



Inferenzstatistik

BSc Psychologie SoSe 2021

Prof. Dr. Dirk Ostwald

(8) Schätzeigenschaften

Die Lektüre von Georgii (2009, Abschnitte 7.4, 7.5, und 7.7) wird empfohlen. Die hier gewählte Darstellung von Schätzereigenschaften bei endlichen Stichproben folgt Wasserman (2004, Chapter 9), Held and Sabanés Bové (2014, Sections 3.1, 3.2, 4.1, 4.2), und DeGroot and Schervish (2012, Section 8.8). Die hier gewählte Darstellung von asymptotischen Schätzereigenschaften folgt Casella and Berger (2012, Section 10.1).

Schätzereigenschaften

- Vorbemerkungen
- Schätzereigenschaften bei endlichen Stichproben
- Asymptotische Schätzereigenschaften
- Eigenschaften von Maximum-Likelihood Schätzern
- Selbstkontrollfragen

Schätzereigenschaften

- **Vorbemerkungen**
- Schätzereigenschaften bei endlichen Stichproben
- Asymptotische Schätzereigenschaften
- Eigenschaften von Maximum-Likelihood Schätzern
- Selbstkontrollfragen

Vorbemerkungen

Wir gehen von einem statistischem parametrischem Produktmodell $\mathcal{M} := \{\mathcal{X}, \mathcal{A}, \{\mathbb{P}_\theta | \theta \in \Theta\}\}$ mit n -dimensionalen Stichprobenraum (z.B. $\mathcal{X} := \mathbb{R}^n$), d -dimensionalen Parameterraum $\Theta \subset \mathbb{R}^d$ und gegebener WMF oder WDF p_θ für alle $\theta \in \Theta$ aus. $X_1, \dots, X_n \sim p_\theta$ bezeichnet die zu $\mathcal{M}$ gehörende Stichprobe unabhängig und identisch verteilter Zufallsvariablen, es gilt also $X_0 \sim p_\theta$ und $X_i \sim p_\theta$ für alle $i = 1, \dots, n$.

Für einen Messraum $(\Sigma, \mathcal{S})$ sei $\hat{\tau} : \mathcal{X} \rightarrow \Sigma$ ein Schätzer von $\tau : \Theta \rightarrow \Sigma$. Wir betrachten Erwartungswerts-Varianz-, und Standardabweichungsschätzer, also Schätzer für

$$\tau : \Theta \rightarrow \Sigma, \theta \mapsto \tau(\theta) \text{ mit } \tau(\theta) := \mathbb{E}_\theta(X_0), \tau(\theta) := \mathbb{V}_\theta(X_0), \text{ und } \tau(\theta) := \mathbb{S}_\theta(X_0) \quad (1)$$

respektive, sowie Parameterschätzer, also Schätzer für

$$\tau : \Theta \rightarrow \Sigma, \tau(\theta) := \theta. \quad (2)$$

In dieser Vorlesungseinheit führen wir frequentistische Schätzereigenschaften ein. Frequentistische Schätzereigenschaften betrachten die Verteilung der Schätzwerte $\hat{\tau}(x_1, \dots, x_n)$ in Abhängigkeit von der Verteilung der Stichprobenwerte $x_1, \dots, x_n$. Weil die Stichprobenwerte zufällig sind, sind auch die Schätzwerte zufällig; ein Schätzer $\hat{\tau}$ ist also eine Zufallsvariable.

Wir unterscheiden zwischen *Schätzereigenschaften bei endlichen Stichproben*, d.h. Eigenschaften von $\hat{\tau}_n$ für ein fixes $n \in \mathbb{N}$ (z.B. $n = 12$) und *Asymptotischen Schätzereigenschaften*, d.h. Eigenschaften von $\hat{\tau}_n$ für unendlich groß werdende Stichproben mit $n \rightarrow \infty$.

Schätzereigenschaften

- Vorbemerkungen
- **Schätzereigenschaften bei endlichen Stichproben**
- Asymptotische Schätzereigenschaften
- Eigenschaften von Maximum-Likelihood Schätzern
- Selbstkontrollfragen

Schätzereigenschaften bei endlichen Stichproben

Ein Schätzer $\hat{\tau}_n$ heißt **erwartungstreu**, wenn sein Erwartungswert dem wahren, aber unbekannten, Wert $\tau(\theta)$ für alle $\theta \in \Theta$ gleicht.

Die **Varianz** eines Schätzers $\hat{\tau}_n$ ist die Varianz der Zufallsvariable $\hat{\tau}_n(X)$; der **Standardfehler** eines Schätzers $\hat{\tau}_n$ ist die Standardabweichung der Zufallsvariable $\hat{\tau}_n(X)$.

Die **Cramér-Rao-Ungleichung** gibt eine untere Schranke für die Varianz erwartungstreuer Schätzer an. Ein erwartungstreuer Schätzer mit Varianz gleich der in der Cramér-Rao-Ungleichung gegebenen unteren Schranke hat die kleinstmögliche Varianz aller erwartungstreuen Schätzer und ist in diesem Sinne ein “optimaler” Schätzer.

Der **mittlere quadratische Fehler** von $\hat{\tau}_n$ ist der Erwartungswert der quadrierten Abweichung von $\hat{\tau}_n(X)$ von $\tau(\theta)$ über Stichproben vom Umfang n .

Schätzereigenschaften bei endlichen Stichproben

Ein Schätzer $\hat{\tau}_n$ heißt **erwartungstreu**, wenn sein Erwartungswert dem wahren, aber unbekannten, Wert $\tau(\theta)$ für alle $\theta \in \Theta$ gleicht.

Die **Varianz** eines Schätzers $\hat{\tau}_n$ ist die Varianz der Zufallsvariable $\hat{\tau}_n(X)$; der **Standardfehler** eines Schätzers $\hat{\tau}_n$ ist die Standardabweichung der Zufallsvariable $\hat{\tau}_n(X)$.

Die **Cramér-Rao-Ungleichung** gibt eine untere Schranke für die Varianz erwartungstreuer Schätzer an. Ein erwartungstreuer Schätzer mit Varianz gleich der in der Cramér-Rao-Ungleichung gegebenen unteren Schranke hat die kleinstmögliche Varianz aller erwartungstreuen Schätzer und ist in diesem Sinne ein “optimaler” Schätzer.

Der **mittlere quadratische Fehler** von $\hat{\tau}_n$ ist der Erwartungswert der quadrierten Abweichung von $\hat{\tau}_n(X)$ von $\tau(\theta)$ über Stichproben vom Umfang n .

Definition (Fehler, Systematischer Fehler, und Erwartungstreue)

$X_1, \dots, X_n \sim p_\theta$ sei eine Stichprobe und $\hat{\tau}_n$ sei ein Schätzer für τ .

- Der *Fehler* von $\hat{\tau}_n$ ist definiert als

$$\hat{\tau}_n(X) - \tau(\theta). \quad (3)$$

- Der *systematische Fehler (Bias)* von $\hat{\tau}_n$ ist definiert als

$$B(\hat{\tau}_n) := \mathbb{E}_\theta(\hat{\tau}_n(X)) - \tau(\theta). \quad (4)$$

- $\hat{\tau}_n$ heißt *erwartungstreu (unbiased)*, wenn

$$B(\hat{\tau}_n) = 0 \Leftrightarrow \mathbb{E}_\theta(\hat{\tau}_n(X)) = \tau(\theta) \text{ für alle } \theta \in \Theta, n \in \mathbb{N}. \quad (5)$$

Andernfalls heißt $\hat{\tau}_n$ *verzerrt (biased)*.

Bemerkungen

- Der Fehler hängt von einer Realisation der Stichprobe ab.
- Der systematische Fehler ist der erwartete Fehler über viele Stichprobenrealisationen.
- Ein Parameterschätzer ist erwartungstreu, wenn $\mathbb{E}_\theta(\hat{\theta}_n(X)) = \theta$.

Theorem (Erwartungstreue von Stichprobenmittel und Stichprobenvarianz)

$X_1, \dots, X_n \sim p_\theta$ sei die Stichprobe eines statistischen parametrischen Produktmodells $\mathcal{M}$. Dann gilt:

- Das Stichprobenmittel

$$\bar{X}_n := \frac{1}{n} \sum_{i=1}^n X_i \quad (6)$$

ist ein erwartungstreuer Schätzer des Erwartungswerts $\mathbb{E}_\theta(X_0)$.

- Die Stichprobenvarianz

$$S_n^2 := \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2 \quad (7)$$

ist ein erwartungstreuer Schätzer der Varianz $\mathbb{V}_\theta(X_0)$.

Beweis

Um die Notation zu vereinfachen, definieren wir $\mathbb{E} := \mathbb{E}_\theta$ und $\mathbb{V} := \mathbb{V}_\theta$. Mit der Linearität von Erwartungswerten ergibt sich dann

$$\mathbb{E}(\bar{X}_n) = \mathbb{E}\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n} \sum_{i=1}^n \mathbb{E}(X_i) = \frac{1}{n} \sum_{i=1}^n \mathbb{E}(X_0) = \frac{1}{n} n \mathbb{E}(X_0) = \mathbb{E}(X_0).$$

Dies zeigt die Erwartungstreue des Stichprobenmittels als Schätzer des Erwartungswertes.

Um die Erwartungstreue der Stichprobenvarianz zu zeigen, halten wir zunächst fest, dass

$$\mathbb{V}(\bar{X}_n) = \mathbb{V}\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n^2} \sum_{i=1}^n \mathbb{V}(X_i) = \frac{1}{n^2} \sum_{i=1}^n \mathbb{V}(X_0) = \frac{1}{n^2} n \mathbb{V}(X_0) = \frac{\mathbb{V}(X_0)}{n}.$$

gilt. Weiterhin halten wir ohne Beweis fest, dass

$$\sum_{i=1}^n (X_i - \bar{X}_n)^2 = \sum_{i=1}^n (X_i - \mu - \bar{X}_n + \mu)^2 = \sum_{i=1}^n (X_i - \mu)^2 - n(\bar{X}_n - \mu)^2.$$

Beweis

Es ergibt sich somit

$$\begin{aligned}\mathbb{E}\left((n-1)S_n^2\right) &= \mathbb{E}\left(\sum_{i=1}^n (X_i - \bar{X}_n)^2\right) \\&= \mathbb{E}\left(\sum_{i=1}^n (X_i - \mu)^2 - n(\bar{X}_n - \mu)^2\right) \\&= \sum_{i=1}^n \mathbb{E}\left((X_i - \mu)^2\right) - n\mathbb{E}\left((\bar{X}_n - \mu)^2\right) \\&= n\mathbb{V}(X_0) - n\mathbb{V}(\bar{X}_n) \\&= n\mathbb{V}(X_0) - n\frac{\mathbb{V}(X_0)}{n} \\&= n\mathbb{V}(X_0) - \mathbb{V}(X_0) \\&= (n-1)\mathbb{V}(X_0)\end{aligned}$$

Schließlich ergibt sich

$$\mathbb{E}(S_n^2) = \mathbb{E}\left(\frac{1}{n-1}(n-1)S_n^2\right) = \frac{1}{n-1}\mathbb{E}\left((n-1)S_n^2\right) = \frac{1}{n-1}(n-1)\mathbb{V}(X_0) = \mathbb{V}(X_0)$$

und damit die Erwartungstreue der Stichprobenvarianz als Schätzer der Varianz.

□

Theorem (Verzerrtheit der Stichprobenstandardabweichung)

$X_1, \dots, X_n \sim p_\theta$ sei die Stichprobe eines statistischen parametrischen Produktmodells $\mathcal{M}$. Dann ist die Stichprobenstandardabweichung

$$S_n := \sqrt{S_n^2} \quad (8)$$

ein verzerrter Schätzer der Standardabweichung $\mathbb{S}_\theta(X_0)$.

Beweis

Wir halten zunächst fest, dass $\sqrt{\cdot}$ eine strikt konkave Funktion und $\sigma^2 > 0$ ist. Dann aber gilt mit der Jensenschen Ungleichung $\mathbb{E}(f(X)) < f(\mathbb{E}(X))$ für strikt konkave Funktionen, dass

$$\mathbb{E}(S) = \mathbb{E}\left(\sqrt{S_n^2}\right) < \sqrt{\mathbb{E}(S_n^2)} = \sqrt{\mathbb{V}_\theta(X_0)} = \mathbb{S}_\theta(X_0). \quad (9)$$

□

Bemerkung

- Nichtlineare Transformationen von erwartungstreuen Schätzern liefern oft verzerrte Schätzer.

Simulation ($X_1, \dots, X_{12} \sim N(\mu, \sigma^2)$ mit $n = 12$, $\mu = 1.7$, $\sigma^2 = 2$, $\sigma \approx 1.41$)

```
# Modellformulierung
mu      = 1.7                # wahrer, aber unbekannter, Erwartungswertparameter
sigsqr  = 2                  # wahrer, aber unbekannter, Varianzparameter
n       = 12                 # Stichprobengroesse n
ns      = 1e4                # Anzahl der Simulationen
X_bar   = rep(NaN, ns)       # Stichprobenmittellarray
S_sqr   = rep(NaN, ns)       # Stichprobenvarianzarray
S       = rep(NaN, ns)       # Stichprobenstandardabweichungarray

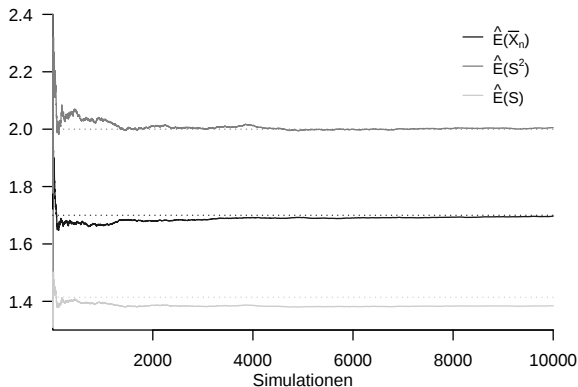
# Simulationsiterationen
for(s in 1:ns){

  # Stichprobenrealisation von  $X_1, \dots, X_{12}$ 
  x      = rnorm(n, mu, sqrt(sigsqr))

  # Erwartungswert-, Varianz-, Standardabweichungsschaetzer
  X_bar[s] = mean(x)          # Stichprobenmittel
  S_sqr[s] = var(x)           # Stichprobenvarianz
  S[s]     = sd(x)            # Stichprobenstandardabweichung
}

# Erwartungswertschaetzung
E_hat_X_bar = cumsum(X_bar)/(1:ns) #  $\mathbb{E}(\bar{X}_n)$  Schaetzungen
E_hat_S_sqr = cumsum(S_sqr)/(1:ns) #  $\mathbb{E}(S^2)$  Schaetzungen
E_hat_S     = cumsum(S)/(1:ns)    #  $\mathbb{E}(S)$  Schaetzungen
```

Simulation ($X_1, \dots, X_{12} \sim N(\mu, \sigma^2)$ mit $n = 12$, $\mu = 1.7$, $\sigma^2 = 2$, $\sigma \approx 1.41$)



Schätzereigenschaften bei endlichen Stichproben

Ein Schätzer $\hat{\tau}_n$ heißt **erwartungstreu**, wenn sein Erwartungswert dem wahren, aber unbekannten, Wert $\tau(\theta)$ für alle $\theta \in \Theta$ gleicht.

Die **Varianz** eines Schätzers $\hat{\tau}_n$ ist die Varianz der Zufallsvariable $\hat{\tau}_n(X)$; der **Standardfehler** eines Schätzers $\hat{\tau}_n$ ist die Standardabweichung der Zufallsvariable $\hat{\tau}_n(X)$.

Die **Cramér-Rao-Ungleichung** gibt eine untere Schranke für die Varianz erwartungstreuer Schätzer an. Ein erwartungstreuer Schätzer mit Varianz gleich der in der Cramér-Rao-Ungleichung gegebenen unteren Schranke hat die kleinstmögliche Varianz aller erwartungstreuen Schätzer und ist in diesem Sinne ein “optimaler” Schätzer.

Der **mittlere quadratische Fehler** von $\hat{\tau}_n$ ist der Erwartungswert der quadrierten Abweichung von $\hat{\tau}_n(X)$ von $\tau(\theta)$ über Stichproben vom Umfang n .

Definition (Varianz und Standardfehler)

$X_1, \dots, X_n \sim p_\theta$ sei eine Stichprobe und $\hat{\tau}_n$ sei ein Schätzer von τ .

- Die *Varianz* von $\hat{\tau}_n$ ist definiert als

$$\mathbb{V}_\theta(\hat{\tau}_n) := \mathbb{E}_\theta \left((\hat{\tau}_n(X) - \mathbb{E}_\theta(\hat{\tau}_n(X)))^2 \right). \quad (10)$$

- Der *Standardfehler* von $\hat{\tau}_n$ ist definiert als

$$\text{SE}(\hat{\tau}_n) := \sqrt{\mathbb{V}_\theta(\hat{\tau}_n)} \quad (11)$$

Bemerkungen

- Die Varianz eines Schätzers $\hat{\tau}_n$ ist die Varianz der Zufallsvariable $\hat{\tau}_n(X)$.
- Der Standardfehler eines Schätzers $\hat{\tau}_n$ ist die Standardabweichung von $\hat{\tau}_n(X)$.

Theorem (Standardfehler des Stichprobenmittels)

$X_1, \dots, X_n \sim p_\theta$ sei die Stichprobe eines parametrischen statistischen Produktmodells $\mathcal{N}$. Dann ist der *Standardfehler des Stichprobenmittels* gegeben durch

$$\text{SE}(\bar{X}_n) = \frac{\mathbb{S}_\theta(X_0)}{\sqrt{n}}. \quad (12)$$

Der Standardfehler des Stichprobenmittels heißt auch *Standardfehler des Mittelwertes*.

Beweis

Per definitionem und mit $\mathbb{V}_\theta(\bar{X}_n) = \mathbb{V}_\theta(X_0)/n$, ergibt sich

$$\text{SE}(\bar{X}_n) = \sqrt{\mathbb{V}_\theta(\bar{X}_n)} = \sqrt{\frac{\mathbb{V}_\theta(X_0)}{n}} = \frac{\mathbb{S}_\theta(X_0)}{\sqrt{n}}. \quad (13)$$

Bemerkung

- Der Standardfehler des Mittelwerts beschreibt die Variabilität des Stichprobenmittels.
- Da $\mathbb{S}_\theta(X_0)$ unbekannt ist, ist auch $\text{SE}(\bar{X}_n)$ unbekannt.
- Ein verzerrter Schätzer für den Standardfehler des Stichprobenmittels ist gegeben durch

$$\hat{\text{SE}}(\bar{X}_n) = \frac{S_n}{\sqrt{n}}. \quad (14)$$

Beispiel (Standardfehler des Bernoulli Parameter ML Schätzers)

Es sei $X_1, \dots, X_n \sim \text{Bern}(\mu)$ und $\hat{\mu}_n^{\text{ML}}$ der ML Schätzer für μ . Dann ist

$$\text{SE}(\hat{\mu}_n^{\text{ML}}) = \sqrt{\frac{\mu(1-\mu)}{n}}. \quad (15)$$

Beweis

Es gilt

$$\text{SE}(\hat{\mu}_n^{\text{ML}}) = \sqrt{\mathbb{V}_{\mu}(\hat{\mu}_n^{\text{ML}})} = \sqrt{\mathbb{V}_{\mu}\left(\frac{1}{n} \sum_{i=1}^n X_i\right)} = \sqrt{\frac{1}{n^2} \sum_{i=1}^n \mathbb{V}_{\mu}(X_i)} = \sqrt{\frac{n\mu(1-\mu)}{n^2}} = \sqrt{\frac{\mu(1-\mu)}{n}}, \quad (16)$$

wobei die dritte Gleichung mit der Unabhängigkeit der $X_i, i = 1, \dots, n$ und die vierte Gleichung mit der Varianz $\mathbb{V}_{\mu}(X_0) = \mathbb{V}_{\mu}(X_i) = \mu(1-\mu), i = 1, \dots, n$ der Bernoulli Stichprobenvariablen folgt.

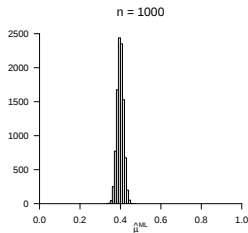
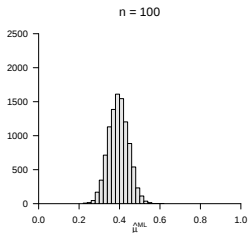
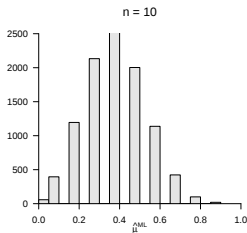
□

Bemerkung

- Ein Schätzer für den Standardfehler $\text{SE}(\hat{\mu}_n^{\text{ML}})$ ist

$$\hat{\text{SE}}(\hat{\mu}_n^{\text{ML}}) = \sqrt{\frac{\hat{\mu}_n^{\text{ML}}(1 - \hat{\mu}_n^{\text{ML}})}{n}}$$

Simulation ($X_1, \dots, X_n \sim \text{Bern}(\mu)$ mit $\mu = 0.4$)



- Die Varianz bzw. der Standardfehler von $\hat{\mu}^{\text{ML}}$ hängen von n ab.

Schätzereigenschaften bei endlichen Stichproben

Ein Schätzer $\hat{\tau}_n$ heißt **erwartungstreu**, wenn sein Erwartungswert dem wahren, aber unbekannten, Wert $\tau(\theta)$ für alle $\theta \in \Theta$ gleicht.

Die **Varianz** eines Schätzers $\hat{\tau}_n$ ist die Varianz der Zufallsvariable $\hat{\tau}_n(X)$; der **Standardfehler** eines Schätzers $\hat{\tau}_n$ ist die Standardabweichung der Zufallsvariable $\hat{\tau}_n(X)$.

Die **Cramér-Rao-Ungleichung** gibt eine untere Schranke für die Varianz erwartungstreuer Schätzer an. Ein erwartungstreuer Schätzer mit Varianz gleich der in der Cramér-Rao-Ungleichung gegebenen unteren Schranke hat die kleinstmögliche Varianz aller erwartungstreuen Schätzer und ist in diesem Sinne ein “optimaler” Schätzer.

Der **mittlere quadratische Fehler** von $\hat{\tau}_n$ ist der Erwartungswert der quadrierten Abweichung von $\hat{\tau}_n(X)$ von $\tau(\theta)$ über Stichproben vom Umfang n .

Vorbemerkungen zur Cramér-Rao-Ungleichung

- Je kleiner die Varianz eines Schätzers, desto besser. Weil aber Stichproben streuen, kann die Varianz von erwartungstreuen Schätzern nicht beliebig klein sein.
- Die **Cramér-Rao-Ungleichung** gibt eine untere Schranke für die Varianz erwartungstreuer Schätzer an. Ein erwartungstreuer Schätzer mit Varianz gleich der in der Cramér-Rao-Ungleichung gegebenen unteren Schranke hat die kleinstmögliche Varianz aller erwartungstreuer Schätzer und ist in diesem Sinne “optimal”.
- Die Cramér-Rao-Ungleichung basiert auf dem Begriff der **Fisher-Information**. Wir diskutieren deshalb zunächst die Begriffe der **Scorefunktion** und der darauf basierenden **Fisher-Information**
- Die vorgestellten Resultate gelten im Allgemeinen nur unter eine Reihe von Annahmen, den sogenannten **Fisher-Regularitätsbedingungen**

Fisher-Regularitätsbedingungen

1. Θ ist ein offenes Intervall, d.h. θ liegt nicht an einer Parameterraumgrenze.
2. Der Träger von p_θ hängt nicht von θ ab.
3. WMFs oder WDF mit unterschiedlichem $\theta \in \Theta$ sind unterschiedlich.
4. Die Likelihood-Funktion ist zweimal stetig differenzierbar.
5. Integration und Differentiation dürfen vertauscht werden.

Definition (Scorefunktion und Fisher-Information)

$X_1, \dots, X_n \sim p_\theta$ sei die Stichprobe eines parametrischen statistischen Produktmodells $\mathcal{M}$ mit eindimensionalem Parameter θ und ℓ_n sei die zugehörige Log-Likelihood-Funktion.

- Die erste Ableitung der Log-Likelihood-Funktion ℓ_n wird *Scorefunktion der Stichprobe* $X_1, \dots, X_n$ genannt und mit

$$S_n(\theta) := \frac{d}{d\theta} \ell_n(\theta). \quad (17)$$

bezeichnet. Für $n = 1$ schreiben wir $S(\theta) := S_1(\theta)$ und nennen $S(\theta)$ *Scorefunktion einer Zufallsvariable*.

- Die negative zweite Ableitung der Log-Likelihood-Funktion ℓ_n wird *Fisher-Information der Stichprobe* $X_1, \dots, X_n$ genannt und mit

$$I_n(\theta) := -\frac{d^2}{d\theta^2} \ell_n(\theta). \quad (18)$$

bezeichnet. Für $n = 1$ schreiben wir $I(\theta) := I_1(\theta)$ und nennen $I(\theta)$ die *Fisher-Information einer Zufallsvariable*.

Definition (Erwartete und beobachtete Fisher-Information)

$X_1, \dots, X_n \sim p_\theta$ sei die Stichprobe eines parametrischen statistischen Produktmodells $\mathcal{M}$ mit eindimensionalem Parameter θ , ℓ_n sei die zugehörige Log-Likelihood-Funktion und $\hat{\theta}_n^{\text{ML}}(X_1, \dots, X_n)$ sei ein ML-Schätzer von θ .

- Die *beobachtete Fisher-Information der Stichprobe* $X_1, \dots, X_n$ ist definiert als

$$I_n \left(\hat{\theta}_n^{\text{ML}} \right) := - \frac{d^2}{d\theta^2} \ell_n \left(\hat{\theta}_n^{\text{ML}} \right), \quad (19)$$

d.h. die beobachtete Fisher-Information der Stichprobe $X_1, \dots, X_n$ ist die Fisher-Information an der Stelle des ML-Schätzers $\hat{\theta}_n^{\text{ML}}(X_1, \dots, X_n)$.

- Die *erwartete Fisher-Information der Stichprobe* $X_1, \dots, X_n$ ist definiert als

$$J_n(\theta) := \mathbb{E}_\theta(I_n(\theta)). \quad (20)$$

Für $n = 1$ schreiben wir $J(\theta) := J_1(\theta)$ und nennen $J(\theta)$ die *erwartete Fisher-Information einer Zufallsvariable*.

Theorem (Additivität der Fisher-Information)

$X_1, \dots, X_n \sim p_\theta$ sei die Stichprobe eines parametrischen statistischen Produktmodells $\mathcal{M}$ mit eindimensionalem Parameter θ , ℓ_n sei die zugehörige Log-Likelihood-Funktion, und $I_n(\theta)$ und $J_n(\theta)$ seien die Fisher-Information und die erwartete Fisher-Information der Stichprobe $X_1, \dots, X_n$, respektive. Dann gilt

$$I_n(\theta) = nI_1(\theta) \text{ und } J_n(\theta) = nJ_1(\theta). \quad (21)$$

Bemerkungen

- Um $I_n(\theta)$ oder $J_n(\theta)$ zu berechnen, genügt es also $I(\theta)$ oder $J(\theta)$ zu berechnen.
- Die Additivität der beobachteten Fisher-Information ist in der Additivität von $I_n(\theta)$ implizit.

Beweis

Wir zeigen das Resultat für die erwartete Fisher-Information, das Resultat für die Fisher-Information ist dann implizit. Per definitionem und mit der Linearität von Ableitungen und Erwartungswerte gilt

$$\begin{aligned} J_n(\theta) &= \mathbb{E}_\theta \left(-\frac{d^2}{d\theta^2} \ell_n(\theta) \right) \\ &= \mathbb{E}_\theta \left(-\frac{d^2}{d\theta^2} \ln \left(\prod_{i=1}^n p_\theta(x_i) \right) \right) \\ &= \mathbb{E}_\theta \left(-\frac{d^2}{d\theta^2} \sum_{i=1}^n \ln p_\theta(x_i) \right) \\ &= \mathbb{E}_\theta \left(-\frac{d^2}{d\theta^2} \sum_{i=1}^n \ln p_\theta(x_1) \right) \tag{22} \\ &= \mathbb{E}_\theta \left(-\frac{d^2}{d\theta^2} \ell_1(\theta) n \right) \\ &= n \mathbb{E}_\theta \left(-\frac{d^2}{d\theta^2} \ell_1(\theta) \right) \\ &= n J_n(\theta). \end{aligned}$$

□

Theorem (Erwartungswert und Varianz der Scorefunktion)

Der Erwartungswert der Scorefunktion einer Zufallsvariable ist

$$\mathbb{E}_{\theta}(S(\theta)) = 0 \quad (23)$$

und die Varianz der Scorefunktion einer Zufallsvariable ist

$$\mathbb{V}_{\theta}(S(\theta)) = J(\theta). \quad (24)$$

Bemerkungen

- Der Erwartungswert der Ableitung der Log-Likelihood-Funktion ist Null.
- Die erwartete Fisher-Information ist gleich der Varianz der Scorefunktion.

Beweis

Wir betrachten nur den Fall, dass p_θ eine WDF ist und zeigen zunächst, dass $\mathbb{E}_\theta(S(\theta)) = 0$ ist:

$$\begin{aligned}\mathbb{E}_\theta(S(\theta)) &= \int S(\theta) p_\theta(x) dx \\&= \int \frac{d}{d\theta} \ell(\theta) p_\theta(x) dx \\&= \int \frac{d}{d\theta} \ln L(\theta) p_\theta(x) dx \\&= \int \frac{1}{L(\theta)} \frac{d}{d\theta} L(\theta) p_\theta(x) dx \\&= \int \frac{1}{p_\theta(x)} \frac{d}{d\theta} L(\theta) p_\theta(x) dx & (25) \\&= \int \frac{d}{d\theta} L(\theta) dx \\&= \frac{d}{d\theta} \int p_\theta(x) dx \\&= \frac{d}{d\theta} 1 \\&= 0.\end{aligned}$$

Mit der Definition der Varianz folgt dann sofort, dass $\mathbb{V}_\theta(S(\theta)) = \mathbb{E}_\theta(S(\theta)^2)$ ist.

Beweis

Als nächstes zeigen wir, dass $J(\theta) = \mathbb{E}_\theta (S(\theta)^2)$ und deshalb $\mathbb{V}_\theta(S(\theta)) = J(\theta)$ ist:

$$\begin{aligned} J(\theta) &= \mathbb{E}_\theta \left(-\frac{d^2}{d\theta^2} \ln L(\theta) \right) \\ &= \mathbb{E}_\theta \left(-\frac{d}{d\theta} \frac{\frac{d}{d\theta} L(\theta)}{L(\theta)} \right) \\ &= \mathbb{E}_\theta \left(-\frac{\frac{d^2}{d\theta^2} L(\theta) L(\theta) - \frac{d}{d\theta} L(\theta) \frac{d}{d\theta} L(\theta)}{L(\theta) L(\theta)} \right) \\ &= -\mathbb{E}_\theta \left(\frac{\frac{d^2}{d\theta^2} L(\theta)}{L(\theta)} \right) + \mathbb{E}_\theta \left(\frac{\left(\frac{d}{d\theta} L(\theta) \right)^2}{(L(\theta))^2} \right) \\ &= -\int \frac{\frac{d^2}{d\theta^2} L(\theta)}{L(\theta)} p_\theta(x) dx + \int \frac{\left(\frac{d}{d\theta} L(\theta) \right)^2}{(L(\theta))^2} p_\theta(x) dx \\ &= -\frac{d^2}{d\theta^2} \int p_\theta(x) dx + \int \left(\frac{1}{L(\theta)} \frac{d}{d\theta} L(\theta) \right)^2 p_\theta(x) dx \\ &= -\frac{d^2}{d\theta^2} 1 + \int \left(\frac{d}{d\theta} \ln L(\theta) \right)^2 p_\theta(x) dx \\ &= \mathbb{E}_\theta (S(\theta)^2) . \end{aligned} \tag{26}$$

□

Beispiel (Scorefunktion und Fisher-Information für den Bernoulli Parameter)

Es sei $X_1, \dots, X_n \sim \text{Bern}(\mu)$ mit $\mu \in]0, 1[$. Dann gilt:

- Die Scorefunktion der Stichprobe ist

$$S_n :]0, 1[\rightarrow \mathbb{R}, \mu \mapsto S_n(\mu) := \frac{1}{\mu} \sum_{i=1}^n x_i - \frac{1}{1-\mu} \left(n - \sum_{i=1}^n x_i \right). \quad (27)$$

- Die Fisher-Information der Stichprobe ist

$$I_n :]0, 1[\rightarrow \mathbb{R}, \mu \mapsto I_n(\mu) := \frac{x}{\mu^2} + \frac{(1-x)^2}{1-\mu}. \quad (28)$$

- Die beobachtete Fisher-Information der Stichprobe ist

$$I_n :]0, 1[\rightarrow \mathbb{R}, \hat{\mu}_n^{\text{ML}} \mapsto I_n(\hat{\mu}_n^{\text{ML}}) := \frac{x}{\hat{\mu}_n^{\text{ML}^2}} + \frac{(1-x)}{1-\hat{\mu}_n^{\text{ML}}}. \quad (29)$$

- Die erwartete Fisher-Information der Stichprobe ist

$$J_n :]0, 1[\rightarrow \mathbb{R}, \mu \mapsto J_n(\mu) := \frac{n}{\mu(1-\mu)}. \quad (30)$$

Beispiel (Scorefunktion und Fisher-Information für den Bernoulli Parameter)

Beweis

Die Scorefunktion wurde bereits in Vorlesung (7) im Kontext der Maximum-Likelihood-Schätzung von μ hergeleitet. Wir betrachten die Fisher-Information einer einzelnen Bernoulli Zufallsvariable X :

$$\begin{aligned} I(\mu) &:= -\frac{d^2}{d\mu^2} \ell_1(\mu) \\ &= -\frac{d^2}{d\mu^2} \ln p_\mu(x) \\ &= -\frac{d^2}{d\mu^2} (x \ln \mu + (1-x) \ln(1-\mu)) \\ &= -\frac{d}{d\mu} \left(\frac{d}{d\mu} (x \ln \mu + (1-x) \ln(1-\mu)) \right) \\ &= -\frac{d}{d\mu} \left(\frac{x}{\mu} + \frac{(1-x)}{1-\mu} \right) \\ &= -\left(-\frac{x}{\mu^2} - \frac{(1-x)^2}{(1-\mu)^2} \right) \\ &= \frac{x}{\mu^2} + \frac{(1-x)^2}{(1-\mu)^2}. \end{aligned} \tag{31}$$

Beispiel (Scorefunktion und Fisher-Information für den Bernoulli Parameter)

Beweis

Damit ergibt sich die erwartete Fisher-Information der Zufallsvariable X als

$$\begin{aligned} J(\mu) &= \mathbb{E}_{\mu}(I(\mu)) \\ &= \mathbb{E}_{\mu} \left(\frac{X}{\mu^2} + \frac{(1-X)^2}{1-\mu} \right) \\ &= \frac{\mathbb{E}_{\mu}(X)}{\mu^2} + \frac{(1 - \mathbb{E}_{\mu}(X))^2}{1-\mu} \\ &= \frac{\mu}{\mu^2} + \frac{(1-\mu)^2}{1-\mu} \\ &= \frac{1}{\mu(1-\mu)}. \end{aligned} \tag{32}$$

Mit der Additivitätseigenschaft der Fisher-Information und der Definition der beobachteten Fisher-Information ergibt sich dann sofort

$$I_n(\mu) = \frac{nx}{\mu^2} + \frac{n(1-x)^2}{1-\mu}, \text{ und } J_n(\mu) = \frac{n}{\mu(1-\mu)}. \tag{33}$$

□

Beispiel (Erwartungswert-Parameter der Normalverteilung)

Es sei $X_1, \dots, X_n \sim N(\mu, \sigma^2)$ und σ^2 als bekannt vorausgesetzt. Dann gilt:

- Die Scorefunktion der Stichprobe ist

$$S_n : \mathbb{R} \rightarrow \mathbb{R}, \mu \mapsto S_n(\mu) := \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu). \quad (34)$$

- Die Fisher-Information der Stichprobe ist

$$I_n : \mathbb{R} \rightarrow \mathbb{R}, \mu \mapsto I_n(\mu) := \frac{n}{\sigma^2}. \quad (35)$$

- Die beobachtete Fisher-Information der Stichprobe ist

$$I_n(\hat{\mu}_n^{\text{ML}}) = \frac{n}{\sigma^2}. \quad (36)$$

- Die erwartete Fisher-Information der Stichprobe ist

$$J_n : \mathbb{R} \rightarrow \mathbb{R}, \mu \mapsto J_n(\mu) := \frac{n}{\sigma^2}. \quad (37)$$

Beispiel (Erwartungswert-Parameter der Normalverteilung)

Beweis

Wir erinnern uns, dass die Log-Likelihood-Funktion der Stichprobe $X_1, \dots, X_n \sim N(\mu, \sigma^2)$ bei bekanntem Varianzparameter σ^2 durch

$$\ell_n : \mathbb{R} \rightarrow \mathbb{R}, \mu \mapsto \ell_n(\mu) := -\frac{n}{2} \ln 2\pi - \frac{n}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2. \quad (38)$$

gegeben ist. Damit ergibt sich die Scorefunktion als

$$S_n(\mu) = \frac{d}{d\mu} \ell_n(\mu) = \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu) \quad (39)$$

Die Fisher-Information der Stichprobe $X_1, \dots, X_n$ ergibt sich als

$$I_n(\mu) = -\frac{d^2}{d\mu^2} \ell_n(\mu) = -\frac{d}{d\mu} S_n(\mu) = -\frac{1}{\sigma^2} \frac{d}{d\mu} \left(\sum_{i=1}^n x_i - n\mu \right) = \frac{n}{\sigma^2}. \quad (40)$$

Die beobachtete Fisher-Information ist die Fisher-Information an der Stelle des Maximum-Likelihood-Schätzers $\hat{\mu}_n^{\text{ML}}$. Die erwartete Fisher-Information schließlich ergibt sich als

$$J_n(\mu) = \mathbb{E}_\mu(I_n(\mu)) = \mathbb{E}_\mu \left(\frac{n}{\sigma^2} \right) = \frac{n}{\sigma^2}. \quad (41)$$

□

Beispiel (Varianz-Parameter der Normalverteilung)

Es sei $X_1, \dots, X_n \sim N(\mu, \sigma^2)$ und μ als bekannt vorausgesetzt. Dann gilt:

- die Scorefunktion ist

$$S_n : \mathbb{R}_{>0} \rightarrow \mathbb{R}, \sigma^2 \mapsto S_n(\sigma^2) := -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (x_i - \mu)^2 \quad (42)$$

- die Fisher-Information der Stichprobe ist

$$I_n : \mathbb{R}_{>0} \rightarrow \mathbb{R}, \sigma^2 \mapsto I_n(\sigma^2) := \frac{1}{\sigma^6} \sum_{i=1}^n (x_i - \mu)^2 - \frac{n}{2\sigma^4} \quad (43)$$

- die beobachtete Fisher-Information der Stichprobe ist

$$I_n(\hat{\sigma}_n^{2\text{ML}}) = \frac{n}{2\hat{\sigma}_{\text{ML}}^4} \quad (44)$$

- die erwartete Fisher-Information der Stichprobe ist

$$J_n : \mathbb{R}_{>0} \rightarrow \mathbb{R}, \sigma^2 \mapsto J_n(\sigma^2) := \frac{n}{2\sigma^4}. \quad (45)$$

Beispiel (Varianz-Parameter der Normalverteilung)

Beweis

Wir erinnern uns, dass die Log-Likelihood-Funktion der Stichprobe $X_1, \dots, X_n \sim N(\mu, \sigma^2)$ bei bekanntem Erwartungswert-Parameter μ durch

$$\ell_n : \mathbb{R}_{>0} \rightarrow \mathbb{R}, \sigma^2 \mapsto \ell_n(\sigma^2) := -\frac{n}{2} \ln 2\pi - \frac{n}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2. \quad (46)$$

gegeben ist. Die Scorefunktion ergibt sich also als

$$S_n(\sigma^2) = \frac{d}{d\sigma^2} \ell_n(\sigma^2) = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (x_i - \mu)^2. \quad (47)$$

Die Fisher-Information der Stichprobe $X_1, \dots, X_n$ ergibt sich als

$$\begin{aligned} I_n(\sigma^2) &= -\frac{d}{d\sigma^2} S_n(\sigma^2) = -\left(\frac{n}{2\sigma^4} - \frac{1}{\sigma^6} \sum_{i=1}^n (x_i - \mu)^2 \right) \\ &= \frac{1}{\sigma^6} \sum_{i=1}^n (x_i - \mu)^2 - \frac{n}{2\sigma^4}. \end{aligned} \quad (48)$$

Beispiel (Varianz-Parameter der Normalverteilung)

Beweis

Die beobachtete Fisher-Information ist die Fisher-Information an der Stelle des Maximum-Likelihood-Schätzers $\hat{\sigma}_n^{2\text{ML}}$.

$$\begin{aligned} I_n(\hat{\sigma}_n^{2\text{ML}}) &= \frac{\sum_{i=1}^n (x_i - \mu)^2}{(\hat{\sigma}_n^{2\text{ML}})^3} - \frac{n}{2 (\hat{\sigma}_n^{2\text{ML}})^2} \\ &= \frac{\sum_{i=1}^n (x_i - \mu)^2}{\frac{1}{n^3} (\sum_{i=1}^n (x_i - \mu)^2)^3} - \frac{n}{2 (\hat{\sigma}_n^{2\text{ML}})^2} \\ &= \frac{1}{\frac{1}{n^3} (\sum_{i=1}^n (x_i - \mu)^2)^2} - \frac{n}{2 (\hat{\sigma}_n^{2\text{ML}})^2} \\ &= \frac{n}{(\hat{\sigma}_n^{2\text{ML}})^2} - \frac{n}{2 (\hat{\sigma}_n^{2\text{ML}})^2} \\ &= \frac{n}{2 (\hat{\sigma}_n^{2\text{ML}})^2} \\ &= \frac{n}{2 \hat{\sigma}_n^4 \text{ML}}. \end{aligned} \tag{49}$$

Beispiel (Varianz-Parameter der Normalverteilung)

Beweis

Die erwartete Fisher-Information ergibt sich als

$$\begin{aligned} J_n(\sigma^2) &= \mathbb{E}_{\sigma^2}(I_n(\sigma^2)) \\ &= \mathbb{E}_{\sigma^2} \left(\frac{1}{\sigma^6} \sum_{i=1}^n (x_i - \mu)^2 - \frac{n}{2\sigma^4} \right) \\ &= \frac{1}{\sigma^6} \sum_{i=1}^n \mathbb{E}_{\sigma^2} \left((x_i - \mu)^2 \right) - \frac{n}{2\sigma^4} \\ &= \frac{1}{\sigma^6} \sum_{i=1}^n \sigma^2 - \frac{n}{2\sigma^4} \\ &= \frac{n\sigma^2}{\sigma^6} - \frac{n}{2\sigma^4} \\ &= \frac{n}{\sigma^4} - \frac{n}{2\sigma^4} \\ &= \frac{n}{2\sigma^4}. \end{aligned} \tag{50}$$

Theorem (Cramér-Rao-Ungleichung)

$\mathcal{M}$ sei ein parametrisches statistisches Modell mit WMF oder WDF p_θ und $\hat{\tau}_n$ sei ein erwartungstreuer Schätzer von $\tau(\theta)$. Dann gilt

$$\mathbb{V}_\theta(\hat{\tau}_n) \geq \frac{\left(\frac{d}{d\theta}\tau(\theta)\right)^2}{J(\theta)}. \quad (51)$$

Im Speziellen gilt für $\tau(\theta) := \theta$ und somit $\hat{\tau}_n = \hat{\theta}_n$ und $\left(\frac{d}{d\theta}\tau(\theta)\right)^2 = 1$, dass

$$\mathbb{V}_\theta(\hat{\theta}_n) \geq \frac{1}{J(\theta)}. \quad (52)$$

Die rechte Seite obiger Ungleichungen heißt *Cramér-Rao-Schranke*.

Bemerkungen

- Die Varianz eines erwartungstreuen Schätzers $\hat{\theta}$ von θ ist größer oder gleich der reziproken erwarteten Fisher-Information $J(\theta)$.
- Wenn $\mathbb{V}_\theta(\hat{\theta}_n) = \frac{1}{J(\theta)}$ ist, ist die Varianz des Schätzers minimal.

Beweis

Wir halten zunächst fest, dass für die Zufallsvariablen $S(\theta)$ und $\hat{\tau}_n$ mit der Korrelationsungleichung und $\mathbb{V}_\theta(S(\theta)) = J(\theta)$ gilt, dass

$$\begin{aligned}\frac{\mathbb{C}_\theta(S_n(\theta), \hat{\tau}_n)^2}{\mathbb{V}_\theta(S(\theta))\mathbb{V}_\theta(\hat{\tau}_n)} &\leq 1 \\ \Leftrightarrow \mathbb{V}_\theta(\hat{\tau}_n) &\geq \frac{\mathbb{C}_\theta(S(\theta), \hat{\tau}_n)^2}{J(\theta)}.\end{aligned}\tag{53}$$

Mit dem Translationstheorem für Kovarianzen, $\mathbb{E}_\theta(S(\theta)) = 0$, und der Erwartungstreue von $\hat{\tau}_n$ ergibt sich dann

$$\mathbb{C}_\theta(S(\theta), \hat{\tau}_n) = \frac{d}{d\theta} \tau(\theta)\tag{54}$$

wie unten gezeigt wird. Also gilt

$$\mathbb{V}_\theta(\hat{\tau}_n) \geq \frac{\left(\frac{d}{d\theta} \tau(\theta)\right)^2}{J(\theta)}.\tag{55}$$

Beweis

Es bleibt zu zeigen, dass $\mathbb{C}_\theta(S(\theta), \hat{\tau}_n) = \frac{d}{d\theta} \tau(\theta)$. Dies ergibt sich mit

$$\begin{aligned}\mathbb{C}_\theta(S(\theta), \hat{\tau}_n) &= \mathbb{E}_\theta(S(\theta) \hat{\tau}_n) - \mathbb{E}_\theta(S(\theta)) \mathbb{E}_\theta(\hat{\tau}_n) \\&= \mathbb{E}_\theta(S(\theta) \hat{\tau}_n) \\&= \int S(\theta) \hat{\tau}_n p_\theta(x) dx \\&= \int \frac{d}{d\theta} \ln L(\theta) \hat{\tau}_n p_\theta(x) dx \\&= \int \frac{\frac{d}{d\theta} L(\theta)}{L(\theta)} \hat{\tau}_n p_\theta(x) dx \\&= \int \frac{\frac{d}{d\theta} L(\theta)}{p_\theta(x)} \hat{\tau}_n p_\theta(x) dx \\&= \int \frac{d}{d\theta} L(\theta) \hat{\tau}_n dx \\&= \frac{d}{d\theta} \int L(\theta) \hat{\tau}_n dx \\&= \frac{d}{d\theta} \int \hat{\tau}_n p_\theta(x) dx \\&= \frac{d}{d\theta} \mathbb{E}_\theta(\hat{\tau}_n) \\&= \frac{d}{d\theta} \tau(\theta)\end{aligned}\tag{56}$$

□

Schätzereigenschaften bei endlichen Stichproben

Ein Schätzer $\hat{\tau}_n$ heißt **erwartungstreu**, wenn sein Erwartungswert dem wahren, aber unbekannten, Wert $\tau(\theta)$ für alle $\theta \in \Theta$ gleicht.

Die **Varianz** eines Schätzers $\hat{\tau}_n$ ist die Varianz der Zufallsvariable $\hat{\tau}_n(X)$; der **Standardfehler** eines Schätzers $\hat{\tau}_n$ ist die Standardabweichung der Zufallsvariable $\hat{\tau}_n(X)$.

Die **Cramér-Rao-Ungleichung** gibt eine untere Schranke für die Varianz erwartungstreuer Schätzer an. Ein erwartungstreuer Schätzer mit Varianz gleich der in der Cramér-Rao-Ungleichung gegebenen unteren Schranke hat die kleinstmögliche Varianz aller erwartungstreuen Schätzer und ist in diesem Sinne ein “optimaler” Schätzer.

Der **mittlere quadratische Fehler** von $\hat{\tau}_n$ ist der Erwartungswert der quadrierten Abweichung von $\hat{\tau}_n(X)$ von $\tau(\theta)$ über Stichproben vom Umfang n .

Definition (Mittlerer quadratischer Fehler)

$X_1, \dots, X_n \sim p_\theta$ sei die Stichprobe eines parametrischen statistischen Produktmodells $\mathcal{M}$ und $\hat{\tau}_n$ ein Schätzer für τ . Dann ist der *mittlere quadratische Fehler* (engl. *mean squared error*) von $\hat{\tau}_n$ definiert als

$$\text{MQF}(\hat{\tau}_n) := \mathbb{E}_\theta \left((\hat{\tau}_n(X) - \tau(\theta))^2 \right). \quad (57)$$

Bemerkungen

- Der MQF von $\hat{\tau}_n$ ist die erwartete quadrierte Abweichung von $\hat{\tau}_n(X)$ von $\tau(\theta)$.
- Die Varianz von $\hat{\tau}_n$ ist die erwartete quadrierte Abweichung von $\hat{\tau}_n$ von $\mathbb{E}_\theta(\hat{\tau}_n(X))$.
- $\mathbb{E}_\theta(\hat{\tau}_n(X))$ kann mit $\tau(\theta)$ übereinstimmen, muss es aber nicht.

Theorem (Zerlegung des mittleren quadratischen Fehlers)

$X_1, \dots, X_n \sim p_\theta$ sei die Stichprobe eines parametrischen statistischen Produktmodells $\mathcal{M}$, $\hat{\tau}_n$ sei ein Schätzer für τ , und $\text{MQF}(\hat{\tau}_n)$ sei der mittlere quadratische Fehler von $\hat{\tau}_n$. Dann gilt

$$\text{MQF}(\hat{\tau}_n) = \text{B}(\hat{\tau}_n)^2 + \mathbb{V}_\theta(\hat{\tau}_n) \quad (58)$$

Bemerkungen

- $\text{MQF} = \text{Bias}^2 + \text{Varianz}$.
- Der MQF kann als Bias-Variance Abwägungskriterium benutzt werden.
- Kleine Verzerrungen können gegenüber großen Varianz präferiert werden.

Mittlerer quadratischer Fehler

Beweis

Zur Vereinfachung der Notation seien $\tau := \tau(\theta)$, $\hat{\tau}_n := \hat{\tau}_n(X)$ und $\bar{\tau}_n := \mathbb{E}_\theta(\hat{\tau}_n(X))$ zur Vereinfachung der Notation. Dann gilt:

$$\begin{aligned}\mathbb{E}_\theta \left((\hat{\tau}_n - \tau)^2 \right) &= \mathbb{E}_\theta \left((\hat{\tau}_n - \bar{\tau}_n + \bar{\tau}_n - \tau)^2 \right) \\&= \mathbb{E}_\theta \left((\hat{\tau}_n - \bar{\tau}_n)^2 + 2(\hat{\tau}_n - \bar{\tau}_n)(\bar{\tau}_n - \tau) + (\bar{\tau}_n - \tau)^2 \right) \\&= \mathbb{E}_\theta \left((\hat{\tau}_n - \bar{\tau}_n)^2 \right) + 2\mathbb{E}_\theta \left((\hat{\tau}_n - \bar{\tau}_n)(\bar{\tau}_n - \tau) \right) + \mathbb{E}_\theta \left((\bar{\tau}_n - \tau)^2 \right) \\&= \mathbb{E}_\theta \left((\hat{\tau}_n - \bar{\tau}_n)^2 \right) + 2\mathbb{E}_\theta \left(\hat{\tau}_n \bar{\tau}_n - \hat{\tau}_n \tau - \bar{\tau}_n \bar{\tau}_n + \bar{\tau}_n \tau \right) + \mathbb{E}_\theta \left((\bar{\tau}_n - \tau)^2 \right) \\&= \mathbb{E}_\theta \left((\hat{\tau}_n - \bar{\tau}_n)^2 \right) + 2 \left(\bar{\tau}_n \bar{\tau}_n - \bar{\tau}_n \tau \right) + \mathbb{E}_\theta \left((\bar{\tau}_n - \tau)^2 \right) \\&= \mathbb{E}_\theta \left((\hat{\tau}_n - \bar{\tau}_n)^2 \right) + 0 + \mathbb{E}_\theta \left((\bar{\tau}_n - \tau)^2 \right) \\&= \mathbb{E}_\theta \left((\bar{\tau}_n - \tau)^2 \right) + \mathbb{E}_\theta \left((\hat{\tau}_n - \bar{\tau}_n)^2 \right) \\&= \mathbb{E}_\theta \left((\mathbb{E}_\theta(\hat{\tau}_n) - \tau)^2 \right) + \mathbb{E}_\theta \left((\hat{\tau}_n - \mathbb{E}_\theta(\hat{\tau}_n))^2 \right) \\&= (\mathbb{E}_\theta(\hat{\tau}_n) - \tau)^2 + \mathbb{V}_\theta(\hat{\tau}_n) \\&= \mathbf{B}(\hat{\tau}_n)^2 + \mathbb{V}_\theta(\hat{\tau}_n).\end{aligned}$$

□

Schätzereigenschaften

- Vorbemerkungen
- Schätzereigenschaften bei endlichen Stichproben
- **Asymptotische Schätzereigenschaften**
- Eigenschaften von Maximum-Likelihood Schätzern
- Selbstkontrollfragen

Motivation

- Dieser Abschnitt ist eine Kurzeinführung in die *Asymptotische Statistik (AS)*.
- Die AS befasst sich mit dem Verhalten von Statistiken bei großen Stichproben.
- Methoden der AS werden benutzt, um
 - qualitative Schätzereigenschaften zu studieren und
 - Schätzereigenschaften für große Stichprobengrößen zu approximieren.
- Moderne Stichproben sind üblicherweise groß.
- Die Methoden der AS sind also praktisch einsetzbar und gerechtfertigt.
- van der Vaart (1998) gibt eine ausführliche Einführung in die AS.

Asymptotische Schätzereigenschaften

Ein Schätzer $\hat{\tau}_n$ für τ heißt **asymptotisch erwartungstreu**, wenn der Erwartungswert von $\hat{\tau}_n$ für große Stichprobengrößen $n \rightarrow \infty$ gleich dem wahren, aber unbekannten, Wert $\tau(\theta)$ ist.

Ein Schätzer $\hat{\tau}_n$ für τ heißt **konsistent**, wenn für große Stichprobengrößen $n \rightarrow \infty$ die Wahrscheinlichkeit dafür, dass $\hat{\tau}_n(X)$ vom wahren, aber unbekannten, Wert $\tau(\theta)$ **abweicht** beliebig klein wird.

Ein Schätzer $\hat{\tau}_n$ für τ heißt **asymptotisch normalverteilt**, wenn für große Stichprobengrößen $n \rightarrow \infty$, die Verteilung von $\hat{\tau}_n$ durch eine Normalverteilung gegeben ist.

Ein Schätzer $\hat{\tau}_n$ für τ heißt **asymptotisch effizient**, wenn für große Stichprobengrößen $n \rightarrow \infty$ die Verteilung von $\hat{\tau}_n$ durch eine Normalverteilung mit Erwartungswertparameter $\tau(\theta)$ und Varianzparameter gleich der Cramér-Rao-Schranke gegeben ist.

Asymptotische Schätzereigenschaften

Ein Schätzer $\hat{\tau}_n$ für τ heißt **asymptotisch erwartungstreu**, wenn der Erwartungswert von $\hat{\tau}_n$ für große Stichprobengrößen $n \rightarrow \infty$ gleich dem wahren, aber unbekannten, Wert $\tau(\theta)$ ist.

Ein Schätzer $\hat{\tau}_n$ für τ heißt **konsistent**, wenn für große Stichprobengrößen $n \rightarrow \infty$ die Wahrscheinlichkeit dafür, dass $\hat{\tau}_n(X)$ vom wahren, aber unbekannten, Wert $\tau(\theta)$ **abweicht** beliebig klein wird.

Ein Schätzer $\hat{\tau}_n$ für τ heißt **asymptotisch normalverteilt**, wenn für große Stichprobengrößen $n \rightarrow \infty$, die Verteilung von $\hat{\tau}_n$ durch eine Normalverteilung gegeben ist.

Ein Schätzer $\hat{\tau}_n$ für τ heißt **asymptotisch effizient**, wenn für große Stichprobengrößen $n \rightarrow \infty$ die Verteilung von $\hat{\tau}_n$ durch eine Normalverteilung mit Erwartungswertparameter $\tau(\theta)$ und Varianzparameter gleich der Cramér-Rao-Schranke gegeben ist.

Definition (Asymptotische Erwartungstreue)

$X_1, \dots, X_n \sim p_\theta$ sei die Stichprobe eines parametrischen statistischen Produktmodells $\mathcal{M}$ und $\hat{\tau}_n$ sei ein Schätzer für τ . $\hat{\tau}_n$ heißt *asymptotisch erwartungstreu*, wenn

$$\lim_{n \rightarrow \infty} \mathbb{E}_\theta(\hat{\tau}_n(X)) = \tau(\theta) \text{ für alle } \theta \in \Theta. \quad (59)$$

Bemerkungen

- Asymptotisch erwartungstreue Schätzer sind für “unendlich große” Stichproben erwartungstreu.
- Erwartungstreue Schätzer sind immer auch asymptotisch erwartungstreu.

Beispiel (Ein asymptotisch erwartungstreuer Schätzer)

Es sei $\mathcal{M}$ das Normalverteilungsmodell, also $X_1, \dots, X_n \sim N(\mu, \sigma^2)$ und

$$\hat{\sigma}_n^{2\text{ML}} := \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2 \quad (60)$$

sei der ML Schätzer von σ^2 . Mit der Erwartungstreue der Stichprobenvarianz ergibt sich dann

$$\mathbb{E}_{\mu, \sigma^2} \left(\hat{\sigma}_n^{2\text{ML}} \right) = \mathbb{E}_{\mu, \sigma^2} \left(\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2 \right) = \frac{1}{n} \mathbb{E}_{\mu, \sigma^2} \left(\sum_{i=1}^n (X_i - \bar{X}_n)^2 \right) = \frac{n-1}{n} \sigma^2 \neq \sigma^2.$$

$\hat{\sigma}_n^{2\text{ML}}$ ist also ein verzerrter Schätzer von σ^2 . Allerdings gilt $(n-1)/n \rightarrow 1$ für $n \rightarrow \infty$, so dass

$$\lim_{n \rightarrow \infty} \mathbb{E}_{\mu, \sigma^2} \left(\hat{\sigma}_n^{2\text{ML}} \right) = \lim_{n \rightarrow \infty} \frac{n-1}{n} \sigma^2 = \sigma^2 \lim_{n \rightarrow \infty} \frac{n-1}{n} = \sigma^2. \quad (61)$$

Beispiel (Ein asymptotisch erwartungstreuer Schätzer)

Es sei $\mathcal{M}$ das Normalverteilungsmodell, also $X_1, \dots, X_n \sim N(\mu, \sigma^2)$ und

$$\hat{\sigma}_n^{2\text{ML}} := \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2 \quad (60)$$

sei der ML Schätzer von σ^2 . Mit der Erwartungstreue der Stichprobenvarianz ergibt sich dann

$$\mathbb{E}_{\mu, \sigma^2} \left(\hat{\sigma}_n^{2\text{ML}} \right) = \mathbb{E}_{\mu, \sigma^2} \left(\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2 \right) = \frac{1}{n} \mathbb{E}_{\mu, \sigma^2} \left(\sum_{i=1}^n (X_i - \bar{X}_n)^2 \right) = \frac{n-1}{n} \sigma^2 \neq \sigma^2.$$

$\hat{\sigma}_n^{2\text{ML}}$ ist also ein verzerrter Schätzer von σ^2 . Allerdings gilt $(n-1)/n \rightarrow 1$ für $n \rightarrow \infty$, so dass

$$\lim_{n \rightarrow \infty} \mathbb{E}_{\mu, \sigma^2} \left(\hat{\sigma}_n^{2\text{ML}} \right) = \lim_{n \rightarrow \infty} \frac{n-1}{n} \sigma^2 = \sigma^2 \lim_{n \rightarrow \infty} \frac{n-1}{n} = \sigma^2. \quad (61)$$

Im Limit großer Stichprobenumfänge ist der ML Schätzer des Varianzparameters einer Normalverteilung also erwartungstreu.

Simulation $X_1, \dots, X_n \sim N(\mu, \sigma^2)$ mit $\mu = 1, \sigma^2 = 2$

```
# Modellformulierung
mu      = 1                # wahrer, aber unbekannter, Erwartungswertparameter
sigsqr  = 2                # wahrer, aber unbekannter, Varianzparameter
n       = seq(1,100, by = 2) # Stichprobengroessen
ns      = 1e3              # Anzahl Simulation pro Stichprobengroesse
sigsqr_ml = matrix(        # \hat{\sigma^2}^{ML} Array
  rep(NaN, length(n)*ns),
  ncol = length(n))

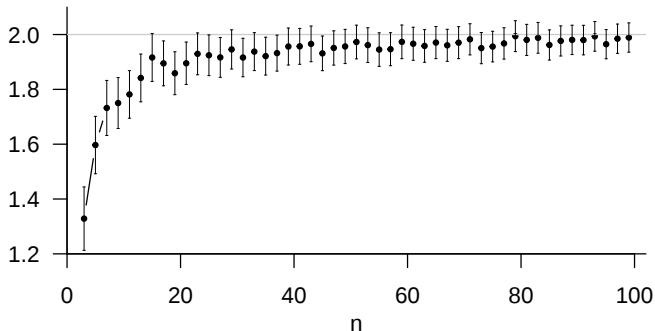
# Stichprobengroesseniterationen
for(i in seq_along(n)){

  # Simulationsiterationen
  for(s in 1:ns){
    # Stichprobenrealisation
    x = rnorm(n[i], mu, sqrt(sigsqr))

    # \hat{\sigma^2}^{ML}
    sigsqr_ml[s,i] = ((n[i]-1)/n[i])*var(x)
  }
}
E_sigsqr_ml = colMeans(sigsqr_ml) # Erwartungswertschaetzung
```

Asymptotische Erwartungstreue

Simulation $X_1, \dots, X_n \sim N(\mu, \sigma^2)$ mit $\mu = 1, \sigma^2 = 2$



Asymptotische Schätzereigenschaften

Ein Schätzer $\hat{\tau}_n$ für τ heißt **asymptotisch erwartungstreu**, wenn der Erwartungswert von $\hat{\tau}_n$ für große Stichprobengrößen $n \rightarrow \infty$ gleich dem wahren, aber unbekannten, Wert $\tau(\theta)$ ist.

Ein Schätzer $\hat{\tau}_n$ für τ heißt **konsistent**, wenn für große Stichprobengrößen $n \rightarrow \infty$ die Wahrscheinlichkeit dafür, dass $\hat{\tau}_n(X)$ vom wahren, aber unbekannten, Wert $\tau(\theta)$ **abweicht** beliebig klein wird.

Ein Schätzer $\hat{\tau}_n$ für τ heißt **asymptotisch normalverteilt**, wenn für große Stichprobengrößen $n \rightarrow \infty$, die Verteilung von $\hat{\tau}_n$ durch eine Normalverteilung gegeben ist.

Ein Schätzer $\hat{\tau}_n$ für τ heißt **asymptotisch effizient**, wenn für große Stichprobengrößen $n \rightarrow \infty$ die Verteilung von $\hat{\tau}_n$ durch eine Normalverteilung mit Erwartungswertparameter $\tau(\theta)$ und Varianzparameter gleich der Cramér-Rao-Schranke gegeben ist.

Definition (Konsistenz)

$X_1, \dots, X_n \sim p_\theta$ sei die Stichprobe eines parametrischen statistischen Produktmodells $\mathcal{M}$ und $\hat{\tau}_n$ sei ein Schätzer von τ . Eine Folge von Schätzern $\hat{\tau}_1, \hat{\tau}_2, \dots$ wird dann eine *konsistente Folge von Schätzern* genannt, wenn für jedes $\epsilon > 0$ und jedes $\theta \in \Theta$ gilt, dass

$$\lim_{n \rightarrow \infty} \mathbb{P}_\theta (|\hat{\tau}_n(X) - \tau(\theta)| \geq \epsilon) = 0.$$

Wenn $\hat{\tau}_1, \hat{\tau}_2, \dots$ eine konsistente Folge von Schätzern ist, dann heißt $\hat{\tau}_n$ *konsistenter Schätzer*.

Bemerkungen

- Für $n \rightarrow \infty$ wird die Wahrscheinlichkeit, dass $\hat{\tau}_n(X)$ beliebig nah bei $\tau(\theta)$ liegt, groß.
- Für $n \rightarrow \infty$ wird die Wahrscheinlichkeit, dass $\hat{\tau}_n(X)$ von $\tau(\theta)$ abweicht, klein.
- Diese Eigenschaften gelten für alle möglichen wahren, aber unbekannten, Parameterwerte.
- Die Konvergenz ist *Konvergenz in Wahrscheinlichkeit*.
- Konsistenz von Schätzern kann direkt oder mit Kriterien nachgewiesen werden.

Theorem (Mittlerer-Quadratischer-Fehler-Kriterium für Konsistenz)

$X_1, \dots, X_n \sim p_\theta$ sei die Stichprobe eines parametrischen statistischen Produktmodells $\mathcal{M}$ und $\hat{\tau}_n$ sei ein Schätzer von τ . Wenn $\lim_{n \rightarrow \infty} \text{MQF}(\hat{\tau}_n) = 0$ gilt, dann ist $\hat{\tau}_n$ ein konsistenter Schätzer

Beweis

Mit der Chebychev-Ungleichung folgt, dass

$$\mathbb{P}_\theta (|\hat{\tau}_n(X) - \tau(\theta)| \geq \epsilon) \leq \frac{\mathbb{E}_\theta ((\hat{\tau}_n(X) - \tau(\theta))^2)}{\epsilon^2} \quad (62)$$

Grenzwertbildung ergibt dann

$$\lim_{n \rightarrow \infty} \mathbb{P}_\theta (|\hat{\tau}_n(X) - \tau(\theta)| \geq \epsilon) \leq \frac{1}{\epsilon^2} \lim_{n \rightarrow \infty} \mathbb{E}_\theta ((\hat{\tau}_n(X) - \tau(\theta))^2). \quad (63)$$

Wenn also $\lim_{n \rightarrow \infty} \mathbb{E}_\theta ((\hat{\tau}_n(X) - \tau(\theta))^2) = 0$ gilt, dann gilt mit $\mathbb{P}_\theta (|\hat{\tau}_n(X) - \tau(\theta)| \geq \epsilon) \geq 0$, dass

$$\lim_{n \rightarrow \infty} \mathbb{P}_\theta (|\hat{\tau}_n(X) - \tau(\theta)| \geq \epsilon) = 0. \quad (64)$$

Also ist $\hat{\tau}_n$ ein konsistenter Schätzer.

□

Theorem (Bias-Varianz-Kriterium für Konsistenz)

$X_1, \dots, X_n \sim p_\theta$ sei die Stichprobe eines parametrischen statistischen Produktmodells $\mathcal{M}$ und $\hat{\tau}_n$ sei ein Schätzer von τ . Wenn

$$\lim_{n \rightarrow \infty} B(\hat{\tau}_n) = 0 \text{ und } \lim_{n \rightarrow \infty} \mathbb{V}_\theta(\hat{\tau}_n) = 0 \quad (65)$$

gilt, dann ist $\hat{\tau}_n$ ein konsistenter Schätzer

Beweis

Wenn $n \rightarrow \infty$, dann gilt $B(\hat{\tau}_n) \rightarrow 0$, also auch $B(\hat{\tau}_n)^2 \rightarrow 0$. Wenn für $n \rightarrow \infty$ sowohl $B(\hat{\tau}_n)^2 \rightarrow 0$ als auch $\mathbb{V}_\theta(\hat{\tau}_n) \rightarrow 0$, dann gilt auch $\lim_{n \rightarrow \infty} \text{MQF}(\hat{\tau}_n) = 0$. Also gilt mit dem MQF-Kriterium, dass $\hat{\tau}_n$ konsistent ist.

□

Beispiel (Konsistenz des Stichprobenmittels)

Für $X_1, \dots, X_n \sim N(\mu, \sigma^2)$ folgt die Konsistenz des Stichprobenmittels als Schätzer für $\mathbb{E}(X_0)$ aus dem Bias-Varianz-Kriterium durch

$$B(\bar{X}_n) = 0 \text{ und } \lim_{n \rightarrow \infty} \mathbb{V}_\theta(\bar{X}_n) = \lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{V}(X_0) = 0. \quad (66)$$

Konsistenz

Simulation $\bar{X}_n$ bei $X_1, \dots, X_n \sim N(\mu, \sigma^2)$ mit $\mu = 1, \sigma^2 = 2$

```
# Modellformulierung
mu      = 1                                # w.a.u. \mu Wert
sigsqr  = 2                                # w.a.u. \sigma^2 Wert
n       = seq(1, 1e3, by = 10)            # Stichprobengroesse n
eps     = c(0.15, 0.10, 0.05)            # \epsilon Werte
ne      = length(eps)                     # Anzahl \epsilon Werte
nn      = length(n)                       # Anzahl Stichprobengroessen
ns      = 1e3                             # Anzahl Simulationen
E       = array(rep(NaN, nn, ne, ns),     # Ereignisindikatorarray
               dim = c(nn, ne, ns))

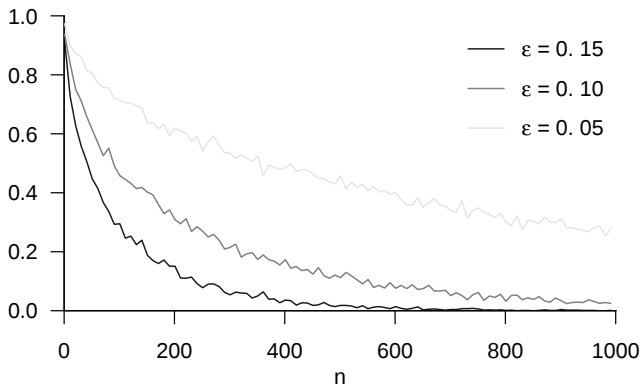
# Simulation
for(e in seq_along(eps)){                 # \epsilon Iterationen
  for(i in seq_along(n)){                 # n Iterationen
    for(s in 1:ns){                       # Simulationsiterationen

      # Stichprobenrealisationen
      x = rnorm(n[i], mu, sqrt(sigsqr))
      if(abs(mean(x) - mu) >= eps[e]){    # |X-bar - \mu| \ge \epsilon
        E[i, e, s] = 1
      } else {                             # |X-bar - \mu| < \epsilon
        E[i, e, s] = 0
      }
    }
  }
}

# Schaetzung von \mathbb{P}(|\hat{\tau}_n(X) - \tau(\theta)| \ge \epsilon)
P_hat   = apply(E, c(1, 2), mean)
```

Simulation $\bar{X}_n$ bei $X_1, \dots, X_n \sim N(\mu, \sigma^2)$ mit $\mu = 1, \sigma^2 = 2$

$$\hat{\mathbb{P}}(|\bar{X}_n - \mu| \geq \epsilon)$$



Asymptotische Schätzereigenschaften

Ein Schätzer $\hat{\tau}_n$ für τ heißt **asymptotisch erwartungstreu**, wenn der Erwartungswert von $\hat{\tau}_n$ für große Stichprobengrößen $n \rightarrow \infty$ gleich dem wahren, aber unbekannten, Wert $\tau(\theta)$ ist.

Ein Schätzer $\hat{\tau}_n$ für τ heißt **konsistent**, wenn für große Stichprobengrößen $n \rightarrow \infty$ die Wahrscheinlichkeit dafür, dass $\hat{\tau}_n(X)$ vom wahren, aber unbekannten, Wert $\tau(\theta)$ **abweicht** beliebig klein wird.

Ein Schätzer $\hat{\tau}_n$ für τ heißt **asymptotisch normalverteilt**, wenn für große Stichprobengrößen $n \rightarrow \infty$, die Verteilung von $\hat{\tau}_n$ durch eine Normalverteilung gegeben ist.

Ein Schätzer $\hat{\tau}_n$ für τ heißt **asymptotisch effizient**, wenn für große Stichprobengrößen $n \rightarrow \infty$ die Verteilung von $\hat{\tau}_n$ durch eine Normalverteilung mit Erwartungswertparameter $\tau(\theta)$ und Varianzparameter gleich der Cramér-Rao-Schranke gegeben ist.

Definition (Asymptotische Normalität)

$X_1, \dots, X_n \sim p_\theta$ sei die Stichprobe eines parametrischen statistischen Produktmodells $\mathcal{M}$ und $\hat{\theta}_n$ sei ein Parameterschätzer für θ . Weiterhin sei $\tilde{\theta} \sim N(\mu, \sigma^2)$ eine normalverteilte Zufallsvariable mit Erwartungswertparameter μ und Varianzparameter σ^2 . Wenn $\hat{\theta}_n$ in Verteilung gegen $\tilde{\theta}$ konvergiert, dann heißt $\hat{\theta}_n$ *asymptotisch normalverteilt* und wir schreiben

$$\hat{\theta}_n \stackrel{a}{\sim} N(\mu, \sigma^2). \quad (67)$$

Bemerkung

- Konvergenz in Verteilung heißt $\lim_{n \rightarrow \infty} P_n(\hat{\theta}_n) = P(\tilde{\theta})$.

Asymptotische Schätzereigenschaften

Ein Schätzer $\hat{\tau}_n$ für τ heißt **asymptotisch erwartungstreu**, wenn der Erwartungswert von $\hat{\tau}_n$ für große Stichprobengrößen $n \rightarrow \infty$ gleich dem wahren, aber unbekannten, Wert $\tau(\theta)$ ist.

Ein Schätzer $\hat{\tau}_n$ für τ heißt **konsistent**, wenn für große Stichprobengrößen $n \rightarrow \infty$ die Wahrscheinlichkeit dafür, dass $\hat{\tau}_n(X)$ vom wahren, aber unbekannten, Wert $\tau(\theta)$ **abweicht** beliebig klein wird.

Ein Schätzer $\hat{\tau}_n$ für τ heißt **asymptotisch normalverteilt**, wenn für große Stichprobengrößen $n \rightarrow \infty$, die Verteilung von $\hat{\tau}_n$ durch eine Normalverteilung gegeben ist.

Ein Schätzer $\hat{\tau}_n$ für τ heißt **asymptotisch effizient**, wenn für große Stichprobengrößen $n \rightarrow \infty$ die Verteilung von $\hat{\tau}_n$ durch eine Normalverteilung mit Erwartungswertparameter $\tau(\theta)$ und Varianzparameter gleich der Cramér-Rao-Schranke gegeben ist,

Definition (Asymptotische Effizienz)

$X_1, \dots, X_n \sim p_\theta$ sei die Stichprobe eines parametrischen statistischen Produktmodells $\mathcal{M}$ und $\hat{\theta}_n$ sei ein Parameterschätzer für θ . Weiterhin sei $J_n(\theta)$ die erwartete Fisher-Information der Stichprobe $X_1, \dots, X_n$. Wenn gilt, dass

$$\hat{\theta}_n \stackrel{a}{\sim} N(\theta, J_n(\theta)^{-1}), \quad (68)$$

dann heißt $\hat{\theta}_n$ *asymptotisch effizient*.

Bemerkungen

- Asymptotische Effizienz impliziert asymptotische Normalität.
- Asymptotische Effizienz impliziert asymptotische Erwartungstreue.
- Die Varianz der asymptotischen Verteilung heißt *asymptotische Varianz*.
- Die Varianz eines asymptotisch effizienten Schätzers ist gleich der Cramér-Rao-Schranke.
- Der Begriff der *Effizienz* wird in der Literatur nicht einheitlich verwendet.

Schätzereigenschaften

- Vorbemerkungen
- Schätzereigenschaften bei endlichen Stichproben
- Asymptotische Schätzereigenschaften
- **Eigenschaften von Maximum-Likelihood Schätzern**
- Selbstkontrollfragen

Theorem (Eigenschaften von ML Schätzern)

$X_1, \dots, X_n \sim p_\theta$ sei die Stichprobe eines parametrischen statistischen Produktmodells $\mathcal{M}$ und $\hat{\theta}_n^{\text{ML}}$ sei ein ML Schätzer für θ . Dann gilt, dass $\hat{\theta}_n^{\text{ML}}$

- (1) nicht notwendigerweise erwartungstreu, aber
- (2) konsistent,
- (3) asymptotisch normalverteilt,
- (4) asymptotisch erwartungstreu, und
- (5) asymptotisch effizient ist.

Für einen Beweis verweisen wir auf Held and Sabanés Bové (2014, Abschnitt 3.4)

Simulation der asymptotische Effizienz des Bernoulli ML Parameterschätzers

- Es sei $X_1, \dots, X_n \sim \text{Bern}(\mu)$.
- Der ML Schätzer von μ ist $\hat{\mu}_n^{\text{ML}} = \frac{1}{n} \sum_{i=1}^n X_i$.
- Die erwartete Fisher-Information der Stichprobe $X_1, \dots, X_n$ ist

$$J_n(\mu) = \frac{n}{\mu(1-\mu)}. \quad (69)$$

- Also gilt

$$\hat{\mu}_n^{\text{ML}} \overset{a}{\sim} N\left(\mu, \frac{\mu(1-\mu)}{n}\right). \quad (70)$$

Eigenschaften von Maximum-Likelihood Schätzern

Simulation der asymptotische Effizienz des Bernoulli ML Parameterschätzers

```
# Modellformulierung
mu          = 0.4                                     # w.a.u. Parameterwert
n_all       = c(1e1,5e1,1e2)                         # Stichprobengroesse n
ns          = 1e4                                     # Anzahl der Simulationen
mu_hat_ML   = matrix(                                # ML Schaetzerarray
  rep(NA,
    length(n_all)*ns),
  nrow = length(n_all))

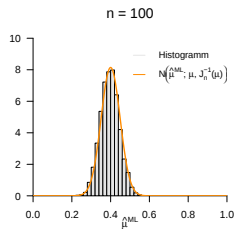
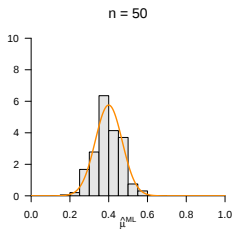
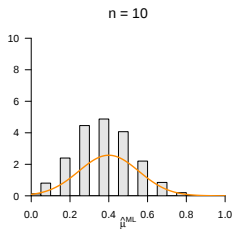
mu_hat_ML_r = 1e3                                     # ML Schaetzerraumaufloesung
mu_hat_ML_x = seq(0,1,len = mu_hat_ML_r)             # ML Schaetzerraum
mu_hat_ML_p = matrix(                                # ML WDF Array
  rep(NA, length(n_all)*mu_hat_ML_r),
  nrow = length(n_all))

# Stichprobengroesseniterationen
for(i in seq_along(n_all)){

  # Simulationsiterationen
  for(s in 1:ns){
    x          = rbinom(n_all[i],1,mu)               # Stichprobenrealisation
    mu_hat_ML[i,s] = mean(x)                         # ML Schaetzer
  }

  # WDF der asymptotischen Verteilung
  mu_hat_ML_p[i,] = dnorm(mu_hat_ML_x, mu, sqrt(mu*(1-mu)/n_all[i]))
}
```

Simulation der asymptotische Effizienz des Bernoulli ML Parameterschätzers



Schätzereigenschaften

- Vorbemerkungen
- Schätzereigenschaften bei endlichen Stichproben
- Asymptotische Schätzereigenschaften
- Eigenschaften von Maximum-Likelihood Schätzern
- **Selbstkontrollfragen**

1. Definieren und erläutern Sie den Begriff der Erwartungstreue eines Schätzers.
2. Definieren Sie die Begriffe der Varianz und des Standardfehlers eines Schätzers.
3. Definieren Sie den Begriff der Scorefunktion einer Stichprobe.
4. Definieren Sie den Begriff der Fisher-Information einer Stichprobe.
5. Definieren Sie den Begriff der erwarteten Fisher-Information einer Stichprobe.
6. Geben Sie das Theorem zur Cramér-Rao-Ungleichung wieder.
7. Erläutern Sie den Begriff der Cramér-Rao-Schranke.
8. Definieren Sie den Begriff des MQFs eines Schätzers.
9. Geben Sie das Theorem zur Zerlegung des MQFs wieder.
10. Erläutern Sie den Begriff des asymptotischen Erwartungstreue eines Schätzers
11. Erläutern Sie den Begriff der Konsistenz eines Schätzers.
12. Erläutern Sie den Begriff der asymptotischen Normalität eines Schätzers.
13. Erläutern Sie den Begriff der asymptotischen Effizienz eines Schätzers.
14. Nennen Sie fünf Eigenschaften eines ML Schätzers.

Literatur

Casella, G. and Berger, R. (2012). *Statistical Inference*. Duxbury.

DeGroot, M. H. and Schervish, M. J. (2012). *Probability and Statistics*. Addison-Wesley, Boston, 4th ed edition.

Georgii, H.-O. (2009). *Stochastik: Einführung in die Wahrscheinlichkeitstheorie und Statistik*. DeGruyter-Lehrbuch. de Gruyter, Berlin, 4., überarb. und erw. aufl edition.

Held, L. and Sabanés Bové, D. (2014). *Applied Statistical Inference*. Springer Berlin Heidelberg, Berlin, Heidelberg.

van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, UK ; New York, NY, USA.

Wasserman, L. (2004). *All of Statistics*.