Tutorium

Wahrscheinlichketstheorie und Frequentistische Inferenz

BSc Psychologie WiSe 2022/23

Belinda Fleischmann

Orga

Tutorium zu Wahrscheinlichkeitstheorie und Frequentistische Inferenz

Mittwochs	**Tutorium**	Donnerstags	Vorlesung
12.10.22	(1) Einführung Tut. + Mathe Gdlg.	13.10.22	(1) Einführung
19.10.22	(1) Einführung (VO SKF)	20.10.22	(2) Wahrscheinlichkeitsräume
26.10.22	(2) Wahrscheinlichkeitsräume	27.10.22	(3) Elementare Wahrscheinlichkeiten
02.11.22	(3) Elementare Wahrscheinlichkeiten	03.11.22	(4) Zufallsvariablen I
09.11.22	(4) Zufallsvariablen	10.11.22	(4) Zufallsvariablen II
16.11.22	(4) Zufallsvariablen	17.11.22	(5) Multivariate Verteilungen
23.11.22	(5) Multivariate Verteilungen	24.11.22	(6) Erwartungswert und Kovarianz
30.11.22	(6) Erwartungswert und Kovarianz	01.12.22	(7) Ungleichungen und Grenzwerte
07.12.22	(7) Ungleichungen und Grenzwerte	08.12.22	(8) Transformationen d. Normalverteilung
14.12.22	(8) Transformationen d. Normalverteilung	15.12.22	(9) Grundbegr. Frequentistischer Inferenz
21.12.22	(9) Grundbegr. Frequentistischer Inferenz	22.12.22	Weihnachtspause
	Weihnachtspause		
04.01.23	*Ersatztermin im Januar*	05.01.23	(10) Parameterschätzung
11.01.23	**(10) Parameterschätzung**	12.01.23	(11) Konfidenzintervalle
18.01.23	(11) Konfidenzintervalle	19.01.23	(12) Hypothesentests I
25.01.23	(12) Hypothesentests I	26.01.23	(12) Hypothesentests II
27.01.23	(12) Hypothesentests II		
02.02.23	Klausur		

(10) Parameterschätzung

Selbstkontrollfragen

1. Definieren und erläutern Sie den Begriff des Parameterpunktschätzers.
2. Definieren Sie den Begriff der Likelihood-Funktion.
3. Definieren Sie den Begriff der Log-Likelihood-Funktion.
4. Definieren Sie den Begriff des Maximum-Likelihood Schätzers.
5. Erläutern Sie das Vorgehen zur Maximum-Likelihood Schätzung für ein parametrisches Produktmodell.
6. Geben Sie den Maximum-Likelihood Schätzer für den Parameter μ des Bernoullimodells an.
7. Geben Sie den Maximum-Likelihood Schätzer für den Parameter μ des Normalverteilungmodels an.
8. Geben Sie den Maximum-Likelihood Schätzer für den Parameter σ^2 des Normalverteilungmodels an.
9. Definieren und erläutern Sie den Begriff der Erwartungstreue eines Schätzers.
10. Definieren Sie die Begriffe der Varianz und des Standardfehlers eines Schätzers.
11. Erläutern Sie den Begriff des asymptotischen Erwartungstreue eines Schätzers
12. Erläutern Sie den Begriff der Konsistenz eines Schätzers.
13. Erläutern Sie den Begriff der asymptotischen Normalität eines Schätzers.
14. Nennen Sie vier Eigenschaften eines Maximum-Likelihood Schätzers.

1. Definieren und erläutern Sie den Begriff des Parameterpunktschätzers.

Definition (Parameterpunktschätzer)

$\mathcal{M} := (\mathcal{Y}, \mathcal{A}, \{\mathbb{P}_\theta | \theta \in \Theta\})$ sei ein statistisches Modell, $(\Theta, \mathcal{S})$ 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.

weitere Erläuterung

- Parameterpunktschätzer sind Abbildungen, die Werte aus dem Datenraum ($\mathcal{Y}$) eines statistisches Modells nehmen und Werte im Parameterraum (Θ) annehmen.
- Intuitiv ist ein Parameterschätzer ein konkreter Zahlenwert, der den Wert eines wahren, aber unbekannten Parameters "schätzt".

2. Definieren Sie den Begriff der Likelihood-Funktion.

Definition (Likelihood-Funktion)

$\mathcal{M}$ sei ein parametrisches statistisches Produktmodell mit WMF oder WDF p_θ, also $\upsilon_1, \ldots, \upsilon_n \sim p_\theta$. Dann ist die *Likelihood Funktion* definiert als

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

Bemerkungen

- L_n ist eine Funktion des Parameters eines statistischen Modells.
- Werte von L_n sind die gemeinsamen Wahrscheinlichkeitsmassen bzw. -dichten von Datenwerten $y_1, \ldots, y_n$.
- Generell gibt es keinen Grund anzunehmen, dass L_n über Θ zu 1 integriert.
- Die Likelihood-Funktion ist also keine WMF oder WDF.

3. Definieren Sie den Begriff der Log-Likelihood-Funktion.

Definition (Log-Likelihood-Funktion)

$\mathcal{M}$ sei ein parametrisches statistisches Produktmodell mit WMF oder WDF p_θ, also $\upsilon_1, ..., \upsilon_n \sim p_\theta$ und Likelihood-Funktion L_n. Dann ist die *Log-Likelihood-Funktion* ist definiert als

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

Bemerkungen

- Die Log-Likelihood-Funktion ist die logarithmierte Likelihood-Funktion.
- Das Arbeiten mit der Log-Likelihood-Funktion ist oft einfacher als mit der Likelihood Funktion.

4. Definieren Sie den Begriff des Maximum-Likelihood Schätzers.

Definition (Maximum-Likelihood Schätzer)

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

$$\hat{\theta}_n^{\mathsf{ML}} : \mathcal{Y} \to \Theta, y \mapsto \hat{\theta}_n^{\mathsf{ML}}(y) := \underset{\theta \in \Theta}{\arg\max}\, L_n(\theta) = \underset{\theta \in \Theta}{\arg\max}\, \ell_n(\theta). \tag{3}$$

Bemerkungen

- $L_n(\theta) := \prod_{i=1}^n p_\theta(y_i)$ hängt von $y := (y_1, \ldots, y_n)$ ab, also hängt auch $\hat{\theta}_n^{\mathsf{ML}}(y)$ von y ab.
- Weil $\ln$ monoton steigend ist, entspricht eine Maximumstelle von ℓ_n einer Maximumstelle von L_n.
- Multiplikation von L_n 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 (vgl. Vorkurs Einheit (5) Differentialrechnung - Extremwerte und Extremstellen).

Achtung

- $\arg\max_{\theta \in \Theta} L_n(\theta)$ ist eine Maximum-*stelle* und *nicht* das Maximum.
- Das Maximum, also der maximale Wert, den $L_n(\theta)$ annehmen kann, wird mit $\max_{\theta \in \Theta} L_n(\theta)$ bezeichnet.
- $\arg\max_{\theta \in \Theta} L_n(\theta)$ ergibt *denjenigen* Wert für $\theta \in \Theta$, für den $L_n(\theta)$ maximal wird.

5. Erläutern Sie das Vorgehen zur Maximum-Likelihood Schätzung für ein parametrisches Produktmodell.

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.

6. Geben Sie den Maximum-Likelihood Schätzer für den Parameter μ des Bernoullimodells an.

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

7. Geben Sie den Maximum-Likelihood Schätzer für den Parameter μ des Normalverteilungsmodels an.

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

8. Geben Sie den Maximum-Likelihood Schätzer für den Parameter σ^2 des Normalverteilungsmodels an.

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

Ein Bsp. für L_n und der Unterschied zwischen $p(x)$ und $p_\theta(y_i)$

Wir schauen uns die Likelihood-Funktion L_n am Beispiel einer Bernoulli-ZV und dessen Realisierungen genauer an.

- Mal angenommen, wir haben einen **Datensatz** $y = (y_1, \ldots y_n) = (0, 1, 1, 0, 1)$ von $n = 5$ Münzwürfen mit einer **verbogenen Münze** vorliegen.
- Da die Münze verbogen ist, können wir davon ausgehen, dass die Wahrscheinlichkeiten für Kopf oder Zahl nicht 0.5 sind, aber **wir wissen nicht, wie hoch die Wahrscheinlichkeiten sind** (anders ausgedrückt, was der wahre, aber unbekannte Parameter der Verteilung bzw. WMF/WDF ist, wenn wir die Daten als Realisierungen einer ZV modellieren).
- Wir nehmen an, dass dieser Datensatz Realisierungen einer Zufallsvariable entspricht und modellieren den Münzwurf mithilfe einer **Bernoulli-Zufallsvariable**.
- Konkret bedeutet das, wir nehmen an, dass der uns vorliegender Datensatz $y = (y_1, \ldots y_n)$ **eine der möglichen Realisierungen** einer Stichprobe $\upsilon_1, \ldots, \upsilon_n$ ist, wobei jedes υ_i einer Bernoulli-Verteilung unterliegt. Das schreiben wir mit $\upsilon_1, \ldots, \upsilon_n \sim \text{Bern}(\mu)$

Wiederholung aus (4) Zufallsvariablen:

Definition (Bernoulli-Zufallsvariable)

Es sei ξ eine Zufallsvariable mit Ergebnisraum $\mathcal{X} = \{0, 1\}$ und WMF

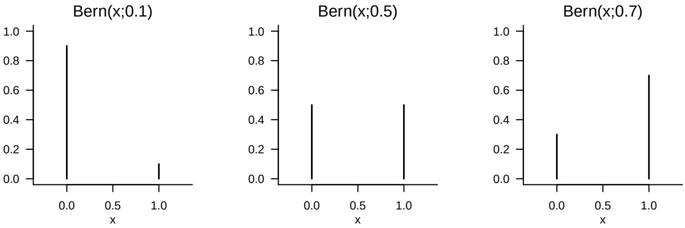
$$p : \mathcal{X} \to [0, 1], x \mapsto p(x) := \mu^x (1 - \mu)^{1-x} \text{ mit } \mu \in [0, 1]. \tag{4}$$

Dann sagen wir, dass ξ einer *Bernoulli-Verteilung mit Parameter* $\mu \in [0, 1]$ unterliegt und nennen ξ eine *Bernoulli-Zufallsvariable*. Wir kürzen dies mit $\xi \sim \text{Bern}(\mu)$ ab. Die WMF einer Bernoulli-Zufallsvariable bezeichnen wir mit

$$\text{Bern}(x; \mu) := \mu^x (1 - \mu)^{1-x}. \tag{5}$$

Ein Bsp. für L_n und der Unterschied zwischen $p(x)$ und $p_\theta(y_i)$ (fortgeführt)

- Die **WMF** $p(x)$ ordnet jedem Wert $x \in \mathcal{X} = \{0, 1\}$, den eine Zufallsvariable υ_i annehmen kann, eine Wahrscheinlichkeitsmasse zu. Deren funktionale Form $\mu^x(1-\mu)^{1-x}$ ist von μ geprägt.
- Bei einem Münzwurf mit ungezinktem Würfel wäre $\mu = 0.5$, und dann wären $p(0) = 0.5^0(1-0.5)^{1-0} = 1 \cdot 0.5^1 = 0.5$ und $p(1) = 0.5^1(1-0.5)^{1-1} = 0.5 \cdot 0.5^0 = 0.5$



- Wir wissen nicht, was der der wahre, aber unbekannte Parameterwert μ in unserem Modell einer verbogenen Münze ist.
- Bei der **Parameterschätzung** im Rahmen der frequentistischne Inferenz geben wir basierend auf Daten y eine *Schätzung* dafür ab, was der wahre, aber unbekannte Parameterwert *sein könnte*.
- Eine Methode, eine solche Schätzung vorzunehmen, ist die **ML-Schätzung**.
- Dabei betrachten wir gewissermaßen verschiedene potentielle Werte für μ und suchen den Wert, für den die Likelihood-Funktion $L_n(\mu)$ maximal wird.

Ein Bsp. für L_n und der Unterschied zwischen $p(x)$ und $p_\theta(y_i)$ (fortgeführt)

- Gemäß des Vorgehens zur Gewinnung von ML-Schätzern formulieren wir hierfür zunächst die **Likelihood-Funktion**

$$
\begin{aligned}
L_n(\theta) &:= \prod_{i=1}^{n} p_\theta(y_i) \\
&= \prod_{i=1}^{n} \mu^{y_i}(1-\mu)^{1-y_i} = \mu^{\sum_{i=1}^{n} y_i}(1-\mu)^{n-\sum_{i=1}^{n} y_i}.
\end{aligned}
$$

- Die Likelihood-Funktion ist das Produkt der Wahrscheinlichkeitsdichten bzw. -massen aller Datenpunkte gegebenen eines bestimmten Parameterwertes $p_\theta(y_i)$.
- Hierbei verwenden wir die funktionale Form der WMF der Bernoulli-ZV.
- **Beispiele für** $p_\theta(y_i)$**-Werte** für die Realiserung $y_i = 1$:
  - wenn $\theta = \mu = 0.5$, ergibt sich $p_{0.5}(1) = 0.5^1(1-0.5)^{1-1} = 0.5$
  - wenn $\theta = \mu = 0.1$, ergibt sich $p_{0.1}(1) = 0.1^1(1-0.1)^{1-1} = 0.1$
  - wenn $\theta = \mu = 0.7$, ergibt sich $p_{0.7}(1) = 0.1^1(1-0.7)^{1-1} = 0.7$
- $p_\theta(y_i)$ ist also die Wahrscheinlichkeit des Datenpunkts y_i (Realisierung der ZV), wenn der Parameter θ einen bestimmten Wert hätte und damit auch die ZV eine bestimmte “theoretische” Verteilung hätte.Anders ausgedrückt, die Wahrscheinlichkeit der Daten unter einem angenommenen Modell.

Ein Bsp. für L_n und der Unterschied zwischen $p(x)$ und $p_\theta(y_i)$ (fortgeführt)

Um es noch ein bisschen konkreter zu machen: In unserem Beispiel, wenn wir $n = 5$ Münzwurfergebnisse $y = (y_1, ...y_n) = (0, 1, 1, 0, 1)$ vorliegen haben, könnten wir die Likehoods für verschiedene potentielle Parameterwerte berechnen.

Beispiele für Likelihood-Werte $L_n(\theta)$

- $L_n(0.5) = p_{0.5}(0) \cdot p_{0.5}(1) \cdot p_{0.5}(1) \cdot p_{0.5}(0) \cdot p_{0.1}(1) = 0.5 \cdot 0.5 \cdot 0.5 \cdot 0.5 \cdot 0.5 = 0.03125$
- $L_n(0.1) = p_{0.1}(0) \cdot p_{0.1}(1) \cdot p_{0.1}(1) \cdot p_{0.1}(0) \cdot p_{0.1}(1) = 0.9 \cdot 0.1 \cdot 0.1 \cdot 0.9 \cdot 0.1 \approx 0.00081$
- $L_n(0.7) = p_{0.7}(0) \cdot p_{0.7}(1) \cdot p_{0.7}(1) \cdot p_{0.7}(0) \cdot p_{0.7}(1) = 0.3 \cdot 0.7 \cdot 0.7 \cdot 0.3 \cdot 0.7 = 0.03087$
- $L_n(0.6) = p_{0.6}(0) \cdot p_{0.6}(1) \cdot p_{0.6}(1) \cdot p_{0.6}(0) \cdot p_{0.6}(1) = 0.4 \cdot 0.6 \cdot 0.6 \cdot 0.4 \cdot 0.6 = 0.03456$

Wenn wir also nur diese vier Parameterwerte vergleichen würden, würden wir die höchste Likelihood für $\mu = 0.6$ erhalten. Anders ausgedrückt, die Wahrscheinlichkeit, diese Daten zu beobachten ist am höchsten unter dem Modell einer Bernoulli-ZV mit Parameter $\mu = 0.6$.

In der Praxis ist das manuelle Bestimmen und Vergleichen verschiedener L_n überflüssig, weil uns für bestimmte, Modelle (wie etwa Bernoulli oder Normalverteilung) Schätzformeln vorliegen, von denen wir aufgrund der Ergebnisse analytischer Optimierung wissen, dass diese den ML-Schätzern entsprechen. Für Bernoulli ist das $\hat{\mu}_n^{ML} := \frac{1}{n}\sum_{i=1}^n y_i$.

In unserem Beispiel könnten wir also dem ML-Schätzer für μ bestimmen mit

$$\hat{\mu}_n^{ML} = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{1}{5}(0, 1, 1, 0, 1) = \frac{1}{5} \cdot 3 = 0.6$$

9. Definieren und erläutern Sie den Begriff der Erwartungstreue eines Schätzers.

Definition (Fehler, Systematischer Fehler, und Erwartungstreue)

$\upsilon := \upsilon_1, ..., \upsilon_n \sim p_\theta$ sei eine Stichprobe und $\hat{\tau}_n$ sei ein Schätzer für τ.

- Der *Fehler* von $\hat{\tau}_n$ ist definiert als
$$\hat{\tau}_n(\upsilon) - \tau(\theta). \tag{6}$$
- Der *systematische Fehler (Bias)* von $\hat{\tau}_n$ ist definiert als
$$\mathrm{B}(\hat{\tau}_n) := \mathbb{E}_\theta(\hat{\tau}_n(\upsilon)) - \tau(\theta). \tag{7}$$
- $\hat{\tau}_n$ heißt *erwartungstreu (unbiased)*, wenn
$$\mathrm{B}(\hat{\tau}_n) = 0 \Leftrightarrow \mathbb{E}_\theta(\hat{\tau}_n(\upsilon)) = \tau(\theta) \text{ für alle } \theta \in \Theta, n \in \mathbb{N}. \tag{8}$$
Andernfalls heißt $\hat{\tau}_n$ *verzerrt (biased)*.

- Der Fehler hängt von einer Realisation ab.
- Der systematische Fehler ist der erwartete Fehler über viele Stichprobenrealisationen.
- Ein Schätzer $\hat{\tau}_n$ heißt *erwartungstreu*, wenn sein Erwartungswert dem wahren, aber unbekannten, Wert $\tau(\theta)$ für alle $\theta \in \Theta$ gleicht.
- Ein Parameterschätzer ist erwartungstreu, wenn $\mathbb{E}_\theta(\hat{\theta}_n(\upsilon)) = \theta$

Wdhl.: Standardannahmen Frequentistischer Inferenz

$\mathcal{M}$ sei ein statistisches Modell mit Stichprobe $\upsilon_1, ..., \upsilon_n \sim p_\theta$. **Es wird angenommen, dass ein konkret vorliegender Datensatz $y = (y_1, ..., y_n) \in \mathbb{R}^n$ eine der möglichen Realisierungen von $\upsilon_1, ..., \upsilon_n \sim p_\theta$ ist.** Aus Frequentistischer Sicht kann man eine Studie unter gleichen Umständen 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 $\bar{y}_n^{(1)} = \frac{1}{n}\sum_{i=1}^n y_i^{(1)}$

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

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

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

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

Um die Qualität statistischer Methoden zu beurteilen betrachtet die Frequentistische Statistik deshalb die Wahrscheinlichkeitsverteilungen von Schätzern und Statistiken unter Annahme von $\upsilon_1, ..., \upsilon_n \sim p_\theta$. Was zum Beispiel ist die Verteilung der $\bar{y}_n^{(1)}$, $\bar{y}_n^{(2)}$, $\bar{y}_n^{(3)}$, $\bar{y}_n^{(4)}$, ... also die Verteilung der Zufallsvariable $\bar{\upsilon}_n$?

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 Normalfall nur eines vorliegenden Datensatzes, kann sie auch "schlecht" sein.

10. Definieren Sie die Begriffe der Varianz und des Standardfehlers eines Schätzers.

Definition (Varianz und Standardfehler)

$\upsilon := \upsilon_1, ..., \upsilon_n \sim p_\theta$ sei eine Stichprobe und $\hat{\tau}_n$ sei ein Schätzer von τ.

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

$$\mathbb{V}_\theta(\hat{\tau}_n) := \mathbb{E}_\theta\left((\hat{\tau}_n(\upsilon) - \mathbb{E}_\theta(\hat{\tau}_n(\upsilon)))^2\right). \quad (9)$$

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

$$\mathrm{SE}(\hat{\tau}_n) := \sqrt{\mathbb{V}_\theta(\hat{\tau}_n)} \quad (10)$$

Bemerkungen

- Die Varianz eines Schätzers $\hat{\tau}_n$ ist die Varianz der Zufallsvariable $\hat{\tau}_n(\upsilon)$.
- Der Standardfehler eines Schätzers $\hat{\tau}_n$ ist die Standardabweichung von $\hat{\tau}_n(\upsilon)$.
- Die Varianz von Schätzern basiert auf der Varianz der Daten, ist aber nicht das gleiche.

Simulation von $\upsilon_1, ..., \upsilon_n \sim \text{Bern}(\mu)$ 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),      # Maximum-Likelihood Schätzearray
                     nrow = length(n_all))

# Stichprobengroessenitertationen
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 $\hat{\mu}_n^{\text{ML}}$ hängen von n ab.