



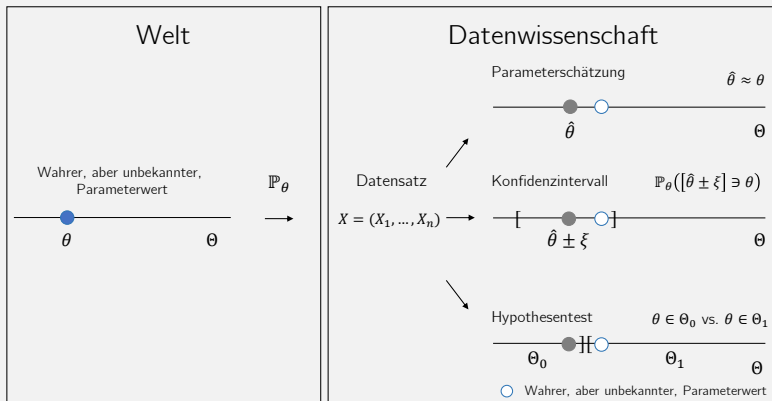
Wahrscheinlichkeitstheorie und Frequentistische Inferenz

BSc Psychologie WiSe 2021/22

Prof. Dr. Dirk Ostwald

(10) Parameterschätzung

Modell und Standardprobleme der Frequentistischen Inferenz



(1) Parameterschätzung

Ziel der Parameterschätzung ist es, einen möglichst guten Tipp für den wahren, aber unbekannten, Parameterwert (oder eine Funktion dessen) abzugeben, typischerweise basierend auf der Beobachtung einer Realisierung von $X_1, \dots, X_n \sim p_\theta$.

(2) Konfidenzintervalle

Das Ziel der Bestimmung von Konfidenzintervallen ist es, basierend auf der Verteilung möglicher Parameterschätzwerte eine quantitative Aussage über die mit dem Schätzwert assoziierte Unsicherheit zu treffen.

(3) Hypothesentests

Das Ziel der Auswertung von Hypothesentests ist es, basierend auf der angenommenen Verteilung der Beobachtungen $X_1, \dots, X_n$ in einer möglichst sinnvollen Form zu entscheiden, ob der wahre, aber unbekannte Parameterwert, in einer von zwei sich gegenseitig ausschließenden Untermengen des Parameterraumes, welche man als Hypothesen bezeichnet, liegt.

Standardannahmen Frequentistischer Inferenz

$\mathcal{M}$ sei ein statistisches Modell mit $X_1, \dots, X_n \sim \mathbb{P}_\theta$. Es wird angenommen, dass ein vorliegender Datensatz eine der möglichen Realisierungen von $X_1, \dots, X_n \sim \mathbb{P}_\theta$ ist. Aus frequentistischer Sicht kann man die Erhebung von Datensätzen unendlich oft wiederholen und zu jedem Datensatz Statistiken auswerten.

Datensatz (1) : $x^{(1)} = (x_1^{(1)}, x_2^{(1)}, \dots, x_n^{(1)})$, Statistik (1): $S : \mathbb{R}^n \rightarrow \Sigma, x^{(1)} \mapsto S(x^{(1)})$

Datensatz (2) : $x^{(2)} = (x_1^{(2)}, x_2^{(2)}, \dots, x_n^{(2)})$, Statistik (2): $S : \mathbb{R}^n \rightarrow \Sigma, x^{(2)} \mapsto S(x^{(2)})$

Datensatz (3) : $x^{(3)} = (x_1^{(3)}, x_2^{(3)}, \dots, x_n^{(3)})$, Statistik (3): $S : \mathbb{R}^n \rightarrow \Sigma, x^{(3)} \mapsto S(x^{(3)})$

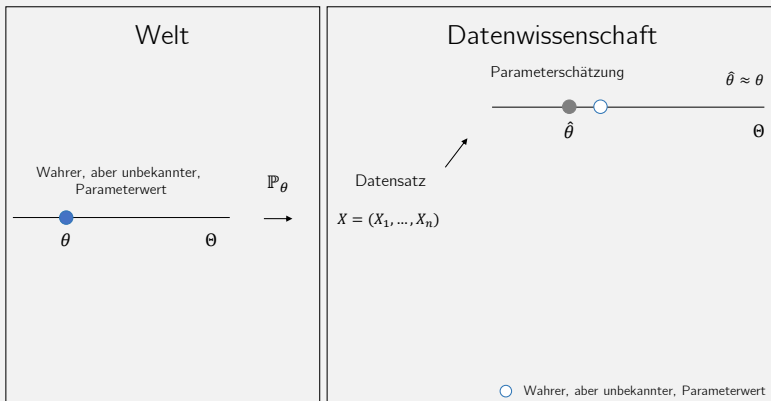
Datensatz (4) : $x^{(4)} = (x_1^{(4)}, x_2^{(4)}, \dots, x_n^{(4)})$, Statistik (4): $S : \mathbb{R}^n \rightarrow \Sigma, x^{(4)} \mapsto S(x^{(4)})$

...

Um die Qualität statistischer Methoden zu beurteilen betrachtet die frequentistische Statistik deshalb die Wahrscheinlichkeitsverteilungen von Statistiken und Schätzern unter der Annahme von $X_1, \dots, X_n \sim p_\theta$.

Wenn eine statistische Methode im Sinne der frequentistischen Standardannahmen "gut" ist, dann heißt das also, dass sie bei häufiger Anwendung "im Mittel gut" ist. Im Einzelfall, also im realen Normalfall nur eines vorliegenden Datensatzes, kann sie auch "schlecht" sein.

Modell und Standardprobleme der Frequentistischen Inferenz



Grundbegriffe

Maximum-Likelihood Schätzer

Schätzereigenschaften bei endlichen Stichproben

Asymptotische Schätzereigenschaften

Eigenschaften von Maximum-Likelihood Schätzern

Selbstkontrollfragen

Grundbegriffe

Maximum-Likelihood Schätzer

Schätzereigenschaften bei endlichen Stichproben

Asymptotische Schätzereigenschaften

Eigenschaften von Maximum-Likelihood Schätzern

Selbstkontrollfragen

Definition (Parameterpunktschätzer)

$\mathcal{M} := (\mathcal{X}, \mathcal{A}, \{\mathbb{P}_\theta | \theta \in \Theta\})$ sei ein statistisches Modell, $(\Theta, \mathcal{S})$ sei ein Messraum, und $\hat{\theta} : \mathcal{X} \rightarrow \Theta$ sei eine Abbildung. Dann nennen wir $\hat{\theta}$ einen *Parameterpunktschätzer* für θ .

Bemerkungen

- Parameterpunktschätzer nennt man auch einfach *Parameterschätzer*.
- Parameterpunktschätzer sind Schätzer mit $\tau := \text{id}_\Theta$
- Parameterschätzer nehmen Zahlwerte in Θ an.
- Notationstechnisch wird oft nicht zwischen $\hat{\theta}$ und $\hat{\theta}(x)$ unterschieden.

Prinzipien zur Gewinnung von Parameterschätzern

Die Definition eines Parameterschätzers macht keine Aussage darüber, wie man Parameterschätzer findet. Zur Gewinnung von Parameterschätzern in statistischen Modellen haben sich deshalb verschiedene Prinzipien etabliert. Populäre Prinzipien zur Gewinnung von Parameterschätzern sind

- Momentenmethode ($\approx$ est. 1890)
- Maximum-Likelihood Methode ($\approx$ est. 1920)
- M-, Z-, W-Schätzung ($\approx$ est. 1960)

Per se garantiert keine der obengenannten Methoden, dass die mit ihrer Hilfe generierten Parameterschätzer in einem wohldefinierten Sinn gute Schätzer sind.

Die Eigenschaften von durch die Maximum-Likelihood Methode generierten Schätzern sind generell wünschenswert. Wir betrachten also in der Folge nur die Maximum-Likelihood Methode genauer. Mithilfe der Maximum-Likelihood Methode generierte Parameterpunktschätzer nennen wir *Maximum-Likelihood (ML) Schätzer*.

Grundbegriffe

Maximum-Likelihood Schätzer

Schätzereigenschaften bei endlichen Stichproben

Asymptotische Schätzereigenschaften

Eigenschaften von Maximum-Likelihood Schätzern

Selbstkontrollfragen

Definition (Likelihood-Funktion und Log-Likelihood-Funktion)

$\mathcal{M}$ sei ein parametrisches statistisches Produktmodell mit WMF oder WDF p_θ , so dass $X_1, \dots, X_n \sim p_\theta$. Dann ist die *Likelihood Funktion* definiert als

$$L_n : \Theta \rightarrow [0, \infty[, \theta \mapsto L_n(\theta) := \prod_{i=1}^n p_\theta(x_i). \quad (1)$$

und die *Log-Likelihood-Funktion* ist definiert als

$$\ell_n : \Theta \rightarrow \mathbb{R}, \theta \mapsto \ell_n(\theta) := \ln L_n(\theta). \quad (2)$$

Bemerkungen

- L_n ist eine Funktion des Parameters eines statistischen Modells.
- Werte von L_n sind die gemeinsamen W-Massen bzw. W-Dichten von Daten $x_1, \dots, x_n$.
- Generell gibt es keinen Grund anzunehmen, dass L_n über Θ zu 1 integriert.
- Die Likelihood-Funktion ist also keine WMF oder WDF.
- Die Log-Likelihood-Funktion ist die logarithmierte Likelihood-Funktion.

Definition (Maximum-Likelihood Schätzer)

$\mathcal{M}$ sei ein parametrisches statistisches Produktmodell mit Parameter $\theta \in \Theta$. Ein *Maximum-Likelihood (ML) Schätzer* von θ ist definiert als

$$\hat{\theta}_n^{\text{ML}} : \mathcal{X} \rightarrow \Theta, x \mapsto \hat{\theta}_n^{\text{ML}}(x) := \arg \max_{\theta \in \Theta} L_n(\theta) = \arg \max_{\theta \in \Theta} \ell_n(\theta). \quad (3)$$

Bemerkungen

- $L_n(\theta) := \prod_{i=1}^n p_\theta(x_i)$ hängt von $x_1, \dots, x_n$ ab, also hängt auch $\hat{\theta}_n^{\text{ML}}(x)$ von $x_1, \dots, x_n$ ab.
- Weil $\ln$ monoton steigend ist, entspricht eine Maximumstelle von ℓ_n einer Maximumstelle von L_n .
- Das Arbeiten mit der Log-Likelihood-Funktion ist oft einfacher als mit der Likelihood Funktion.
- Multiplikation von L_n mit einer positive Konstante, die nicht von θ abhängt, verändert einen ML Schätzer nicht, konstante additive Terme in der Log-Likelihood können also vernachlässigt werden.
- Maximum-Likelihood Schätzung ist ein Optimierungsproblem

Vorgehen zur Gewinnung von Maximum-Likelihood Schätzern

- (1) Formulierung der Log-Likelihood-Funktion.
- (2) Auswertung der Ableitung der Log-Likelihood-Funktion und Nullsetzen.
- (3) Auflösen nach potentiellen Maximumstellen.

Dabei nutzt man typischerweise

- Methoden der analytische Optimierung in klassischen Beispielen und
- Methoden der numerische Optimierung im Anwendungskontext.

Beispiel (Bernoullimodell)

$\mathcal{M}$ sei das Bernoullimodell, also $X_1, \dots, X_n \sim \text{Bern}(\mu)$.

(1) Formulierung der Log-Likelihood-Funktion

Es gilt

$$L_n :]0, 1[\rightarrow]0, 1[, \mu \mapsto L_n(\mu) := \prod_{i=1}^n \mu^{x_i} (1 - \mu)^{1-x_i} = \mu^{\sum_{i=1}^n x_i} (1 - \mu)^{n - \sum_{i=1}^n x_i}. \quad (4)$$

Logarithmieren ergibt

$$\ell_n :]0, 1[\rightarrow \mathbb{R}, \mu \mapsto \ell_n(\mu) = \ln \mu \sum_{i=1}^n x_i + \ln(1 - \mu) \left(n - \sum_{i=1}^n x_i \right). \quad (5)$$

Beispiel (Bernoullimodell)

(2) Auswertung der Ableitung der Log-Likelihood-Funktion und Nullsetzen

Es gilt

$$\begin{aligned}\frac{d}{d\mu} \ell_n(\mu) &= \frac{d}{d\mu} \left(\ln \mu \sum_{i=1}^n x_i + \ln(1 - \mu) \left(n - \sum_{i=1}^n x_i \right) \right) \\ &= \frac{d}{d\mu} \ln \mu \sum_{i=1}^n x_i + \frac{d}{d\mu} \ln(1 - \mu) \left(n - \sum_{i=1}^n x_i \right) \\ &= \frac{1}{\mu} \sum_{i=1}^n x_i - \frac{1}{1 - \mu} \left(n - \sum_{i=1}^n x_i \right).\end{aligned}\tag{6}$$

Die sogenannte *Maximum-Likelihood Gleichung* ergibt sich in diesem Beispiel also zu

$$\frac{1}{\hat{\mu}_n^{\text{ML}}} \sum_{i=1}^n x_i - \frac{1}{1 - \hat{\mu}_n^{\text{ML}}} \left(n - \sum_{i=1}^n x_i \right) = 0.\tag{7}$$

Beispiel (Bernoullimodell)

(3) Auflösen nach potentiellen Maximumstellen

Es gilt

$$\begin{aligned} & \frac{1}{\hat{\mu}_n^{\text{ML}}} \sum_{i=1}^n x_i - \frac{1}{1 - \hat{\mu}_n^{\text{ML}}} \left(n - \sum_{i=1}^n x_i \right) = 0 \\ \Leftrightarrow & \hat{\mu}_n^{\text{ML}} (1 - \hat{\mu}_n^{\text{ML}}) \left(\frac{1}{\hat{\mu}_n^{\text{ML}}} \sum_{i=1}^n x_i - \frac{1}{1 - \hat{\mu}_n^{\text{ML}}} \left(n - \sum_{i=1}^n x_i \right) \right) = 0 \\ \Leftrightarrow & \sum_{i=1}^n x_i - \hat{\mu}_n^{\text{ML}} \sum_{i=1}^n x_i - n \hat{\mu}_n^{\text{ML}} + \hat{\mu}_n^{\text{ML}} \sum_{i=1}^n x_i = 0 \quad (8) \\ \Leftrightarrow & n \hat{\mu}_n^{\text{ML}} = \sum_{i=1}^n x_i \\ \Leftrightarrow & \hat{\mu}_n^{\text{ML}} = \frac{1}{n} \sum_{i=1}^n x_i. \end{aligned}$$

Beispiel (Bernoullimodell)

$\hat{\mu}_n^{\text{ML}} = \frac{1}{n} \sum_{i=1}^n x_i$ ist also ein potentieller Maximum-Likelihood Schätzer von μ . Dies kann durch Betrachten der zweiten Ableitung von ℓ_n verifiziert werden, worauf hier verzichtet werden soll.

$$\hat{\mu}_n^{\text{ML}} : \{0, 1\}^n \rightarrow [0, 1], x \mapsto \hat{\mu}_n^{\text{ML}}(x) := \frac{1}{n} \sum_{i=1}^n x_i \quad (9)$$

ist also ein Maximum-Likelihood Schätzer von μ im Bernoullimodell.

Beispiel (Normalverteilungsmodell)

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

(1) Formulierung der Log-Likelihood-Funktion

Es gilt

$$\begin{aligned} L_n : \mathbb{R} \times \mathbb{R}_{>0} &\rightarrow \mathbb{R}_{>0}, (\mu, \sigma^2) \mapsto L_n(\mu, \sigma^2) := \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x_i - \mu)^2\right) \\ &= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2\right). \end{aligned} \quad (10)$$

Logarithmieren ergibt

$$\ell_n : \mathbb{R} \times \mathbb{R}_{>0} \rightarrow \mathbb{R}, (\mu, \sigma^2) \mapsto \ell_n(\mu, \sigma^2) = -\frac{n}{2} \ln 2\pi - \frac{n}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2. \quad (11)$$

Beispiel (Normalverteilungsmodell)

(2) Auswertung der Ableitung der Log-Likelihood-Funktion und Nullsetzen

Es ergibt sich

$$\frac{d}{d\mu} \ell_n(\mu) = -\frac{d}{d\mu} \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 = -\frac{1}{2\sigma^2} \sum_{i=1}^n \frac{d}{d\mu} (x_i - \mu)^2 = \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu). \quad (12)$$

Weiterhin ergibt sich

$$\frac{d}{d\sigma^2} \ell_n(\sigma^2) = -\frac{n}{2} \frac{d}{d\sigma^2} \ln \sigma^2 - \frac{d}{d\sigma^2} \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (x_i - \mu)^2. \quad (13)$$

Die Maximum-Likelihood Gleichungen haben also die Form

$$\begin{aligned} \sum_{i=1}^n (x_i - \hat{\mu}_n^{\text{ML}}) &= 0 \\ -\frac{n}{2\hat{\sigma}_n^{2\text{ML}}} + \frac{1}{2\hat{\sigma}_n^{4\text{ML}}} \sum_{i=1}^n (x_i - \mu)^2 &= 0 \end{aligned} \quad (14)$$

Beispiel (Normalverteilungsmodell)

(3) Auflösen nach potentiellen Maximumstellen

Es ergibt sich

$$\sum_{i=1}^n (x_i - \hat{\mu}_n^{\text{ML}}) = 0 \Leftrightarrow \sum_{i=1}^n x_i = n \hat{\mu}_n^{\text{ML}} \Leftrightarrow \hat{\mu}_n^{\text{ML}} = \frac{1}{n} \sum_{i=1}^n x_i. \quad (15)$$

Also ist $\hat{\mu}_n^{\text{ML}} = n^{-1} \sum_{i=1}^n x_i$ ein potentieller Maximum-Likelihood Schätzer von μ . Einsetzen ergibt dann weiterhin

$$\begin{aligned} -\frac{n}{2\hat{\sigma}_n^{2\text{ML}}} + \frac{1}{2\hat{\sigma}_n^{4\text{ML}}} \sum_{i=1}^n (x_i - \hat{\mu}_n^{\text{ML}})^2 &= 0 \\ \Leftrightarrow -n\hat{\sigma}_n^{2\text{ML}} + \sum_{i=1}^n (x_i - \hat{\mu}_n^{\text{ML}})^2 &= 0 \end{aligned} \quad (16)$$

$$\Leftrightarrow \hat{\sigma}_n^{2\text{ML}} = \frac{1}{n} \sum_{i=1}^n (x_i - \hat{\mu}_n^{\text{ML}})^2.$$

$\hat{\sigma}_n^{2\text{ML}} = n^{-1} \sum_{i=1}^n (x_i - \hat{\mu}_n^{\text{ML}})^2$ ist also potentieller Maximum-Likelihood Schätzer von σ^2 .

Beispiel (Normalverteilungsmodell)

Beide potentiellen Maximum-Likelihood Schätzer können durch Betrachten der zweiten Ableitung von ℓ_n verifiziert werden, worauf hier verzichtet werden soll.

Also sind

$$\hat{\mu}_n^{\text{ML}} : \mathbb{R}^n \rightarrow \mathbb{R}, x \mapsto \hat{\mu}_n^{\text{ML}}(x) := \frac{1}{n} \sum_{i=1}^n x_i \quad (17)$$

und

$$\hat{\sigma}_n^{2\text{ML}} : \mathbb{R}^n \rightarrow \mathbb{R}_{\geq 0}, x \mapsto \hat{\sigma}_n^{2\text{ML}}(x) := \frac{1}{n} \sum_{i=1}^n (x_i - \hat{\mu}_n^{\text{ML}})^2. \quad (18)$$

die Maximum-Likelihood Schätzer von μ und σ^2 im Normalverteilungsmodell. $\hat{\mu}_n^{\text{ML}}$ ist identisch mit dem Stichprobenmittel $\bar{X}_n$, $\hat{\sigma}_n^{2\text{ML}}$ ist dagegen nicht identisch mit der Stichprobenvarianz S_n^2 .

Anwendungsbeispiel

Experimentelle Bedingung
(Gruppen von $n = 50$)

Psychotherapie

Klassisch

Pre-BDI



Post-BDI



Online

Pre-BDI



Post-BDI



Anwendungsbeispiel

Standardannahmen der Frequentistischen Inferenz

Wir legen das Normalverteilungsmodell zugrunde, d.h. wir nehmen an, dass die BDI Werte Realisierungen von u.i.v. Zufallsvariablen

$$X_{ijk} \sim N\left(\mu_{ij}, \sigma^2\right), i = 1, 2, j = 1, 2, k = 1, \dots, n_i \quad (19)$$

wobei $i \in \{1, 2\}$ die experimentelle Bedingung (1 = Klassisch, 2 = Online), $j \in \{1, 2\}$ den Zeitpunkt der Messung (1 = Pre, 1 = Post) und $k \in \mathbb{N}_i$ den Proband:innen Index in der i ten experimentellen Bedingung bezeichnen sollend Dies entspricht der Annahme, dass sich der BDI Wert einer Proband:in durch Addition einer normalverteilten Fehlervariable mit Erwartungswertparameter 0 und Varianzparameter σ^2 zu den innerhalb einer Versuchsbedingung und einer Messung identischen Wert μ_{ij} ergibt.

Standardprobleme der Frequentistischen Inferenz

- (1) Was sind sinnvolle Tipps für die wahren, aber unbekannten, Parameterwerte μ_{ij} und σ^2 ?
- (2) Wie hoch ist im frequentistischen Sinn die mit diesen Tipps assoziierte Unsicherheit?
- (3) Entscheiden wir uns sinnvollerweise für die Hypothese, dass gilt

$$(\mu_{12} - \mu_{11}) - (\mu_{21} - \mu_{22}) \neq 0? \quad (20)$$

Maximum-Likelihood Schätzer

Anwendungsbeispiel

Wir beschränken uns hier zunächst auf die Pre.BDI Daten der Klassischen Bedingung, formal

$$X_k \sim N(\mu, \sigma) \text{ mit } X_k := X_{11k} \text{ für } k = 1, \dots, n \text{ und } \mu := \mu_{11}. \quad (21)$$

```
# Einlesen und Auswahl der Daten
fname = file.path(getwd(), "10_Daten", "psychotherapie_datensatz.csv")
D = read.table(fname, sep = ",")
X = D$Pre.BDI[D$Bedingung == "Klassisch"] # i = 1 (Klassisch), j = 1 (Pre) Daten

# Maximum-Likelihood Schätzung der Gruppen-spezifischen Erwartungswertparameter \mu_{11}
mu_hat_11 = mean(X) # mean(x) berechnet das Stichprobenmittel des Datensatzes x
print(mu_hat_11) # Ausgabe
```

> [1] 18.2

```
# Maximum-Likelihood Schätzung der Gruppen-spezifischen Varianzparameter \sigma^2_{ij}
n = length(X) # Anzahl der Datenpunkte
sigsqr_hat = ((n-1)/n)*var(X) # var(x) berechnet die Stichprobenvarianz des Datensatzes x
print(sigsqr_hat)
```

> [1] 3.43

Basierend auf dem Prinzip der Maximum-Likelihood Schätzung sind also

$$\hat{\mu}_{50}^{\text{ML}} = 18.19, \text{ und } \hat{\sigma}_{50}^{2\text{ML}} = 3.43 \quad (22)$$

sinnvolle Tipps für μ und σ^2 basierend auf den vorliegenden 50 Datenpunkten.

Grundbegriffe

Maximum-Likelihood Schätzer

Schätzereigenschaften bei endlichen Stichproben

Asymptotische Schätzereigenschaften

Eigenschaften von Maximum-Likelihood Schätzern

Selbstkontrollfragen

Vorbemerkungen zu Frequentistischen Schätzereigenschaften

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 Parameteraum $\Theta \subset \mathbb{R}^d$ 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 (Σ, S) sei $\hat{\tau} : \mathcal{X} \rightarrow \Sigma$ ein Schätzer von $\tau : \Theta \rightarrow \Sigma$. Wir betrachten Erwartungswerts-, Varianz-, und Standardabweichungsschätzer, also Schätzer für

$$\tau : \Theta \rightarrow \Sigma, \theta \mapsto \tau(\theta) \text{ mit } \tau(\theta) := \mathbb{E}_\theta(X_0), \tau(\theta) := \mathbb{V}_\theta(X_0), \text{ und } \tau(\theta) := \mathbb{S}_\theta(X_0) \quad (23)$$

respektive, sowie Parameterschätzer, also Schätzer für

$$\tau : \Theta \rightarrow \Sigma, \tau(\theta) := \theta. \quad (24)$$

In der Folge 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 wie oben gesehen 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 $n \rightarrow \infty$.

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

$$\hat{\tau}_n(X) - \tau(\theta). \quad (25)$$

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

$$B(\hat{\tau}_n) := \mathbb{E}_\theta(\hat{\tau}_n(X)) - \tau(\theta). \quad (26)$$

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

$$B(\hat{\tau}_n) = 0 \Leftrightarrow \mathbb{E}_\theta(\hat{\tau}_n(X)) = \tau(\theta) \text{ für alle } \theta \in \Theta, n \in \mathbb{N}. \quad (27)$$

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 $\mathcal{M}$.

- Das Stichprobenmittel

$$\bar{X}_n := \frac{1}{n} \sum_{i=1}^n X_i \quad (28)$$

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

- Die Stichprobenvarianz

$$S_n^2 := \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2 \quad (29)$$

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

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

Es ergibt sich somit

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

Schließlich ergibt sich

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

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

$$S_n := \sqrt{S_n^2} \quad (30)$$

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

$$\mathbb{E}(S) = \mathbb{E}\left(\sqrt{S_n^2}\right) < \sqrt{\mathbb{E}(S_n^2)} = \sqrt{\mathbb{V}_\theta(X_0)} = \mathbb{S}_\theta(X_0). \quad (31)$$

□

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

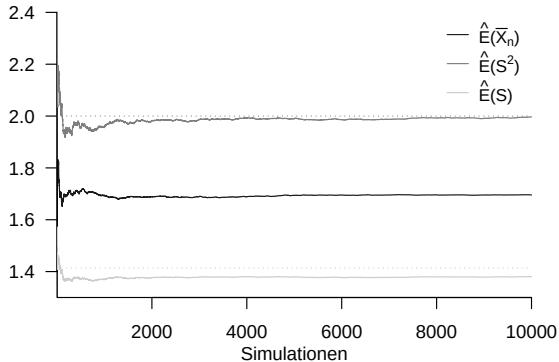
# Simulationsiterationen
for(s in 1:ns){

  # 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 $n = 12$, $\mu = 1.7$, $\sigma^2 = 2$, $\sigma \approx 1.41$)



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

$$\mathbb{V}_\theta(\hat{\tau}_n) := \mathbb{E}_\theta \left((\hat{\tau}_n(X) - \mathbb{E}_\theta(\hat{\tau}_n(X)))^2 \right). \quad (32)$$

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

$$\text{SE}(\hat{\tau}_n) := \sqrt{\mathbb{V}_\theta(\hat{\tau}_n)} \quad (33)$$

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

$$\text{SE}(\bar{X}_n) = \frac{\mathbb{S}_\theta(X_0)}{\sqrt{n}}. \quad (34)$$

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

$$\text{SE}(\bar{X}_n) = \sqrt{\mathbb{V}_\theta(\bar{X}_n)} = \sqrt{\frac{\mathbb{V}_\theta(X_0)}{n}} = \frac{\mathbb{S}_\theta(X_0)}{\sqrt{n}}. \quad (35)$$

Bemerkungen

- 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 $\hat{\text{SE}}(\bar{X}_n) = \frac{s_n}{\sqrt{n}}$.

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

$$\text{SE}(\hat{\mu}_n^{\text{ML}}) = \sqrt{\frac{\mu(1-\mu)}{n}}. \quad (36)$$

Beweis

Es gilt

$$\begin{aligned} \text{SE}(\hat{\mu}_n^{\text{ML}}) &= \sqrt{\mathbb{V}_{\mu}(\hat{\mu}_n^{\text{ML}})} = \sqrt{\mathbb{V}_{\mu}\left(\frac{1}{n} \sum_{i=1}^n X_i\right)} \\ &= \sqrt{\frac{1}{n^2} \sum_{i=1}^n \mathbb{V}_{\mu}(X_i)} = \sqrt{\frac{n\mu(1-\mu)}{n^2}} = \sqrt{\frac{\mu(1-\mu)}{n}}, \end{aligned} \quad (37)$$

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. $\square$

Bemerkung

- Ein Schätzer für den Standardfehler $\text{SE}(\hat{\mu}_n^{\text{ML}})$ ist $\hat{\text{SE}}(\hat{\mu}_n^{\text{ML}}) = \sqrt{\frac{\hat{\mu}_n^{\text{ML}}(1-\hat{\mu}_n^{\text{ML}})}{n}}$

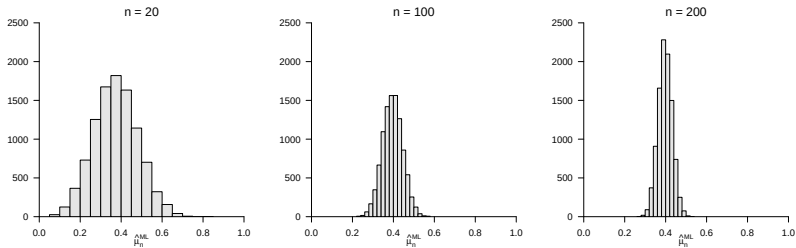
Simulation ($X_1, \dots, X_n \sim \text{Bern}(\mu)$ mit $\mu = 0.4$)

```
# Modellformulierung
mu           = 0.4                                # wahrer, aber unbekannter, Parameterwert
n_all        = c(20,100,200)                     # Stichprobengrößen n
ns           = 1e4                                # Anzahl der Simulationen
mu_hat_ML    = matrix(rep(NaN, length(n_all)*ns),  # ML Schätzerarray
                      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 von  $X_1, \dots, X_n$ 
    mu_hat_ML[i,s] = mean(x)                      # Stichprobenmittel
  }
}
```

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



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

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 erwartungstreu 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

$$S_n(\theta) := \frac{d}{d\theta} \ell_n(\theta). \quad (38)$$

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

$$I_n(\theta) := -\frac{d^2}{d\theta^2} \ell_n(\theta). \quad (39)$$

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 $\hat{\theta}_n^{\text{ML}}$ sei ein ML-Schätzer von θ .

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

$$I_n \left(\hat{\theta}_n^{\text{ML}} \right) := - \frac{d^2}{d\theta^2} \ell_n \left(\hat{\theta}_n^{\text{ML}} \right), \quad (40)$$

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

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

$$J_n(\theta) := \mathbb{E}_\theta(I_n(\theta)). \quad (41)$$

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

$$I_n(\theta) = nI_1(\theta) \text{ und } J_n(\theta) = nJ_1(\theta). \quad (42)$$

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.
- Für einen Beweis verweisen wir auf den Appendix.

Theorem (Erwartungswert und Varianz der Scorefunktion)

Der Erwartungswert der Scorefunktion einer Zufallsvariable ist

$$\mathbb{E}_\theta(S(\theta)) = 0 \quad (43)$$

und die Varianz der Scorefunktion einer Zufallsvariable ist

$$\mathbb{V}_\theta(S(\theta)) = J(\theta). \quad (44)$$

Bemerkungen * Der Erwartungswert der Ableitung der Log-Likelihood-Funktion ist Null. * Die erwartete Fisher-Information ist gleich der Varianz der Scorefunktion. * Für einen Beweis verweisen wir auf den Appendix.

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

$$S_n :]0, 1[\rightarrow \mathbb{R}, \mu \mapsto S_n(\mu) := \frac{1}{\mu} \sum_{i=1}^n x_i - \frac{1}{1-\mu} \left(n - \sum_{i=1}^n x_i \right). \quad (45)$$

- Die Fisher-Information der Stichprobe ist

$$I_n :]0, 1[\rightarrow \mathbb{R}, \mu \mapsto I_n(\mu) := \frac{x}{\mu^2} + \frac{(1-x)^2}{1-\mu}. \quad (46)$$

- Die beobachtete Fisher-Information der Stichprobe ist

$$I_n :]0, 1[\rightarrow \mathbb{R}, \hat{\mu}_n^{\text{ML}} \mapsto I_n \left(\hat{\mu}_n^{\text{ML}} \right) := \frac{x}{\hat{\mu}_n^{\text{ML}2}} + \frac{(1-x)}{1-\hat{\mu}_n^{\text{ML}}}. \quad (47)$$

- Die erwartete Fisher-Information der Stichprobe ist

$$J_n :]0, 1[\rightarrow \mathbb{R}, \mu \mapsto J_n(\mu) := \frac{n}{\mu(1-\mu)}. \quad (48)$$

Für einen Beweis verweisen wir auf den Appendix.

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

$$S_n : \mathbb{R} \rightarrow \mathbb{R}, \mu \mapsto S_n(\mu) := \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu). \quad (49)$$

- Die Fisher-Information der Stichprobe ist

$$I_n : \mathbb{R} \rightarrow \mathbb{R}, \mu \mapsto I_n(\mu) := \frac{n}{\sigma^2}. \quad (50)$$

- Die beobachtete Fisher-Information der Stichprobe ist

$$I_n(\hat{\mu}_n^{\text{ML}}) = \frac{n}{\sigma^2}. \quad (51)$$

- Die erwartete Fisher-Information der Stichprobe ist

$$J_n : \mathbb{R} \rightarrow \mathbb{R}, \mu \mapsto J_n(\mu) := \frac{n}{\sigma^2}. \quad (52)$$

Für einen Beweis verweisen wir auf den Appendix.

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

$$S_n : \mathbb{R}_{>0} \rightarrow \mathbb{R}, \sigma^2 \mapsto S_n(\sigma^2) := -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (x_i - \mu)^2 \quad (53)$$

- die Fisher-Information der Stichprobe ist

$$I_n : \mathbb{R}_{>0} \rightarrow \mathbb{R}, \sigma^2 \mapsto I_n(\sigma^2) := \frac{1}{\sigma^6} \sum_{i=1}^n (x_i - \mu)^2 - \frac{n}{2\sigma^4} \quad (54)$$

- die beobachtete Fisher-Information der Stichprobe ist

$$I_n(\hat{\sigma}_n^{2\text{ML}}) = \frac{n}{2\hat{\sigma}_{\text{ML}}^4} \quad (55)$$

- die erwartete Fisher-Information der Stichprobe ist

$$J_n : \mathbb{R}_{>0} \rightarrow \mathbb{R}, \sigma^2 \mapsto J_n(\sigma^2) := \frac{n}{2\sigma^4}. \quad (56)$$

Für einen Beweis verweisen wir auf den Appendix.

Theorem (Cramér-Rao-Ungleichung)

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

$$\mathbb{V}_\theta(\hat{\tau}_n) \geq \frac{\left(\frac{d}{d\theta}\tau(\theta)\right)^2}{J(\theta)}. \quad (57)$$

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

$$\mathbb{V}_\theta(\hat{\theta}_n) \geq \frac{1}{J(\theta)}. \quad (58)$$

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

$$\begin{aligned} \frac{\mathbb{C}_\theta(S_n(\theta), \hat{\tau}_n)^2}{\mathbb{V}_\theta(S(\theta))\mathbb{V}_\theta(\hat{\tau}_n)} &\leq 1 \\ \Leftrightarrow \mathbb{V}_\theta(\hat{\tau}_n) &\geq \frac{\mathbb{C}_\theta(S(\theta), \hat{\tau}_n)^2}{J(\theta)}. \end{aligned} \quad (59)$$

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

$$\mathbb{C}_\theta(S(\theta), \hat{\tau}_n) = \frac{d}{d\theta} \tau(\theta) \quad (60)$$

wie unten gezeigt wird. Also gilt

$$\mathbb{V}_\theta(\hat{\tau}_n) \geq \frac{\left(\frac{d}{d\theta} \tau(\theta)\right)^2}{J(\theta)}. \quad (61)$$

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

$$\begin{aligned}\mathbb{C}_\theta(S(\theta), \hat{\tau}_n) &= \mathbb{E}_\theta(S(\theta)\hat{\tau}_n) - \mathbb{E}_\theta(S(\theta))\mathbb{E}_\theta(\hat{\tau}_n) = \mathbb{E}_\theta(S(\theta)\hat{\tau}_n) \\&= \int S(\theta) \hat{\tau}_n p_\theta(x) dx \\&= \int \frac{d}{d\theta} \ln L(\theta) \hat{\tau}_n p_\theta(x) dx \\&= \int \frac{\frac{d}{d\theta} L(\theta)}{L(\theta)} \hat{\tau}_n p_\theta(x) dx \\&= \int \frac{\frac{d}{d\theta} L(\theta)}{p_\theta(x)} \hat{\tau}_n p_\theta(x) dx & (62) \\&= \int \frac{d}{d\theta} L(\theta) \hat{\tau}_n dx \\&= \frac{d}{d\theta} \int L(\theta) \hat{\tau}_n dx \\&= \frac{d}{d\theta} \int \hat{\tau}_n p_\theta(x) dx = \frac{d}{d\theta} \mathbb{E}_\theta(\hat{\tau}_n) = \frac{d}{d\theta} \tau(\theta).\end{aligned}$$

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

$$\text{MQF}(\hat{\tau}_n) := \mathbb{E}_\theta \left((\hat{\tau}_n(X) - \tau(\theta))^2 \right). \quad (63)$$

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

$$\text{MQF}(\hat{\tau}_n) = \text{B}(\hat{\tau}_n)^2 + \mathbb{V}_\theta(\hat{\tau}_n). \quad (64)$$

Bemerkungen

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

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:

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

Grundbegriffe

Maximum-Likelihood Schätzer

Schätzereigenschaften bei endlichen Stichproben

Asymptotische Schätzereigenschaften

Eigenschaften von Maximum-Likelihood Schätzern

Selbstkontrollfragen

Vorbemerkungen zu Asymptotischen Schätzereigenschaften

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.

Vaart (1998) gibt eine ausführliche Einführung in die AS.

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

$$\lim_{n \rightarrow \infty} \mathbb{E}_\theta(\hat{\tau}_n(X)) = \tau(\theta) \text{ für alle } \theta \in \Theta. \quad (65)$$

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

$$\hat{\sigma}_n^{2\text{ML}} := \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2 \quad (66)$$

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

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

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

$$\lim_{n \rightarrow \infty} \mathbb{E}_{\mu, \sigma^2} \left(\hat{\sigma}_n^{2\text{ML}} \right) = \lim_{n \rightarrow \infty} \frac{n-1}{n} \sigma^2 = \sigma^2 \quad \lim_{n \rightarrow \infty} \frac{n-1}{n} = 1. \quad (67)$$

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(NA, length(n)*ns),
  ncol = length(n))

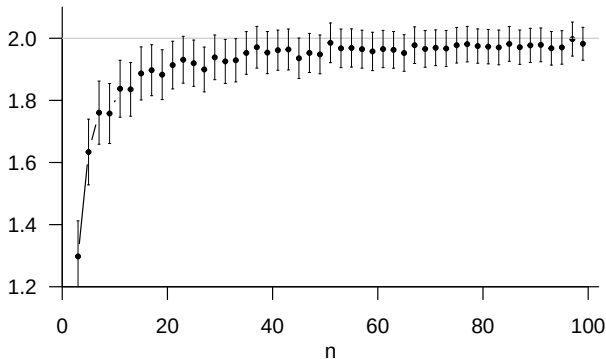
# Stichprobengroesseniterationen
for(i in seq_along(n)){

  # 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
```

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



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

$$\mathbb{P}_\theta (|\hat{\tau}_n(X) - \tau(\theta)| \geq \epsilon) \leq \frac{\mathbb{E}_\theta \left((\hat{\tau}_n(X) - \tau(\theta))^2 \right)}{\epsilon^2} \quad (68)$$

Grenzwertbildung ergibt dann

$$\lim_{n \rightarrow \infty} \mathbb{P}_\theta (|\hat{\tau}_n(X) - \tau(\theta)| \geq \epsilon) \leq \frac{1}{\epsilon^2} \lim_{n \rightarrow \infty} \mathbb{E}_\theta \left((\hat{\tau}_n(X) - \tau(\theta))^2 \right). \quad (69)$$

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

$$\lim_{n \rightarrow \infty} \mathbb{P}_\theta (|\hat{\tau}_n(X) - \tau(\theta)| \geq \epsilon) = 0. \quad (70)$$

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

$$\lim_{n \rightarrow \infty} B(\hat{\tau}_n) = 0 \text{ und } \lim_{n \rightarrow \infty} \mathbb{V}_\theta(\hat{\tau}_n) = 0 \quad (71)$$

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 $\lim_{n \rightarrow \infty} \text{MQF}(\hat{\theta}_n) = 0$. 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

$$B(\bar{X}_n) = 0 \text{ und } \lim_{n \rightarrow \infty} \mathbb{V}_\theta(\bar{X}_n) = \lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{V}(X_0) = 0. \quad (72)$$

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

```
# Modellformulierung
mu      = 1
sigsqr  = 2
n       = seq(1,1e3,by = 10)
eps     = c(0.15, 0.10, 0.05)
ne      = length(eps)
nn      = length(n)
ns      = 1e3
E       = array(rep(NA,n,ne,ns),
               dim = c(nn,ne,ns))

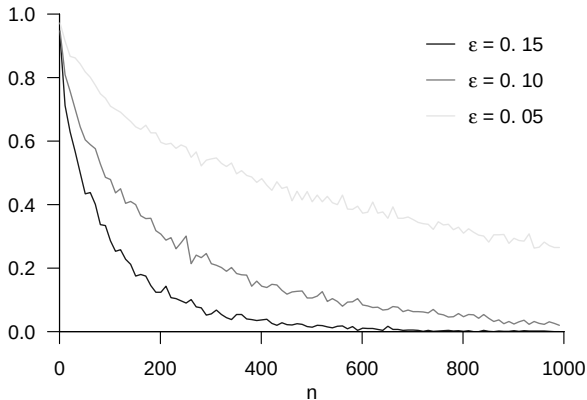
# Simulation
for(e in seq_along(eps)){
  for(i in seq_along(n)){
    for(s in 1:ns){

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

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

w.a.u. μ Wert
 # w.a.u. σ^2 Wert
 # Stichprobengroesse n
 # ϵ Werte
 # Anzahl ϵ Werte
 # Anzahl Stichprobengroessen
 # Anzahl Simulationen
 # Ereignisindikatorarray
 # ϵ Iterationen
 # n Iterationen
 # Simulationsiterationen
 # $|\bar{X}_n - \mu| \geq \epsilon$
 # $|\bar{X}_n - \mu| < \epsilon$

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



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

$$\hat{\theta}_n \stackrel{a}{\sim} N(\mu, \sigma^2). \quad (73)$$

Bemerkung

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

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

$$\hat{\theta}_n \stackrel{a}{\sim} N\left(\theta, J_n(\theta)^{-1}\right), \quad (74)$$

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.

Grundbegriffe

Maximum-Likelihood Schätzer

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 Parametschätzers

- 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

$$J_n(\mu) = \frac{n}{\mu(1 - \mu)}. \quad (75)$$

- Also gilt

$$\hat{\mu}_n^{\text{ML}} \overset{a}{\sim} N\left(\mu, \frac{\mu(1 - \mu)}{n}\right). \quad (76)$$

Eigenschaften von Maximum-Likelihood Schätzern

Simulation der asymptotischen Effizienz des Bernoulli ML Parameterschätzers

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

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

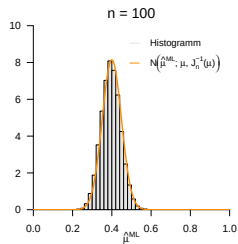
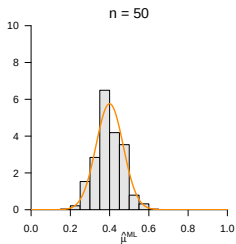
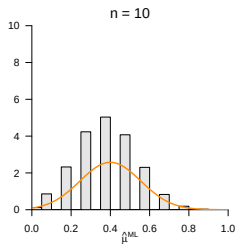
# Stichprobengrößeniterationen
for(i in seq_along(n_all)){

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

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

Eigenschaften von Maximum-Likelihood Schätzern

Simulation der asymptotischen Effizienz des Bernoulli ML Parameterschätzers



Selbstkontrollfragen

1. Definieren und erläutern Sie den Begriff des Parameterpunktschätzers.
2. Definieren Sie die Begriffe der Likelihood-Funktion und der Log-Likelihood Funktion.
3. Definieren Sie den Begriff des Maximum-Likelihood Schätzers.
4. Erläutern Sie das Vorgehen zur ML Schätzung für ein parametrisches statistisches Produktmodell.
5. Geben Sie den ML Schätzer für den Parameter μ des Bernoullimodells an.
6. Geben Sie den ML Schätzer für den Parameter μ des Normalverteilungmodells an.
7. Geben Sie den ML Schätzer für den Parameter σ^2 des Normalverteilungmodells an.
8. Definieren und erläutern Sie den Begriff der Erwartungstreue eines Schätzers.
9. Definieren Sie die Begriffe der Varianz und des Standardfehlers eines Schätzers.
10. Definieren Sie den Begriff der Scorefunktion einer Stichprobe.
11. Definieren Sie den Begriff der Fisher-Information einer Stichprobe.
12. Definieren Sie den Begriff der erwarteten Fisher-Information einer Stichprobe.
13. Geben Sie das Theorem zur Cramér-Rao-Ungleichung wieder.
14. Erläutern Sie den Begriff der Cramér-Rao-Schranke.
15. Definieren Sie den Begriff des MQFs eines Schätzers.
16. Geben Sie das Theorem zur Zerlegung des MQFs wieder.
17. Erläutern Sie den Begriff des asymptotischen Erwartungstreue eines Schätzers
18. Erläutern Sie den Begriff der Konsistenz eines Schätzers.
19. Erläutern Sie den Begriff der asymptotischen Normalität eines Schätzers.
20. Erläutern Sie den Begriff der asymptotischen Effizienz eines Schätzers.
21. Nennen Sie fünf Eigenschaften eines ML Schätzers.

Appendix

Beweis der Additivität der Fisher-Information

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

$$\begin{aligned} J_n(\theta) &= \mathbb{E}_\theta \left(-\frac{d^2}{d\theta^2} \ell_n(\theta) \right) \\ &= \mathbb{E}_\theta \left(-\frac{d^2}{d\theta^2} \ln \left(\prod_{i=1}^n p_\theta(x_i) \right) \right) \\ &= \mathbb{E}_\theta \left(-\frac{d^2}{d\theta^2} \sum_{i=1}^n \ln p_\theta(x_i) \right) \\ &= \mathbb{E}_\theta \left(-\frac{d^2}{d\theta^2} \sum_{i=1}^n \ln p_\theta(x_1) \right) \\ &= \mathbb{E}_\theta \left(-\frac{d^2}{d\theta^2} \ell_1(\theta) n \right) \\ &= n \mathbb{E}_\theta \left(-\frac{d^2}{d\theta^2} \ell_1(\theta) \right) \\ &= n J_n(\theta). \end{aligned} \tag{77}$$

Appendix

Beweis von Erwartungswert und Varianz der Scorefunktion

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

$$\begin{aligned}\mathbb{E}_\theta(S(\theta)) &= \int S(\theta) p_\theta(x) dx \\&= \int \frac{d}{d\theta} \ell(\theta) p_\theta(x) dx \\&= \int \frac{d}{d\theta} \ln L(\theta) p_\theta(x) dx \\&= \int \frac{1}{L(\theta)} \frac{d}{d\theta} L(\theta) p_\theta(x) dx & (78) \\&= \int \frac{1}{p_\theta(x)} \frac{d}{d\theta} L(\theta) p_\theta(x) dx \\&= \int \frac{d}{d\theta} L(\theta) dx \\&= \frac{d}{d\theta} \int p_\theta(x) dx = \frac{d}{d\theta} 1 = 0.\end{aligned}$$

Appendix

Mit der Definition der Varianz folgt dann sofort, dass $\mathbb{V}_\theta(S(\theta)) = \mathbb{E}_\theta \left(S(\theta)^2 \right)$ ist. Als nächstes zeigen wir, dass $J(\theta) = \mathbb{E}_\theta \left(S(\theta)^2 \right)$ und deshalb $\mathbb{V}_\theta(S(\theta)) = J(\theta)$ ist:

$$\begin{aligned} J(\theta) &= \mathbb{E}_\theta \left(-\frac{d^2}{d\theta^2} \ln L(\theta) \right) \\ &= \mathbb{E}_\theta \left(-\frac{d}{d\theta} \frac{\frac{d}{d\theta} L(\theta)}{L(\theta)} \right) \\ &= \mathbb{E}_\theta \left(-\frac{\frac{d^2}{d\theta^2} L(\theta) L(\theta) - \frac{d}{d\theta} L(\theta) \frac{d}{d\theta} L(\theta)}{L(\theta) L(\theta)} \right) \\ &= -\mathbb{E}_\theta \left(\frac{\frac{d^2}{d\theta^2} L(\theta)}{L(\theta)} \right) + \mathbb{E}_\theta \left(\frac{\left(\frac{d}{d\theta} L(\theta) \right)^2}{(L(\theta))^2} \right) \\ &= -\int \frac{\frac{d^2}{d\theta^2} L(\theta)}{L(\theta)} p_\theta(x) dx + \int \frac{\left(\frac{d}{d\theta} L(\theta) \right)^2}{(L(\theta))^2} p_\theta(x) dx \\ &= -\frac{d^2}{d\theta^2} \int p_\theta(x) dx + \int \left(\frac{1}{L(\theta)} \frac{d}{d\theta} L(\theta) \right)^2 p_\theta(x) dx \\ &= -\frac{d^2}{d\theta^2} 1 + \int \left(\frac{d}{d\theta} \ln L(\theta) \right)^2 p_\theta(x) dx = \mathbb{E}_\theta \left(S(\theta)^2 \right). \end{aligned} \tag{79}$$

Beweis der Scorefunktion und Fisher-Information für den Bernoulli Parameter

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

$$\begin{aligned} I(\mu) &:= -\frac{d^2}{d\mu^2} \ell_1(\mu) \\ &= -\frac{d^2}{d\mu^2} \ln p_\mu(x) \\ &= -\frac{d^2}{d\mu^2} (x \ln \mu + (1-x) \ln(1-\mu)) \\ &= -\frac{d}{d\mu} \left(\frac{d}{d\mu} (x \ln \mu + (1-x) \ln(1-\mu)) \right) \\ &= -\frac{d}{d\mu} \left(\frac{x}{\mu} + \frac{(1-x)}{1-\mu} \right) \\ &= -\left(-\frac{x}{\mu^2} - \frac{(1-x)}{1-\mu} \right) \\ &= \frac{x}{\mu^2} + \frac{(1-x)^2}{1-\mu} . \end{aligned} \tag{80}$$

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

$$\begin{aligned} J(\mu) &= \mathbb{E}_{\mu}(I(\mu)) \\ &= \mathbb{E}_{\mu} \left(\frac{X}{\mu^2} + \frac{(1-X)^2}{1-\mu} \right) \\ &= \frac{\mathbb{E}_{\mu}(X)}{\mu^2} + \frac{(1 - \mathbb{E}_{\mu}(X))^2}{1-\mu} \\ &= \frac{\mu}{\mu^2} + \frac{(1-\mu)^2}{1-\mu} \\ &= \frac{1}{\mu(1-\mu)}. \end{aligned} \tag{81}$$

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

$$I_n(\mu) = \frac{nx}{\mu^2} + \frac{n(1-x)^2}{1-\mu}, \text{ und } J_n(\mu) = \frac{n}{\mu(1-\mu)}. \tag{82}$$

Appendix

Beweis der Scorefunktion und Fisher-Information für den Erwartungswertparameter der Normalverteilung

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

$$\ell_n : \mathbb{R} \rightarrow \mathbb{R}, \mu \mapsto \ell_n(\mu) := -\frac{n}{2} \ln 2\pi - \frac{n}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2. \quad (83)$$

gegeben ist. Damit ergibt sich die Scorefunktion als

$$S_n(\mu) = \frac{d}{d\mu} \ell_n(\mu) = \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu) \quad (84)$$

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

$$I_n(\mu) = -\frac{d^2}{d\mu^2} \ell_n(\mu) = -\frac{d}{d\mu} S_n(\mu) = -\frac{1}{\sigma^2} \frac{d}{d\mu} \left(\sum_{i=1}^n x_i - n\mu \right) = \frac{n}{\sigma^2}. \quad (85)$$

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

$$J_n(\mu) = \mathbb{E}_\mu(I_n(\mu)) = \mathbb{E}_\mu \left(\frac{n}{\sigma^2} \right) = \frac{n}{\sigma^2}. \quad (86)$$

Beweis der Scorefunktion und Fisher-Information für den Varianzparameter der Normalverteilung

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

$$\ell_n : \mathbb{R}_{>0} \rightarrow \mathbb{R}, \sigma^2 \mapsto \ell_n(\sigma^2) := -\frac{n}{2} \ln 2\pi - \frac{n}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2. \quad (87)$$

gegeben ist. Die Scorefunktion ergibt sich also als

$$S_n(\sigma^2) = \frac{d}{d\sigma^2} \ell_n(\sigma^2) = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (x_i - \mu)^2. \quad (88)$$

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

$$\begin{aligned} I_n(\sigma^2) &= -\frac{d}{d\sigma^2} S_n(\sigma^2) = -\left(\frac{n}{2\sigma^4} - \frac{1}{\sigma^6} \sum_{i=1}^n (x_i - \mu)^2 \right) \\ &= \frac{1}{\sigma^6} \sum_{i=1}^n (x_i - \mu)^2 - \frac{n}{2\sigma^4}. \end{aligned} \quad (89)$$

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

$$\begin{aligned} I_n(\hat{\sigma}_n^{2\text{ML}}) &= \frac{\sum_{i=1}^n (x_i - \mu)^2}{\left(\hat{\sigma}_n^{2\text{ML}}\right)^3} - \frac{n}{2 \left(\hat{\sigma}_n^{2\text{ML}}\right)^2} \\ &= \frac{\sum_{i=1}^n (x_i - \mu)^2}{\frac{1}{n^3} \left(\sum_{i=1}^n (x_i - \mu)^2\right)^3} - \frac{n}{2 \left(\hat{\sigma}_n^{2\text{ML}}\right)^2} \\ &= \frac{1}{\frac{1}{n^3} \left(\sum_{i=1}^n (x_i - \mu)^2\right)^2} - \frac{n}{2 \left(\hat{\sigma}_n^{2\text{ML}}\right)^2} \\ &= \frac{n}{\left(\hat{\sigma}_n^{2\text{ML}}\right)^2} - \frac{n}{2 \left(\hat{\sigma}_n^{2\text{ML}}\right)^2} \\ &= \frac{n}{2 \left(\hat{\sigma}_n^{2\text{ML}}\right)^2} \\ &= \frac{n}{2 \hat{\sigma}_n^{4\text{ML}}} . \end{aligned} \tag{90}$$

Die erwartete Fisher-Information ergibt sich als

$$\begin{aligned} J_n(\sigma^2) &= \mathbb{E}_{\sigma^2}(I_n(\sigma^2)) \\ &= \mathbb{E}_{\sigma^2} \left(\frac{1}{\sigma^6} \sum_{i=1}^n (x_i - \mu)^2 - \frac{n}{2\sigma^4} \right) \\ &= \frac{1}{\sigma^6} \sum_{i=1}^n \mathbb{E}_{\sigma^2} \left((x_i - \mu)^2 \right) - \frac{n}{2\sigma^4} \\ &= \frac{1}{\sigma^6} \sum_{i=1}^n \sigma^2 - \frac{n}{2\sigma^4} \\ &= \frac{n\sigma^2}{\sigma^6} - \frac{n}{2\sigma^4} \\ &= \frac{n}{\sigma^4} - \frac{n}{2\sigma^4} \\ &= \frac{n}{2\sigma^4}. \end{aligned} \tag{91}$$

References

- Held, Leonhard, and Daniel Sabanés Bové. 2014. *Applied Statistical Inference*. Berlin, Heidelberg: Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-642-37887-4>.
- Vaart, A. W. van der. 1998. *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge, UK ; New York, NY, USA: Cambridge University Press.