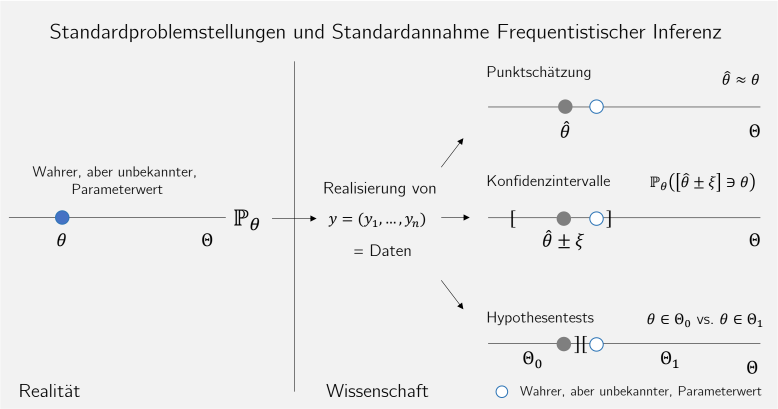
Wahrscheinlichkeitstheorie und Frequentistische Inferenz

BSc Psychologie WiSe 2024/25

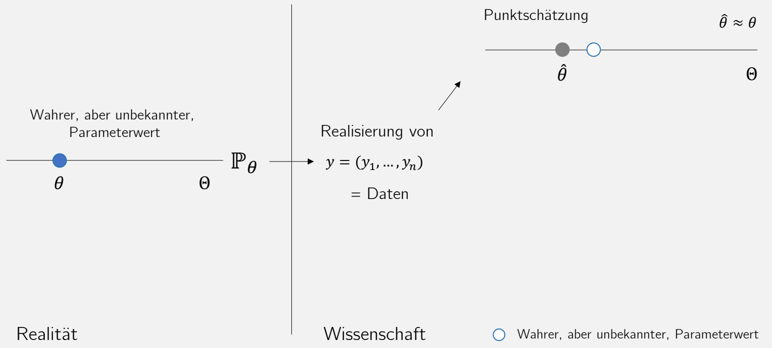
Prof. Dr. Dirk Ostwald

(9) Punktschätzung



Standardproblemstellungen und Standardannahme Frequentistischer Inferenz
Wahrer, aber unbekannter, Parameterwert
θ
Θ
$\mathbb{P}_\theta$
Realisierung von
$y = (y_1, \dots, y_n)$
= Daten
Punktschätzung
$\hat{\theta} \approx \theta$
$\hat{\theta}$
Θ
Konfidenzintervalle
$\mathbb{P}_\theta([\hat{\theta} \pm \xi] \ni \theta)$
$\hat{\theta} \pm \xi$
Θ
Hypothesentests
$\theta \in \Theta_0$ vs. $\theta \in \Theta_1$
Θ_0
Θ_1
Θ
Realität
Wissenschaft
Wahrer, aber unbekannter, Parameterwert

Standardproblemstellungen und Standardannahme Frequentistischer Inferenz



Standardannahme der Frequentistischen Inferenz

$\mathcal{M}$ sei ein Frequentistisches Inferenzmodell mit Zufallsvektor y.

Es wird angenommen, dass ein vorliegender Datensatz eine der möglichen Realisierungen von y ist.

Aus Frequentistischer Sicht kann man eine Studie unendlich oft wiederholen und zu jedem Datensatz Schätzer oder Statistiken auswerten, z.B. das Stichprobenmittel:

Datensatz (1) : $y^{(1)} = \left(y_1^{(1)}, y_2^{(1)}, ..., y_n^{(1)}\right)$ mit $\bar{y}^{(1)} = \frac{1}{n}\sum_{i=1}^n y_i^{(1)}$

Datensatz (2) : $y^{(2)} = \left(y_1^{(2)}, y_2^{(2)}, ..., y_n^{(2)}\right)$ mit $\bar{y}^{(2)} = \frac{1}{n}\sum_{i=1}^n y_i^{(2)}$

Datensatz (3) : $y^{(3)} = \left(y_1^{(3)}, y_2^{(3)}, ..., y_n^{(3)}\right)$ mit $\bar{y}^{(3)} = \frac{1}{n}\sum_{i=1}^n y_i^{(3)}$

Datensatz (4) : $y^{(4)} = \left(y_1^{(4)}, y_2^{(4)}, ..., y_n^{(4)}\right)$ mit $\bar{y}^{(4)} = \frac{1}{n}\sum_{i=1}^n y_i^{(4)}$

Datensatz (5) : $y^{(5)} = ...$

Um die Qualität ihrer Methoden zu beurteilen betrachtet die Frequentistische Inferenz deshalb die Wahrscheinlichkeitsverteilungen von Schätzern und Statistiken. Was zum Beispiel ist die Verteilung der $\bar{y}^{(1)}$, $\bar{y}^{(2)}$, $\bar{y}^{(3)}$, $\bar{y}^{(4)}$, ... also die Verteilung der Zufallsvariable $\bar{y}_n$?

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

Maximum-Likelihood-Schätzer

Schätzereigenschaften bei endlichen Stichproben

Asymptotische Schätzereigenschaften

Eigenschaften von Maximum-Likelihood-Schätzern

Selbstkontrollfragen

Definition (Parameterpunktschätzer)

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

Bemerkungen

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

Maximum-Likelihood-Schätzer

Prinzipien zur Gewinnung von Parameterschätzern

Die Definition eines Parameterschätzers macht keine Aussage darüber, wie man Parameterschätzer findet. Zur Gewinnung von Parameterschätzern in Frequentistischen Inferenzmodellen 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)

Perse 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*.

Maximum-Likelihood-Schätzer

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

$\mathcal{M}$ sei ein parametrisches Produktmodell mit WMF oder WDF p_θ. Dann ist die *Likelihood-Funktion* definiert als

$$L : \Theta \to [0, \infty[, \theta \mapsto L(\theta) := \prod_{i=1}^{n} p_\theta(y_i) \tag{1}$$

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

$$\ell : \Theta \to \mathbb{R}, \theta \mapsto \ell(\theta) := \ln L(\theta). \tag{2}$$

Bemerkungen

- L ist eine Funktion des Parameters eines Frequentistischen Inferenzmodells.
- Werte von L sind die gemeinsamen Wahrscheinlichkeitsmassen bzw. -dichten von Datenwerten $y_1, ..., y_n$.
- Generell gibt es keinen Grund anzunehmen, dass L 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 Produktmodell mit Parameter $\theta \in \Theta$. Ein *Maximum-Likelihood-Schätzer* von θ ist definiert als

$$\hat{\theta}^{\mathsf{ML}} : \mathcal{Y} \to \Theta, y \mapsto \hat{\theta}^{\mathsf{ML}}(y) := \arg\max_{\theta \in \Theta} L(\theta) = \arg\max_{\theta \in \Theta} \ell(\theta). \tag{3}$$

Bemerkungen

- $L(\theta) := \prod_{i=1}^{n} p_\theta(y_i)$ hängt von $y := (y_1, ..., y_n)$ ab, also hängt auch $\hat{\theta}^{\mathsf{ML}}(y)$ von y ab.
- Weil $\ln$ monoton steigend ist, entspricht eine Maximumstelle von ℓ einer Maximumstelle von L.
- Das Arbeiten mit der Log-Likelihood-Funktion ist oft einfacher als mit der Likelihood-Funktion.
- Multiplikation von L mit einer positiven Konstante, die nicht von θ abhängt, verändert einen Maximum-Likelihood-Schätzer nicht, konstante additive Terme in der Log-Likelihood können also vernachlässigt werden.
- Maximum-Likelihood Schätzung ist ein Optimierungsproblem

Maximum-Likelihood-Schätzer

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 analytischen Optimierung in klassischen Beispielen
- Methoden der numerischen Optimierung im Anwendungskontext

Theorem (Maximum-Likelihood-Schätzer des Bernoullimodells)

Es sei $y_1, ..., y_n \sim \text{Bern}(\mu)$ die Stichprobe des Bernoullimodells. Dann ist

$$\hat{\mu}^{\text{ML}} : \{0,1\}^n \to [0,1], y \mapsto \hat{\mu}^{\text{ML}}(y) := \frac{1}{n}\sum_{i=1}^{n} y_i \tag{4}$$

ein Maximum-Likelihood-Schätzer von μ.

Bemerkung

- $\hat{\mu}^{\text{ML}}$ ist offenbar mit dem Stichprobenmittel identisch.

Maximum-Likelihood-Schätzer

Beweis

Wir formulieren zunächst die Log-Likelihood-Funktion. Für die Likelihood-Funktion gilt

$$L :]0,1[\to]0,1[, \mu \mapsto L(\mu) := \prod_{i=1}^{n} \mu^{y_i} (1-\mu)^{1-y_i} = \mu^{\sum_{i=1}^{n} y_i} (1-\mu)^{n - \sum_{i=1}^{n} y_i}. \quad (5)$$

Logarithmieren ergibt

$$\ell :]0,1[\to \mathbb{R}, \mu \mapsto \ell(\mu) = \ln \mu \sum_{i=1}^{n} y_i + \ln(1-\mu)\left(n - \sum_{i=1}^{n} y_i\right). \quad (6)$$

Wir werten dann die Ableitung der Log-Likelihood-Funktion aus. Es gilt

$$\begin{aligned}
\frac{d}{d\mu}\ell(\mu) &= \frac{d}{d\mu}\left(\ln \mu \sum_{i=1}^{n} y_i + \ln(1-\mu)\left(n - \sum_{i=1}^{n} y_i\right)\right) \\
&= \frac{d}{d\mu}\ln \mu \sum_{i=1}^{n} y_i + \frac{d}{d\mu}\ln(1-\mu)\left(n - \sum_{i=1}^{n} y_i\right) \quad (7) \\
&= \frac{1}{\mu}\sum_{i=1}^{n} y_i - \frac{1}{1-\mu}\left(n - \sum_{i=1}^{n} y_i\right).
\end{aligned}$$

Nullsetzen ergibt dann folgende *Maximum-Likelihood-Gleichung* als notwendige Bedingung für einen Maximum-Likelihood-Schätzer des Bernoullimodelparameters:

$$\frac{1}{\hat{\mu}^{\mathsf{ML}}}\sum_{i=1}^{n} y_i - \frac{1}{1-\hat{\mu}^{\mathsf{ML}}}\left(n - \sum_{i=1}^{n} y_i\right) = 0. \quad (8)$$

Beweis (fortgeführt)

Auflösen der Maximum-Likelihood-Gleichung nach $\hat{\mu}^{\mathsf{ML}}$ ergibt dann

$$
\begin{aligned}
\frac{1}{\hat{\mu}^{\mathsf{ML}}}\sum_{i=1}^n y_i - \frac{1}{1-\hat{\mu}^{\mathsf{ML}}}\left(n - \sum_{i=1}^n y_i\right) & = 0 \\
\Leftrightarrow \hat{\mu}^{\mathsf{ML}}(1-\hat{\mu}^{\mathsf{ML}})\left(\frac{1}{\hat{\mu}^{\mathsf{ML}}}\sum_{i=1}^n y_i - \frac{1}{1-\hat{\mu}^{\mathsf{ML}}}\left(n - \sum_{i=1}^n y_i\right)\right) & = 0 \\
\Leftrightarrow \sum_{i=1}^n y_i - \hat{\mu}^{\mathsf{ML}}\sum_{i=1}^n y_i - n\hat{\mu}^{\mathsf{ML}} + \hat{\mu}^{\mathsf{ML}}\sum_{i=1}^n y_i & = 0 \\
\Leftrightarrow n\hat{\mu}^{\mathsf{ML}} & = \sum_{i=1}^n y_i \\
\Leftrightarrow \hat{\mu}^{\mathsf{ML}} & = \frac{1}{n}\sum_{i=1}^n y_i .
\end{aligned}
\tag{9}
$$

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

$\square$

Theorem (Maximum-Likelihood-Schätzer des Normalverteilungsmodells)

Es sei $y_1, ..., y_n \sim N\left(\mu, \sigma^2\right)$ die Stichprobe des Normalverteilungsmodells. Dann sind

$$\hat{\mu}^{\mathsf{ML}} : \mathbb{R}^n \to \mathbb{R}, y \mapsto \hat{\mu}^{\mathsf{ML}}(y) := \frac{1}{n}\sum_{i=1}^{n} y_i \tag{10}$$

und

$$\hat{\sigma}^{2^{\mathsf{ML}}} : \mathbb{R}^n \to \mathbb{R}_{\geq 0}, y \mapsto \hat{\sigma}_n^{2^{\mathsf{ML}}}(y) := \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{\mu}^{\mathsf{ML}}\right)^2. \tag{11}$$

Maximum-Likelihood-Schätzer von μ und σ^2.

Bemerkungen

- $\hat{\mu}^{\mathsf{ML}}$ ist identisch mit dem Stichprobenmittel $\bar{y}$.
- $\hat{\sigma}^{2^{\mathsf{ML}}}$ ist nicht identisch mit der Stichprobenvarianz S^2.

Maximum-Likelihood-Schätzer

Beweis

Wir formulieren zunächst die Log-Likelihood-Funktion. Für die Likelihood-Funktion ergibt sich

$$
\begin{aligned}
L : \mathbb{R} \times \mathbb{R}_{>0} \to \mathbb{R}_{>0}, (\mu, \sigma^2) \mapsto L(\mu, \sigma^2) &:= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(y_i - \mu)^2\right) \\
&= \left(2\pi\sigma^2\right)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mu)^2\right).
\end{aligned}
\tag{12}
$$

Logarithmieren ergibt dann

$$
\ell : \mathbb{R} \times \mathbb{R}_{>0} \to \mathbb{R}, (\mu, \sigma^2) \mapsto \ell(\mu, \sigma^2) = -\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mu)^2. \tag{13}
$$

Die Bestimmung der partiellen Ableitung der Log-Likelihood-Funktion hinsichtlich μ ergibt dann

$$
\frac{\partial}{\partial\mu}\ell(\mu, \sigma^2) = -\frac{\partial}{\partial\mu}\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mu)^2 = -\frac{1}{2\sigma^2}\sum_{i=1}^{n}\frac{\partial}{\partial\mu}(y_i - \mu)^2 = \frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i - \mu) \tag{14}
$$

und Bestimmung der partiellen Ableitung der Log-Likelihood-Funktion hinsichtlich σ^2 ergibt

$$
\frac{\partial}{\partial\sigma^2}\ell(\mu, \sigma^2) = -\frac{n}{2}\frac{\partial}{\partial\sigma^2}\ln\sigma^2 - \frac{\partial}{\partial\sigma^2}\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mu)^2 = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(y_i - \mu)^2. \tag{15}
$$

Das System der Maximum-Likelihood Gleichungen hat also die Form

$$
\sum_{i=1}^{n}(y_i - \hat{\mu}^{\mathsf{ML}}) = 0 \text{ und } -\frac{n}{2\hat{\sigma}^{2\mathsf{ML}}} + \frac{1}{2\hat{\sigma}^{4\mathsf{ML}}}\sum_{i=1}^{n}(y_i - \mu)^2 = 0 \tag{16}
$$

Maximum-Likelihood-Schätzer

Beweis (fortgeführt)

Lösen des Systems der Maximum-Likelihood Gleichungen ergibt dann

$$\sum_{i=1}^{n}(y_i - \hat{\mu}^{\mathsf{ML}}) = 0 \Leftrightarrow \sum_{i=1}^{n} y_i = n\hat{\mu}^{\mathsf{ML}} \Leftrightarrow \hat{\mu}^{\mathsf{ML}} = \frac{1}{n}\sum_{i=1}^{n} y_i. \tag{17}$$

Also ist

$$\hat{\mu}^{\mathsf{ML}} = \frac{1}{n}\sum_{i=1}^{n} y_i \tag{18}$$

ein potentieller Maximum-Likelihood-Schätzer von μ. Einsetzen in die zweite Maximum-Likelihood Gleichung ergibt dann

$$\begin{aligned}
-\frac{n}{2\hat{\sigma}^{2^{\mathsf{ML}}}} + \frac{1}{2\hat{\sigma}^{4^{\mathsf{ML}}}}\sum_{i=1}^{n}(y_i - \hat{\mu}^{\mathsf{ML}})^2 = 0 \\
\Leftrightarrow -n\hat{\sigma}^{2^{\mathsf{ML}}} + \sum_{i=1}^{n}(y_i - \hat{\mu}^{\mathsf{ML}})^2 = 0 \\
\Leftrightarrow \hat{\sigma}^{2^{\mathsf{ML}}} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{\mu}^{\mathsf{ML}})^2.
\end{aligned} \tag{19}$$

Also ist

$$\hat{\sigma}^{2^{\mathsf{ML}}} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{\mu}^{\mathsf{ML}})^2 \tag{20}$$

ein potentieller Maximum-Likelihood-Schätzer von σ^2. Beide potentiellen Maximum-Likelihood-Schätzer können durch Betrachten der zweiten Ableitung von ℓ verifiziert werden, worauf hier verzichtet werden soll.

Beispiel | Evidenzbasierte Evaluation von Psychotherapie bei Depression

Pre-BDI

n = 12

BDI-II Fragebogen

Name | Alter | Geschlecht m / w | Datum

Anleitung: Dieser Fragebogen enthält 21 Gruppen von Aussagen. Bitte lesen Sie jede dieser Gruppen von Aussagen sorgfältig durch und suchen Sie sich dann in jeder Gruppe **eine Aussage** heraus, die am besten beschreibt, wie Sie sich **in den letzten zwei Wochen, einschließlich heute, gefühlt haben**. Kreuzen Sie die Zahl neben der Aussage an, die Sie sich herausgesucht haben (0, 1, 2 oder 3). Falls in einer Gruppe mehrere Aussagen gleichermaßen auf Sie zutreffen, kreuzen Sie die Aussage mit der höheren Zahl an. Achten Sie bitte darauf, dass Sie in jeder Gruppe nicht mehr als eine Aussage ankreuzen, das gilt auch für Gruppe 16 (Veränderungen der Schlafgewohnheiten) oder Gruppe 18 (Veränderungen des Appetits).

1.) Traurigkeit
0 Ich bin nicht traurig.
1 Ich bin oft traurig.
2 Ich bin ständig traurig.
3 Ich bin so traurig oder unglücklich, dass ich es nicht aushalte.

2.) Pessimismus
0 Ich sehe nicht mutlos in die Zukunft.
1 Ich sehe mutloser in die Zukunft als sonst.
2 Ich bin mutlos und erwarte nicht, dass meine Situation besser wird.
3 Ich glaube, dass meine Zukunft hoffnungslos ist und nur noch schlechter wird.

3.) Versagensgefühle
0 Ich fühle mich nicht als Versager.
1 Ich habe häufiger Versagensgefühle.
2 Wenn ich zurückblicke, sehe ich eine Menge Fehlschläge.
3 Ich habe das Gefühl, als Mensch ein völliger Versager zu sein.

4.) Verlust von Freude
0 Ich kann die Dinge genauso gut genießen wie früher.
1 Ich kann die Dinge nicht mehr so genießen wie früher.
2 Dinge, die mir früher Freude gemacht haben, kann ich kaum mehr genießen.
3 Dinge, die mir früher Freude gemacht haben, kann ich überhaupt nicht mehr genießen.

5.) Schuldgefühle
0 Ich habe keine besonderen Schuldgefühle.
1 Ich habe oft Schuldgefühle wegen Dingen, die ich getan habe oder hätte tun sollen.
2 Ich habe die meiste Zeit Schuldgefühle.
3 Ich habe ständig Schuldgefühle.

6.) Bestrafungsgefühle
0 Ich habe nicht das Gefühl, für etwas bestraft zu sein.
1 Ich habe das Gefühl, vielleicht bestraft zu werden.
2 Ich erwarte, bestraft zu werden.
3 Ich habe das Gefühl, bestraft zu sein.

7.) Selbstablehnung
0 Ich halte von mir genauso viel wie immer.
1 Ich habe Vertrauen in mich verloren.
2 Ich bin von mir enttäuscht.
3 Ich lehne mich völlig ab.

8.) Selbstvorwürfe
0 Ich kritisiere oder tadle mich nicht mehr als sonst.
1 Ich bin mir gegenüber kritischer als sonst.
2 Ich kritisiere mich für all meine Mängel.
3 Ich gebe mir die Schuld für alles Schlimme, was passiert.

9.) Selbstmordgedanken
0 Ich denke nicht daran, mir etwas anzutun.
1 Ich denke manchmal an Selbstmord, aber ich würde es nicht tun.
2 Ich möchte mich am liebsten umbringen.
3 Ich würde mich umbringen, wenn ich die Gelegenheit dazu hätte.

10.) Weinen
0 Ich weine nicht öfter als früher.
1 Ich weine jetzt mehr als früher.
2 Ich weine beim geringsten Anlass.
3 Ich möchte gern weinen, aber ich kann nicht.

PEARSON PsychCorp Summe Seite 1:

Post-BDI

⇒ Pre-Post BDI Score Reduktion

Beispiel | Evidenzbasierte Evaluation von Psychotherapie bei Depression

i	BDI.Reduktion
1	-1
2	3
3	-2
4	9
5	3
6	-2
7	4
8	5
9	5
10	1
11	9
12	4

Maximum-Likelihood-Schätzer

Beispiel | Evidenzbasierte Evaluation von Psychotherapie bei Depression

Für die Pre-Post BDI Score Reduktion y_i der $i$ten von n Patient:innen legen wir das Modell

$$y_i = \mu + \varepsilon_i \text{ mit } \varepsilon_i \sim N(0, \sigma^2) \text{ u.i.v. für } i = 1, ..., n \tag{21}$$

zugrunde. Dabei wird die Pre-Post BDI Reduktion y_i der $i$ten Patient:in also mithilfe einer über die Gruppe von Patient:innen identischen Pre-Post BDI Score Reduktion $\mu \in \mathbb{R}$ und einer Patient:innen-spezifischen normalverteilten Pre-Post BDI Score Reduktionsabweichung ε_i erklärt.

Wie gezeigt ist dieses Modell äquivalent zum Normalverteilungsmodell

$$y_1, ..., y_n \sim N(\mu, \sigma^2). \tag{22}$$

Die Standardproblemstellungen der Frequentistischen Inferenz führen dann auf folgende Fragen:

(1) Was sind sinnvolle Tipps für die wahren, aber unbekannten, Parameterwerte μ und σ^2?
(2) Wie gelingt im Sinne einer Intervallschätzung eine möglichst sichere Schätzung von μ?
(3) Entscheiden wir uns sinnvollerweise für die Hypothese, dass gilt $\mu \neq 0$?

Maximum-Likelihood-Schätzer

Beispiel | Evidenzbasierte Evaluation von Psychotherapie bei Depression

```
# Einlesen und Auswahl der Daten
fname = "./9_Daten/9_Punktschätzung.csv"
D     = read.table(fname, sep = ",", header = T)
y     = D$BDI.Reduktion

# ML Schätzung des Erwartungswertparameters
mu_hat    = mean(y)                            # mean(y) berechnet das Stichprobenmittel
print(mu_hat)                                  # Ausgabe
```

```
[1] 3.166667
```

```
# ML Schätzung des Varianzparameters
n          = length(y)                         # Anzahl der Datenpunkte
sigsqr_hat  = ((n-1)/n)*var(y)                 # var(y) berechnet die Stichprobenvarianz
print(sigsqr_hat)                              # Ausgabe
```

```
[1] 12.63889
```

Basierend auf dem Prinzip der Maximum-Likelihood Schätzung sind also

$$\hat{\mu}^{\text{ML}} = 3.17, \text{ und } \hat{\sigma}^{2^{\text{ML}}} = 12.6 \tag{23}$$

sinnvolle Tipps für μ und σ^2 basierend auf den vorliegenden 12 Datenpunkten.

Maximum-Likelihood-Schätzer

Schätzereigenschaften bei endlichen Stichproben

Asymptotische Schätzereigenschaften

Eigenschaften von Maximum-Likelihood-Schätzern

Selbstkontrollfragen

Vorbemerkungen zu Frequentistischen Schätzereigenschaften

Wir gehen von einem parametrischem Produktmodell $\mathcal{M} := \{\mathcal{Y}, \mathcal{A}, \{\mathbb{P}_\theta | \theta \in \Theta\}\}$ mit n-dimensionalen Stichprobenraum (z.B. $\mathcal{Y} := \mathbb{R}^n$), d-dimensionalen Parameteraum $\Theta \subset \mathbb{R}^d$ und gegebener WMF oder WDF p_θ für alle $\theta \in \Theta$ aus. $y := (y_1, ..., y_n)$ bezeichnet die zu $\mathcal{M}$ gehörende Stichprobe unabhängig und identisch verteilter Zufallsvariablen, es gilt also $y_1 \sim \mathbb{P}_\theta$ und $y_i \sim \mathbb{P}_\theta$ für alle $i = 1, ..., n$.

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

$$\tau : \Theta \to \Sigma,\ \theta \mapsto \tau(\theta) \text{ mit } \tau(\theta) := \mathbb{E}_\theta(y_1), \tau(\theta) := \mathbb{V}_\theta(y_1), \text{ und } \tau(\theta) := \mathbb{S}_\theta(y_1) \tag{24}$$

respektive, sowie Parameterschätzer, also Schätzer für

$$\tau : \Theta \to \Sigma, \tau(\theta) := \theta. \tag{25}$$

In der Folge führen wir *Frequentistische Schätzereigenschaften* ein. Frequentistische Schätzereigenschaften betrachten die Verteilung der Schätzwerte in Abhängigkeit von der Verteilung der Stichprobenwerte. Weil die Stichprobenwerte zufällig sind, sind auch die Schätzwerte zufällig; ein Schätzer $\hat{\tau}$ ist also immer eine eine Zufallsvariable. Wir unterscheiden zwischen *Schätzereigenschaften bei endlichen Stichproben*, das heißt Eigenschaften von $\hat{\tau}$ für ein festes $n \in \mathbb{N}$ und *Asymptotischen Schätzereigenschaften*, das heißt Eigenschaften von $\hat{\tau}$ für unendlich groß werdende Stichproben mit $n \to \infty$.

Schätzereigenschaften bei endlichen Stichproben

Wir betrachten in der Folge zwei Aspekte von Schätzereigenschaften bei endlichen Stichproben:

(1) Erwartungstreue
(2) Varianz und Standardfehler

Intuitiv haben diese folgende Bedeutungen: Ein Schätzer $\hat{\tau}$ heißt *erwartungstreu*, wenn sein Erwartungswert dem wahren, aber unbekannten, Wert $\tau(\theta)$ für alle $\theta \in \Theta$ gleicht. Die *Varianz* eines Schätzers $\hat{\tau}$ ist die Varianz der Zufallsvariable $\hat{\tau}(y)$. Der *Standardfehler* eines Schätzers $\hat{\tau}$ ist die Standardabweichung der Zufallsvariable $\hat{\tau}(y)$.

Für folgende Schätzereigenschaften bei endlichen Stichproben verweisen wir auf das Vorlesungsskript:

(1) Mittlerer Quadratischer Fehler
(2) Cramér-Rao Ungleichung

Intuitiv haben diese folgende Bedeutungen: Der *mittlere quadratische Fehler* von $\hat{\tau}$ ist der Erwartungswert der quadrierten Abweichung von $\hat{\tau}(y)$ von $\tau(\theta)$ über Stichproben vom Umfang n. 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.

Schätzereigenschaften bei endlichen Stichproben

Definition (Fehler, Systematischer Fehler, und Erwartungstreue)

$y = (y_1, ..., y_n)$ sei die Stichprobe eines Frequentistischen Inferenzmodells und $\hat{\tau}$ sei ein Schätzer für τ.

- Der *Fehler* von $\hat{\tau}$ ist definiert als
$$\hat{\tau}(y) - \tau(\theta). \tag{26}$$
- Der *systematische Fehler (Bias)* von $\hat{\tau}$ ist definiert als
$$\mathrm{B}(\hat{\tau}) := \mathbb{E}_\theta(\hat{\tau}(y)) - \tau(\theta). \tag{27}$$
- $\hat{\tau}$ heißt *erwartungstreu (unbiased)*, wenn
$$\mathrm{B}(\hat{\tau}) = 0 \Leftrightarrow \mathbb{E}_\theta(\hat{\tau}(y)) = \tau(\theta) \text{ für alle } \theta \in \Theta \text{ und alle } n \in \mathbb{N}. \tag{28}$$
Andernfalls heißt $\hat{\tau}$ *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 $\mathbb{E}_\theta(\hat{\theta}) = \theta$.

Theorem (Erwartungstreue von Stichprobenmittel und -varianz)

$y = (y_1, ..., y_n)$ sei die Stichprobe eines parametrischen Produktmodells. Dann gelten

- Das Stichprobenmittel
$$\bar{y} := \frac{1}{n}\sum_{i=1}^{n} y_i \tag{29}$$
ist ein erwartungstreuer Schätzer des Erwartungswerts $\mathbb{E}_\theta(y_1)$.

- Die Stichprobenvarianz
$$S^2 := \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2 \tag{30}$$
ist ein erwartungstreuer Schätzer der Varianz $\mathbb{V}_\theta(y_1)$.

Beweis

Mit der Linearität von Erwartungswerten ergibt sich zunächst ergibt sich dann

$$\mathbb{E}_\theta(\bar{y}) = \mathbb{E}_\theta\left(\frac{1}{n}\sum_{i=1}^n y_i\right) = \frac{1}{n}\sum_{i=1}^n \mathbb{E}_\theta\left(y_i\right) = \frac{1}{n}\sum_{i=1}^n \mathbb{E}_\theta\left(y_1\right) = \frac{1}{n} n\mathbb{E}_\theta\left(y_1\right) = \mathbb{E}_\theta\left(y_1\right).$$

Dies zeigt die Erwartungstreue des Stichprobenmittels als Schätzer des Erwartungswertes.

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

$$\mathbb{V}_\theta(\bar{y}) = \mathbb{V}_\theta\left(\frac{1}{n}\sum_{i=1}^n y_i\right) = \frac{1}{n^2}\sum_{i=1}^n \mathbb{V}_\theta\left(y_i\right) = \frac{1}{n^2}\sum_{i=1}^n \mathbb{V}_\theta\left(y_1\right) = \frac{1}{n^2} n\mathbb{V}_\theta\left(y_1\right) = \frac{\mathbb{V}_\theta\left(y_1\right)}{n}.$$

Weiterhin gilt

$$\sum_{i=1}^n \left(y_i - \bar{y}\right)^2 = \sum_{i=1}^n (y_i - \mathbb{E}_\theta(y_1))^2 - n(\bar{y} - \mathbb{E}_\theta(y_1))^2,$$

weil

Schätzereigenschaften bei endlichen Stichproben

$$
\begin{aligned}
\sum_{i=1}^{n}(y_i-\bar{y})^2 &= \sum_{i=1}^{n}\left(y_i-\mathbb{E}_\theta(y_1)-\bar{y}+\mathbb{E}_\theta(y_1)\right)^2 \\
&= \sum_{i=1}^{n}\left((y_i-\mathbb{E}_\theta(y_1))-(\bar{y}-\mathbb{E}_\theta(y_1))\right)^2 \\
&= \sum_{i=1}^{n}(y_i-\mathbb{E}_\theta(y_1))^2-2(\bar{y}-\mathbb{E}_\theta(y_1))\left(\sum_{i=1}^{n}(y_i-\mathbb{E}_\theta(y_1))\right)+\sum_{i=1}^{n}(\bar{y}-\mathbb{E}_\theta(y_1))^2 \\
&= \sum_{i=1}^{n}(y_i-\mathbb{E}_\theta(y_1))^2-2(\bar{y}-\mathbb{E}_\theta(y_1))\left(\sum_{i=1}^{n}y_i-n\mathbb{E}_\theta(y_1)\right)+n(\bar{y}-\mathbb{E}_\theta(y_1))^2 \qquad (31)\\
&= \sum_{i=1}^{n}(y_i-\mathbb{E}_\theta(y_1))^2-2(\bar{y}-\mathbb{E}_\theta(y_1))\left(n\left(\frac{1}{n}\sum_{i=1}^{n}y_i\right)-n\mathbb{E}_\theta(y_1)\right)+n(\bar{y}-\mathbb{E}_\theta(y_1))^2 \\
&= \sum_{i=1}^{n}(y_i-\mathbb{E}_\theta(y_1))^2-2n(\bar{y}-\mathbb{E}_\theta(y_1))^2+n(\bar{y}-\mathbb{E}_\theta(y_1))^2 \\
&= \sum_{i=1}^{n}(y_i-\mathbb{E}_\theta(y_1))^2-n(\bar{y}-\mathbb{E}_\theta(y_1))^2.
\end{aligned}
$$

Es ergibt sich also

$$
\begin{aligned}
\mathbb{E}_\theta\left((n-1)S^2\right) &= \mathbb{E}_\theta\left(\sum_{i=1}^n (y_i - \bar{y})^2\right) \\
&= \mathbb{E}_\theta\left(\sum_{i=1}^n (y_i - \mathbb{E}_\theta(y_1))^2 - n(\bar{y} - \mathbb{E}_\theta(y_1))^2\right) \\
&= \sum_{i=1}^n \mathbb{E}_\theta\left((y_i - \mathbb{E}_\theta(y_1))^2\right) - n\mathbb{E}_\theta\left((\bar{y} - \mathbb{E}_\theta(y_1))^2\right) \\
&= n\mathbb{V}_\theta(y_1) - n\mathbb{V}_\theta(\bar{y}) \\
&= n\mathbb{V}_\theta(y_1) - n\frac{\mathbb{V}_\theta(y_1)}{n} \\
&= n\mathbb{V}_\theta(y_1) - \mathbb{V}_\theta(y_1) \\
&= (n-1)\mathbb{V}_\theta(y_1)
\end{aligned}
$$

Schließlich ergibt sich

$$
\mathbb{E}_\theta(S^2) = \mathbb{E}_\theta\left(\frac{1}{n-1}(n-1)S^2\right) = \frac{1}{n-1}\mathbb{E}_\theta\left((n-1)S^2\right) = \frac{1}{n-1}(n-1)\mathbb{V}_\theta(y_1) = \mathbb{V}_\theta(y_1)
$$

und damit die Erwartungstreue der Stichprobenvarianz als Schätzer der Varianz. $\square$

Schätzereigenschaften bei endlichen Stichproben

Theorem (Verzerrtheit der Stichprobenstandardabweichung)

$y := (y_1, ..., y_n)$ sei die Stichprobe eines parametrischen Produktmodells. Dann ist die Stichprobenstandardabweichung

$$S := \sqrt{S^2} \tag{32}$$

ein verzerrter Schätzer der Standardabweichung $\mathbb{S}_\theta(y_1)$.

<u>Beweis</u>

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

$$\mathbb{E}_\theta(S) = \mathbb{E}_\theta\left(\sqrt{S^2}\right) < \sqrt{\mathbb{E}_\theta(S^2)} = \sqrt{\mathbb{V}_\theta(y_1)} = \mathbb{S}_\theta(y_1). \tag{33}$$

$\square$

Bemerkung

- Nichtlineare Transformationen von erwartungstreuen Schätzern liefern oft verzerrte Schätzer.

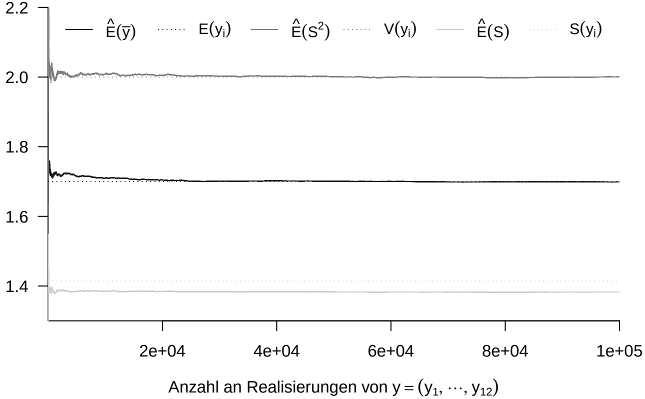
Schätzereigenschaften bei endlichen Stichproben

Simulation von $y_1, ..., y_{12} \sim N(\mu, \sigma^2)$ 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
nsim    = 1e5                       # Anzahl der Simulationen
y_bar   = rep(NaN,nsim)             # Stichprobenmittelarray
s_sqr   = rep(NaN,nsim)             # Stichprobenvarianzarray
s       = rep(NaN,nsim)             # Stichprobenstandardabweichungarray

# Simulationsiterationen
for(sim in 1:nsim){

    # Stichprobenrealisation von y_1,...,y_{12}
    y           = rnorm(n,mu,sqrt(sigsqr))

    # Erwartungswert-, Varianz-, Standardabweichungschätzer
    y_bar[sim] = mean(y)            # Stichprobenmittel
    s_sqr[sim] = var(y)             # Stichprobenvarianz
    s[sim]     = sd(y)              # Stichprobenstandardabweichung
}

# Erwartungswertschaetzung
E_hat_y_bar = cumsum(y_bar)/(1:nsim) # \mathbb{E}(\bar{y}) Schätzungen
E_hat_s_sqr = cumsum(s_sqr)/(1:nsim) # \mathbb{E}(S^2) Schätzungen
E_hat_s     = cumsum(s)    /(1:nsim) # \mathbb{E}(S) Schätzungen
```

Schätzereigenschaften bei endlichen Stichproben

Simulation von $y_1, ..., y_{12} \sim N(\mu, \sigma^2)$ mit $n = 12$, $\mu = 1.7$, $\sigma^2 = 2$, $\sigma \approx 1.41$



Definition (Varianz und Standardfehler)

$y := (y_1, ..., y_n)$ die Stichprobe eines Frequentistischen Inferenzmodells und $\hat{\tau}$ sei ein Schätzer von τ.

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

$$\mathbb{V}_\theta(\hat{\tau}) := \mathbb{E}_\theta\left((\hat{\tau}(y) - \mathbb{E}_\theta(\hat{\tau}(y)))^2\right). \tag{34}$$

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

$$\mathsf{SE}(\hat{\tau}) := \sqrt{\mathbb{V}_\theta(\hat{\tau})} \tag{35}$$

Bemerkungen

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

Theorem (Standardfehler des Stichprobenmittels)

$y := (y_1, ..., y_n)$ sei die Stichprobe eines parametrischen Produktmodells. Dann ist der *Standardfehler des Stichprobenmittels* gegeben durch

$$\mathsf{SE}(\bar{y}) = \frac{\mathbb{S}_\theta(y_1)}{\sqrt{n}}. \tag{36}$$

<u>Beweis</u>

Per definitionem und mit $\mathbb{V}_\theta(\bar{y}) = \mathbb{V}_\theta(y_1)/n$, ergibt sich

$$\mathsf{SE}(\bar{y}) = \sqrt{\mathbb{V}_\theta(\bar{y})} = \sqrt{\frac{\mathbb{V}_\theta(y_1)}{n}} = \frac{\mathbb{S}_\theta(y_1)}{\sqrt{n}}. \tag{37}$$

Bemerkungen

- Der Standardfehler des Mittelwerts beschreibt die Variabilität des Stichprobenmittels.
- Da $\mathbb{S}_\theta(y_1)$ unbekannt ist, ist auch $\mathsf{SE}(\bar{y})$ unbekannt.
- Ein verzerrter Schätzer für den Standardfehler des Stichprobenmittels ist gegeben durch $\hat{\mathsf{SE}}(\bar{y}) = \frac{s}{\sqrt{n}}$.

Beispiel (Standardfehler des Bernoulli Parameter Maximum-Likelihood Schätzes)

Es sei $y := (y_1, ..., y_n)$ die Stichprobe eines Bernoullimodells und $\hat{\mu}^{\text{ML}}$ der Maximum-Likelihood-Schätzer für μ. Dann ist

$$\text{SE}\left(\hat{\mu}^{\text{ML}}\right) = \sqrt{\frac{\mu(1-\mu)}{n}}. \tag{38}$$

<u>Beweis</u>

Es gilt

$$\text{SE}\left(\hat{\mu}^{\text{ML}}\right) = \sqrt{\mathbb{V}_\mu\left(\hat{\mu}^{\text{ML}}\right)} = \sqrt{\mathbb{V}_\mu\left(\frac{1}{n}\sum_{i=1}^n y_i\right)} = \sqrt{\frac{1}{n^2}\sum_{i=1}^n \mathbb{V}_\mu(y_i)} = \sqrt{\frac{n\mu(1-\mu)}{n^2}} = \sqrt{\frac{\mu(1-\mu)}{n}}, \tag{39}$$

wobei die dritte Gleichung mit der Unabhängigkeit der $y_i, i = 1, ..., n$ und die vierte Gleichung mit der Varianz $\mathbb{V}_\mu(y_1) = \mathbb{V}_\mu(y_i) = \mu(1-\mu), i = 1, ..., n$ der Bernoulli Stichprobenvariablen folgt.

Bemerkung

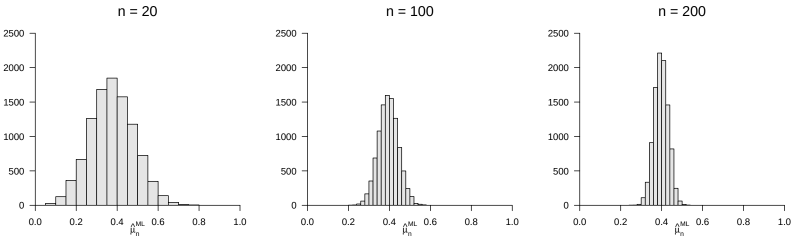
- Ein Schätzer für den Standardfehler $\text{SE}\left(\hat{\mu}^{\text{ML}}\right)$ ist $\hat{\text{SE}}\left(\hat{\mu}^{\text{ML}}\right) = \sqrt{\frac{\hat{\mu}^{\text{ML}}(1-\hat{\mu}^{\text{ML}})}{n}}$

Schätzereigenschaften bei endlichen Stichproben

Simulation von $y_1, ..., y_n \sim \text{Bern}(\mu)$ mit $\mu = 0.4$

```
# Modellformulierung
set.seed(0)                                               # Zufallszahlengeneratorzustand
mu          = 0.4                                         # wahrer, aber unbekannter, Parameterwert
n_all       = c(20,100,200)                               # Stichprobenumfänge
ns          = 1e4                                         # Anzahl der Simulationen
mu_hat_ML   = matrix(rep(NaN, length(n_all)*ns),          # Maximum-Likelihood Schätzearray
                     nrow = length(n_all))

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

    # Simulationsiterationen
    for(s in 1:ns){
        y              = rbinom(n_all[i],1,mu)            # Stichprobenrealisation von y_1,...,y_n
        mu_hat_ML[i,s] = mean(y)                          # Stichprobenmittel
    }
}
```

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



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 Stichprobenumfänge 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.

Asymptotische Schätzereigenschaften

Wir betrachten im Folgenden drei asymptotische Schätzereigenschaften:

(1) Asymptotische Erwartungstreue
(2) Konsistenz
(3) Asymptotische Normalverteilung

Um zu betonen, dass in diesem Abschnitt die Eigenschaften eines Schätzers $\hat{\tau}$ vom Stichprobenumfang abhängen, schreiben wir im Folgenden $\hat{\tau}_n$. Intuitiv haben obige asymptotische Schätzereigenschaften folgende Bedeutungen: Ein Schätzer $\hat{\tau}_n$ für τ heißt *asymptotisch erwartungstreu*, wenn der Erwartungswert von $\hat{\tau}_n$ für große Stichprobenumfänge $n \to \infty$ gleich dem wahren, aber unbekannten, Wert $\tau(\theta)$ ist. Ein Schätzer $\hat{\tau}_n$ für τ heißt *konsistent*, wenn für große Stichprobenumfänge $n \to \infty$ die Wahrscheinlichkeit dafür, dass $\hat{\tau}_n$ vom wahren, aber unbekannten, Wert $\tau(\theta)$ abweicht, beliebig klein wird. Ein Schätzer $\hat{\tau}_n$ für τ heißt *asymptotisch normalverteilt*, wenn für große Stichprobenumfänge $n \to \infty$, die Verteilung von $\hat{\tau}_n$ durch eine Normalverteilung gegeben ist.

Für folgende asymptotische Schätzereigenschaften verweisen wir auf das Vorlesungsskript:

(1) Asymptotische Effizienz

Intuitiv hat diese die folgende Bedeutung: Ein Schätzer $\hat{\tau}_n$ für τ heißt *asymptotisch effizient*, wenn für große Stichprobenumfänge $n \to \infty$ 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)

$y = (y_1, ..., y_n)$ sei die Stichprobe eines parametrischen Produktmodells und $\hat{\tau}_n$ sei ein Schätzer für τ. $\hat{\tau}_n$ heißt *asymptotisch erwartungstreu*, wenn

$$\lim_{n\to\infty} \mathbb{E}_\theta(\hat{\tau}_n(y)) = \tau(\theta) \text{ für alle } \theta \in \Theta. \tag{40}$$

Bemerkungen

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

Theorem (Asymp. Ewartungstreue des Varianzparameterschätzers)

$y := (y_1, ..., y_n)$ sei die Stichprobe eines Normalverteilungsmodells mit Varianzparameter σ^2. Dann ist der Maximum-Likelihood-Schätzer von σ^2,

$$\hat{\sigma}_n^{2^{\mathsf{ML}}} := \frac{1}{n}\sum_{i=1}^n (y_i - \bar{y}_n)^2 \tag{41}$$

nicht erwartungstreu, aber asymptotisch erwartungstreu.

Beweis

Mit der Erwartungstreue der Stichprobenvarianz ergibt sich

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

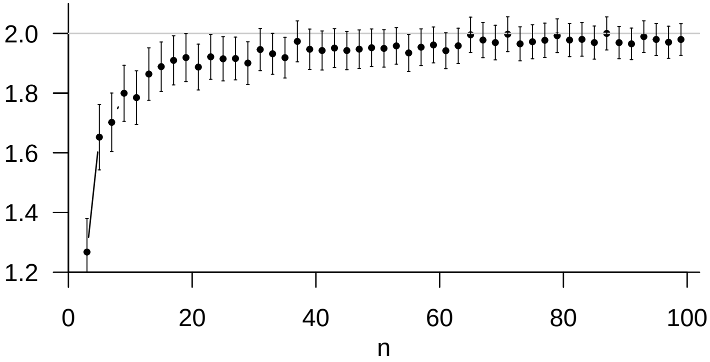
Also gilt $\mathbb{E}_{\mu,\sigma^2}\left(\hat{\sigma}_n^{2^{\mathsf{ML}}}\right) \neq \sigma^2$. $\hat{\sigma}_n^{2^{\mathsf{ML}}}$ ist also ein verzerrter Schätzer von σ^2. Allerdings gilt $(n-1)/n \to 1$ für $n \to \infty$, so dass

$$\lim_{n\to\infty}\mathbb{E}_{\mu,\sigma^2}\left(\hat{\sigma}_n^{2^{\mathsf{ML}}}\right) = \lim_{n\to\infty}\frac{n-1}{n}\sigma^2 = \sigma^2\lim_{n\to\infty}\frac{n-1}{n} = \sigma^2. \tag{42}$$

Simulation von $y_1, ..., y_n \sim N(\mu, \sigma^2)$ 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
        y               = rnorm(n[i], mu, sqrt(sigsqr))

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

Simulation von $y_1, ..., y_n \sim N(\mu, \sigma^2)$ mit $\mu = 1$, $\sigma^2 = 2$



Asymptotische Schätzereigenschaften

Definition (Konsistenz)

$y = (y_1, ..., y_n)$ sei die Stichprobe eines parametrischen Produktmodells und $\hat{\tau}_n$ sei ein Schätzer von τ. Eine Folge von Schätzern $\hat{\tau}_1, \hat{\tau}_2, ...$ wird dann eine *konsistente Folge von Schätzern* genannt, wenn für jedes beliebig kleine $\epsilon > 0$ und jedes $\theta \in \Theta$ gilt, dass

$$\lim_{n\to\infty} \mathbb{P}_\theta \left(|\hat{\tau}_n(y) - \tau(\theta)| \geq \epsilon\right) = 0. \tag{43}$$

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

Bemerkungen

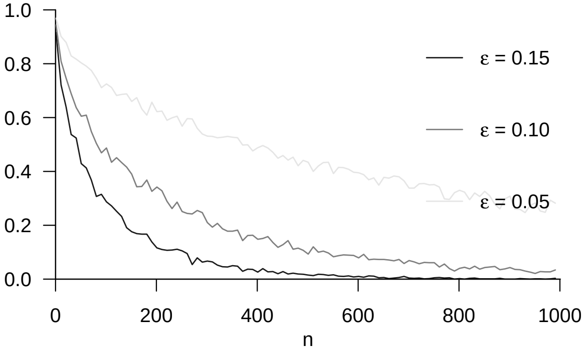
- Für $n \to \infty$ wird die Wahrscheinlichkeit, dass $\hat{\tau}_n(y)$ beliebig nah bei $\tau(\theta)$ liegt, groß.
- Für $n \to \infty$ wird die Wahrscheinlichkeit, dass $\hat{\tau}_n(y)$ von $\tau(\theta)$ abweicht, klein.
- Diese Eigenschaften gelten für alle möglichen wahren, aber unbekannten, Parameterwerte.

Asymptotische Schätzereigenschaften

Simulation der Konsistenz von $\bar{y}_n$ bei $y_1, ..., y_n \sim N(\mu, \sigma^2)$ 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      = 1000                                            # Anzahl Simulationen
E       = array(rep(NaN,nn*ne*ns),dim = c(nn,ne,ns)) # Ereignisindikatorarray


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

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

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

Simulation der Konsistenz von $\bar{y}_n$ bei $y_1, ..., y_n \sim N(\mu, \sigma^2)$ mit $\mu = 1$, $\sigma^2 = 2$



Definition (Asymptotische Normalität)

$y = (y_1, ..., y_n)$ sei die Stichprobe eines parametrischen Produktmodells 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

$$\hat{\theta}_n \overset{a}{\sim} N(\mu, \sigma^2). \tag{44}$$

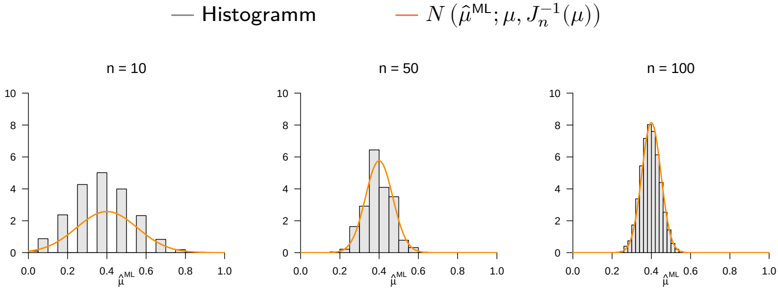
Bemerkung

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

Simulation der asymptotische Normalverteilung des ML Bernoulliparameterschätzers

```
# 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(NaN,
                        length(n_all)*ns),
                        nrow = length(n_all))
mu_hat_ML_r = 1e3                                                   # ML Schaetzerraumaufloesung
mu_hat_ML_y = seq(0,1,len = mu_hat_ML_r)                            # ML Schaetzerraum
mu_hat_ML_p = matrix(rep(NaN, length(n_all)*mu_hat_ML_r),           # ML WDF Array
                     nrow = length(n_all))

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

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

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

Simulation der asymptotische Normalverteilung des ML Bernoulliparameterschätzers



Maximum-Likelihood-Schätzer

Schätzereigenschaften bei endlichen Stichproben

Asymptotische Schätzereigenschaften

Eigenschaften von Maximum-Likelihood-Schätzern

Selbstkontrollfragen

Eigenschaften von Maximum-Likelihood-Schätzern

Theorem (Eigenschaften von Maximum-Likelihood-Schätzern)

Gegeben sei in Frequentistisches Inferenzmodell mit wahrem, aber unbekannten, Parameter θ und $\hat{\theta}_n^{\text{ML}}$ sei ein Maximum-Likelihood-Schätzer für θ. Dann gilt, dass $\hat{\theta}_n^{\text{ML}}$

(1) nicht notwendigerweise erwartungstreu, aber

(2) asymptotisch erwartungstreu,

(3) konsistent und

(4) asymptotisch normalverteilt ist.

Bemerkungen

- Maximum-Likelihood-Schätzer sind überdies asymptotisch effizient.
- Für einen Beweis verweisen wir auf Held and Sabanés Bové (2014), Abschnitt 3.4.

Maximum-Likelihood-Schätzer

Schätzereigenschaften bei endlichen Stichproben

Asymptotische Schätzereigenschaften

Eigenschaften von Maximum-Likelihood-Schätzern

Selbstkontrollfragen

Selbstkontrollfragen

1. Geben Sie die Definition des Begriffs eines Parameterpunktschätzers wieder.
2. Erläutern Sie den Begriff des Parameterpunktschätzers.
3. Geben Sie Definition der Begriffe der Likelihood-Funktion und der Log-Likelihood-Funktion wieder.
4. Geben Sie Definition des Begriffs des Maximum-Likelihood Schätzes wieder.
5. Erläutern Sie das Vorgehen zur Gewinnung von Maximum-Likelihood-Schätzern.
6. Geben Sie das Theorem zum Maximum-Likelihood-Schätzer des Bernoullimodellparameters wieder.
7. Geben Sie das Theorem zu den Maximum-Likelihood-Schätzern der Normalverteilungsmodellparameter wieder.
8. Geben Sie die Definition des Begriffs der Erwartungstreue eines Schätzers wieder.
9. Erläutern Sie den Begriff der Erwartungstreue eines Schätzers.
10. Geben Sie Definition der Begriffe der Varianz und des Standardfehlers eines Schätzers wieder.
11. Erläutern Sie den Begriff der asymptotischen Erwartungstreue eines Schätzers.
12. Erläutern Sie den Begriff der Konsistenz eines Schätzers.
13. Erläutern Sie den Begriff der asymptotischen Normalität eines Schätzers.
14. Geben Sie das Theorem zu den Eigenschaften von Maximum-Likelihood-Schätzern wieder.

Selbstkontrollfragen | Lösungen

1. Siehe Definition (Parameterpunktschätzer).
2. Ein Parameterpunktschätzer gibt basierend auf einer Stichprobe einen skalaren Tipp für den entsprechenden wahren, aber unbekannten, Parameter an.
3. Siehe Definition (Likelihood-Funktion und Log-Likelihood-Funktion).
4. Siehe Definition (Maximum-Likelihood-Schätzer)
5. Um einen Maximum-Likelihood-Schätzer zu gewinnen, formuliert man zunächst die Log-Likelihood-Funktion und bestimmt dann die Nullstellen ihrer Ableitung als potentielle Maximumstellen. Dazu nutzt man in klassischen Beispielen meist die analytische, in der Anwendung meist die numerische Optimierung.
6. Siehe Theorem (Maximum-Likelihood-Schätzer des Bernoullimodells).
7. Siehe Theorem (Maximum-Likelihood-Schätzer des Normalverteilungsmodells).
8. Siehe Definition (Fehler, Systematischer Fehler, und Erwartungstreue).
9. Ein Schätzer heißt erwartungstreu, wenn sein Erwartungswert mit dem von ihm geschätzten wahren, aber unbekannten, Wert identisch ist.
10. Siehe Definition (Varianz und Standardfehler).
11. Ein Schätzer heißt asymptotisch erwartungstreu, wenn sein Erwartungswert für gegen unendlich gehende Stichprobenumfänge mit dem von ihm geschätzten wahren, aber unbekannten, Wert identisch ist.
12. Ein Schätzer heißt konsistent, wenn für große Stichprobenumfänge die Wahrscheinlichkeit dafür, dass der Schätzerwert vom wahren, aber unbekannten, Wert abweicht, beliebig klein wird.
13. Ein Schätzer heißt asymptotisch normalverteilt, die Verteilung des Schätzers für große Stichprobenumfänge durch eine Normalverteilung gegeben ist.
14. Siehe Theorem (Eigenschaften von Maximum-Likelihood-Schätzern).

Referenzen

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.