zur Auswahl.
- Angenommen, alle sind erwartungstreu. Welche sollte man nehmen?

Effizienz

1. Sind T_1 und T_2 zwei erwartungstreue Schätzer für ϑ und gilt $\mathrm{Var}(T_1) < \mathrm{Var}(T_2)$, so heißt T_1 **effizienter** als T_2.
2. T_1 ist **effizient**, wenn T_1 effizienter als jede andere erwartungstreue Schätzfunktion ist.

Beispiele

Beispiel: $X_1, \ldots, X_n$ sei eine einfache Stichprobe. Betrachte

$$T_1 = \frac{X_1 + X_n}{2}, \qquad T_2 = \overline{X}_n.$$

Welche Schätzfunktion ist effizienter für die Schätzung von μ?

Beide Schätzfunktionen sind erwartungstreu: $E(T_2) = E(\overline{X}_n) = \mu$ und

$$E(T_1) = \frac{1}{2}E(X_1 + X_n) = \frac{1}{2}(E(X_1) + E(X_n)) = \frac{2\mu}{2} = \mu$$

Vergleich der Varianzen:

$$\mathrm{Var}(T_1) = \frac{\sigma^2}{2}, \qquad \mathrm{Var}(T_2) = \frac{\sigma^2}{n}$$

Für $n > 2$ ist T_2 effizienter als T_1.

Beispiele

Beispiel

Gelte $X_1, \ldots, X_n \sim G[0, \vartheta]$, $\vartheta > 0$ *unbekannt.*
Zwei erwartungstreue Schätzer für ϑ*:*

$$T_1 = 2\overline{X} \quad \textit{und} \quad T_2 = \frac{n+1}{n} \max(X_1, \ldots, X_n).$$

Welche Schätzfunktion ist effizienter?

MSE

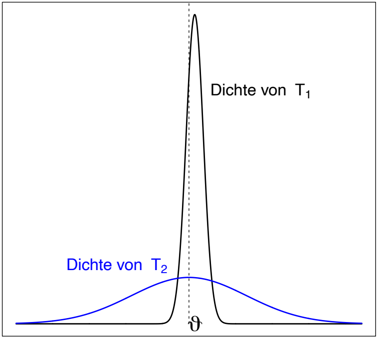

Abbildung: Dargestellt sind Dichten von zwei Schätzern zur Schätzung des Parameters ϑ. T_1 ist zwar verzerrt, hat aber eine viel kleinere Streuung.

MSE

MSE: Mean Squared Error

Der MSE ist das wichtigste Gütemaß für Bewertung und Vergleiche von Schätzern. Er integriert die Varianz (als Streuungsmaß) und den Bias in einer Kennzahl.

MSE

$$\mathrm{MSE}(\widehat{\vartheta}_n; \vartheta) = E_\vartheta(\widehat{\vartheta}_n - \vartheta)^2$$

Additive Zerlegung

Ist $\widehat{\vartheta}_n$ eine Schätzfunktion mit $\mathrm{Var}_\vartheta(\widehat{\vartheta}_n) < \infty$, dann gilt die additive Zerlegung

$$\mathrm{MSE}(\widehat{\vartheta}_n; \vartheta) = \mathrm{Var}_\vartheta(\widehat{\vartheta}_n) + [\mathrm{Bias}(\widehat{\vartheta}_n; \vartheta)]^2.$$

MSE

MS-Effizienz

1. Sind T_1 und T_2 zwei Schätzer für ϑ und gilt $\mathrm{MSE}(T_1; \vartheta) < \mathrm{MSE}(T_2; \vartheta)$, so heißt T_1 **effizienter** als T_2.
2. T_1 ist **effizient**, wenn T_1 effizienter als jede andere erwartungstreue Schätzfunktion ist.

Beispiel: $X_1, \ldots, X_n \sim G(0, \vartheta)$. Effizienzvergleich[1] von

$$T_1 = 2\overline{X}_n, \qquad T_2 = \frac{n+1}{n} \max_{1 \leq i \leq n} X_i.$$

Schritt 1: Berechne $MSE(T_1; \vartheta)$:
Erwartungswert und Varianz von T_1:

$$E(T_1) = \vartheta, \qquad \mathrm{Var}(T_1) = 4\mathrm{Var}(\overline{X}_n) = 4\frac{\sigma^2}{n}$$

mit $\sigma^2 = \mathrm{Var}(X_i) = \frac{\vartheta^2}{12}$. Also: $\mathrm{Var}(T_1) = \frac{\vartheta^2}{3n}$. Damit ist

$$MSE(T_1; \vartheta) = \frac{\vartheta^2}{3n}$$

[1] Diese Aufgabe ist eine gute Übung für das Zusammenspiel von Erwartungswerten, Varianzen, Termumformungen und Berechnung von Integralen (Arbeiten mit Dichten)!

Schritt 2: Berechne $MSE(T_2; \vartheta)$:
Berechne die Varianz von $Z = \max_{1 \leq i \leq n} X_i$ mit Verschiebungssatz:

$$\mathrm{Var}(Z) = E(Z^2) - (E(Z))^2$$

Oben schon berechnet: $E(Z) = E(X_{(n)}) = \frac{n}{n+1}\vartheta$ und

$$f_Z(x) = \frac{n}{\vartheta^n} x^{n-1} \mathbf{1}_{(0,\vartheta)}(x), \qquad x \in \mathbb{R}.$$

$$\begin{aligned}
\Rightarrow E(Z^2) &= \int_{-\infty}^{\infty} x^2 \cdot f_Z(x)\,dx = \int_0^{\vartheta} x^2 \frac{n}{\vartheta^n} x^{n-1}\,dx \\
&= \frac{n}{\vartheta^n} \int_0^{\vartheta} x^{n+1}\,dx = \frac{n}{\vartheta^n} \frac{\vartheta^{n+2}}{n+2} \\
&= \frac{n}{n+2} \vartheta^2
\end{aligned}$$

MSE

Fs. Schritt 2:

$$\mathrm{Var}(Z) = E(Z^2) - (E(Z))^2 = \frac{n}{n+2}\vartheta^2 - \left(\frac{n}{n+1}\vartheta\right)^2$$
$$= \vartheta^2 \left(\frac{n}{n+2} - \frac{n^2}{(n+1)^2}\right)$$
$$= \vartheta^2 \left(\frac{n(n+1)^2 - (n+2)n^2}{(n+2)(n+1)^2}\right)$$

Vereinfachen des Ausdrucks im Zähler des Bruchs:

$$n(n+1)^2 - (n+2)n^2 = n(n^2 + 2n + 1) - (n^3 + 2n^2)$$
$$= (n^3 + 2n^2 + n) - n^3 - 2n^2 = n.$$

Damit folgt:

$$\mathrm{Var}(Z) = \vartheta^2 \frac{n}{(n+2)(n+1)^2}$$

Fs. Schritt 2:

$$\mathrm{Var}(Z) = \vartheta^2 \frac{n}{(n+2)(n+1)^2}$$

Mit $T_2 = \frac{n+1}{n} Z$ ergibt sich

$$\mathrm{Var}(T_2) = \frac{(n+1)^2}{n^2} \cdot \vartheta^2 \frac{n}{(n+2)(n+1)^2} = \frac{\vartheta^2}{n(n+2)}$$

Da T_2 erwartungstreu für ϑ ist, ergeben sich also die folgenden MSEs:

$$MSE(T_1; \vartheta) = \frac{\vartheta^2}{3n} \qquad MSE(T_2; \vartheta) = \frac{\vartheta^2}{n(n+2)}$$

Schritt 4: Vergleich der Ausdrücke:

$$MSE(T_1; \vartheta) > MSE(T_2; \vartheta) \Leftrightarrow \frac{\vartheta^2}{3n} > \frac{\vartheta^2}{n(n+2)} \Leftrightarrow n^2 + 2n > 3n$$

Dies ist für alle $n > 1$ der Fall (für $n = 1$ sind die Ausdrücke gleich).