Stochastik

FH Aachen – Studienstandort Aachen
5. Dezember 2025

Prof. Dr. Florian Heinrichs

Schätzer

Münzwurf

Beispiel 137

- Wir werfen eine Münze 100 Mal und erhalten 33 Mal Kopf und 67 Mal Zahl
- Vermutlich ist die Münze nicht fair, d.h. $\mathbb{P}(\text{“Kopf”}) \neq 0.5$
- Aber: Grundsätzlich ist das Ergebnis möglich
- Die Wahrscheinlichkeit dafür beträgt ca. 0.023%
- Wie hoch ist $p = \mathbb{P}(\text{“Kopf”})$ wirklich?
- Wir können p schätzen: $\hat{p} = \frac{33}{100}$
- Ist $\hat{p}$ ein guter Schätzer?
- Was ist ein guter Schätzer?
- Was ist ein Schätzer?

Statistisches Modell

Definition 60

Sei $\mathcal{X} \subset \mathbb{R}^n$ und $X : \Omega \to \mathcal{X}$ eine Zufallsvariable. Eine *Stichprobe* ist eine Realisierung von $X = (X_1, \dots, X_n)^T$, d.h. $X(\omega)$ für ein $\omega \in \Omega$.

Bemerkung 26

> Die Menge $\mathcal{X}$ wird als *Stichprobenraum* bezeichnet.
> Die Definition ist eine Formalisierung von "Sei $x_1, \dots, x_n$ eine Stichprobe"
> Die Stichprobe $x_1, \dots, x_n$ entspricht den Koordinaten des Vektors $(X_1(\omega), \dots, X_n(\omega))^T = X(\omega)$
> Die Definition ist allgemein: Grundsätzlich können die Zufallsvariablen $X_1, \dots, X_n$ abhängig sein. Wir nehmen im Folgenden an, dass die ZV unabhängig sind.

Statistisches Modell

Beispiel 138: n-facher Münzwurf

- Der Stichprobenraum beim n-fachen Münzwurf ist $\mathcal{X} = \{0, 1\}^n$, wobei $1 \hat{=}$ "Kopf" und $0 \hat{=}$ "Zahl".
- **Bevor** wir werfen:
  - Das Ergebnis im i. Wurf ist eine Zufallsvariable X_i mit Werten in $\{0, 1\}$.
  - Das Ergebnis aller Würfe ist eine Zufallsvariable $X = (X_1, \ldots, X_n)$ in $\{0, 1\}^n$.
- **Nachdem** wir geworfen haben:
  - Das Ergebnis des i. Wurfs $x_i = X_i(\omega)$ ist in $\{0, 1\}$.
  - Das Gesamtergebnis ist eine Stichprobe $(x_1, \ldots, x_n) = X(\omega)$ in $\{0, 1\}^n$.

Ist unsere Annahme erfüllt, dass die Zufallsvariablen $X_1, \ldots, X_n$, d.h. die Koordinaten des Vektors $X \in \mathbb{R}^n$, unabhängig sind?

Statistisches Modell

Definition 61

Sei $\mathcal{X} \subset \mathbb{R}^n$ und $X : \Omega \to \mathcal{X}$ eine Zufallsvariable.

1. Ein *statistisches Modell* ist eine Familie von Wahrscheinlichkeitsmaßen $\mathcal{P}$ auf $\mathcal{X}$.
2. Falls $\mathcal{P} = \{\mathbb{P}_\theta : \theta \in \Theta\}$, heißt Θ der *Parameterraum* von $\mathcal{P}$ und $\theta \in \Theta$ wird als *Parameter* bezeichnet.

Beispiel 138: n-facher Münzwurf

Sei $\mathbb{P}_p(X_1 = 1) = p$ die Randverteilung der unabh. Würfe und $\mathbb{P}_p^n$ die gemeinsame Verteilung, dann ist $\mathcal{P} = \{\mathbb{P}_p^n : p \in (0,1)\}$ ein statistisches Modell für den Münzwurf.

Für uns relevant: Falls $X_1, \ldots, X_n$ unabhängig, jeweils mit Verteilung $\mathbb{P}_\theta$, für $\theta \in \Theta$, dann ist $\mathcal{P} = \{\mathbb{P}_\theta^n : \theta \in \Theta\}$ ein statistisches Modell für $X = (X_1, \ldots, X_n)^T$.

Statistisches Modell

Beispiel 138: n-facher Münzwurf

Zusammenfassung:

> Stichprobenraum: $\mathcal{X} = \{0, 1\}^n$

> $X_i \sim Ber(p)$ (mit Werten in $\{0, 1\}$)

> $X = (X_1, \ldots, X_n)^T$ ist eine Zufallsvariable in $\{0, 1\}^n$

> Stichprobe: $X(\omega) = (x_1, \ldots, x_n)$ in $\{0, 1\}^n$

>> Informell: "Sei $x_1, \ldots, x_n$ eine Stichprobe..."

> Randverteilung der unabhängigen Würfe: $\mathbb{P}_p(X_1 = 1) = p$

> Gemeinsame Verteilung: $\mathbb{P}_p^n$

> Statistisches Modell für den Münzwurf:

$$\mathcal{P} = \{\mathbb{P}_p^n : p \in (0, 1)\}$$

Schätzer

Definition 62

Sei $X : \Omega \to \mathcal{X}$ eine Zufallsvariable mit statistischem Modell $\mathcal{P} = \{\mathbb{P}_\theta : \theta \in \Theta\}$.

1. Eine Abbildung $T_n : \mathcal{X} \to \Theta$ bzw. die Zufallsvariable $T_n(X)$ heißt *Schätzer für* θ.
2. Eine *Schätzung* ist eine Realisierung des Schätzers, d.h. $T_n(X)(\omega)$ für ein $\omega \in \Omega$.
3. Gilt für die wahre Verteilung von X: $\mathbb{P}_X = \mathbb{P}_{\theta_0}$, so heißt θ_0 der *wahre Parameter*.

Interpretation:

> Schätzer: Allgemeines Verfahren/Formel, die $X = (X_1, \ldots, X_n)^T$ einen Parameter θ zuordnet.

> Schätzung: Einsetzen der Stichprobe $x_1, \ldots, x_n$ in den Schätzer.

Schätzer

Beispiel 139: (Fortsetzung)

> Wir werfen eine Münze n Mal: $X = (X_1, \dots, X_n)^T$

> Sei $S_n = \sum_{i=1}^n X_i$ die Anzahl der Würfe mit Ergebnis "Kopf"

> $S_n \sim Bin(n, p)$

> Definiere den Schätzer $\hat{p} = \frac{S_n}{n}$

> Dann gilt: $\mathbb{E}[\hat{p}] = \frac{1}{n}\mathbb{E}[S_n] = \frac{np}{n} = p$

> Außerdem: $\text{var}(\hat{p}) = \frac{1}{n^2}\text{var}(S_n) = \frac{np(1-p)}{n^2} = \frac{p(1-p)}{n} \xrightarrow{n\to\infty} 0$

> Der Schätzer $\hat{p}$ liegt also im Durchschnitt richtig und die Varianz wird immer kleiner

> Wenn wir oft genug werfen ($n \to \infty$), können wir p immer besser schätzen

Schätzer

Bemerkung 27

- Jede Abbildung von einer Stichprobe $X_1, \ldots, X_n$ in das Intervall $[0, 1]$ ist ein Schätzer von p
- Offensichtlich muss ein Schätzer nicht gut sein
  - $T_n(X) = 0$ ist der Schätzer, der immer $p = 0$ schätzt, unabhängig von der Stichprobe
- Gilt für eine Münze $\mathbb{P}(\text{“Kopf”}) = \frac{1}{3}$, so ist $\frac{1}{3}$ der wahre Parameter

Erwartungstreue & Konsistenz

Definition 63

Sei $X : \Omega \to \mathcal{X}$ eine Zufallsvariable mit statistischem Modell $\mathcal{P} = \{\mathbb{P}_\theta : \theta \in \Theta\}$. Für $\theta \in \Theta$ bezeichne $\mathbb{E}_\theta$ den Erwartungswert bzgl. $\mathbb{P}_\theta$ und var_θ die entsprechende Varianz.

1. Die *Verzerrung* (auch: *bias*) eines Schätzers T_n ist
$$Bias_{T_n}(\theta) = \mathbb{E}_\theta[T_n(X)] - \theta.$$
2. Ein Schätzer heißt *unverzerrt* (auch: *erwartungstreu, unbiased*), falls $Bias_{T_n}(\theta) = 0$ für alle $\theta \in \Theta$, also
$$\mathbb{E}_\theta[T_n(X)] = \theta.$$
3. Eine Folge von Schätzern $(T_n)_{n\in\mathbb{Z}}$ heißt *konsistent*, falls $\forall \varepsilon > 0, \theta \in \Theta$
$$\lim_{n\to\infty} \mathbb{P}_\theta(|T_n(X) - \theta| > \varepsilon) = 0$$

Erwartungstreue & Konsistenz

Bemerkung 28

> Ein erwartungstreuer Schätzer liegt im Durchschnitt richtig

> Ein konsistenter Schätzer “konvergiert” gegen den wahren Parameter

> > “Je mehr Daten, desto besser der Schätzer”

> Oft zeigt man die Konsistenz eines Schätzers mit dem Gesetz der großen Zahlen

Beispiel 140: (Fortsetzung)

> Der Schätzer $\hat{p} = \frac{S_n}{n}$ ist

> > Erwartungstreu: $\mathbb{E}[\hat{p}] = p$

> > Konsistent: $\mathbb{P}_p(|\hat{p} - p| > \varepsilon) \leq \frac{\mathsf{var}(\hat{p})}{\varepsilon^2} = \frac{p(1-p)}{n\varepsilon^2} \xrightarrow{n\to\infty} 0$

Erwartungstreue & Konsistenz

Satz 37

Sei $X : \Omega \to \mathcal{X}$ eine Zufallsvariable mit $X = (X_1, \dots, X_n)^T$ für unabhängig, identisch verteilte Zufallsvariablen $X_1, \dots, X_n$.

1. Falls der Erwartungswert von X_1 existiert, ist der Mittelwert $\bar{X}_n$ ein erwartungstreuer und konsistenter Schätzer für $\mathbb{E}[X_1]$.
2. Falls die Varianz von X_1 existiert und der Erwartungswert $\mu = \mathbb{E}[X_1]$ bekannt ist, ist $\hat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \mu)^2$ ein erwartungstreuer und konsistenter Schätzer für $\mathrm{var}(X_1)$.
3. Falls die Varianz von X_1 existiert und der Erwartungswert von X_1 unbekannt ist, ist die Stichprobenvarianz $s_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2$ ein erwartungstreuer und konsistenter Schätzer für $\mathrm{var}(X_1)$.

Erwartungstreue & Konsistenz

Beweis von Satz 37

1. Aufgrund der Linearität des Erwartungswerts gilt

$$\mathbb{E}[\bar{X}_n] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[X_i] = \mathbb{E}[X_1].$$

   Die Konsistenz folgt direkt aus dem starken Gesetz der großen Zahlen.

2. Definiere $Y_i = (X_i - \mu)^2$. Dann ist $\mathbb{E}[Y_i] = \sigma^2$ und die Aussage folgt aus 1.

Erwartungstreue & Konsistenz

Optional 8: (Beweis von Satz 37)

3. Aus $\mathbb{E}[X_i] = \mathbb{E}[\bar{X}_n]$ (vgl. 1.) folgt

$$\mathsf{var}(X_i - \bar{X}_n) = \mathbb{E}\Big[\Big(X_i - \bar{X}_n - \mathbb{E}[X_i - \bar{X}_n]\Big)^2\Big] = \mathbb{E}[(X_i - \bar{X}_n)^2]$$

Damit folgt

$$\begin{aligned}
\mathbb{E}[s_n^2] &= \frac{1}{n-1}\sum_{i=1}^{n} \mathsf{var}(X_i - \bar{X}_n) \\
&= \frac{1}{n-1}\sum_{i=1}^{n} \mathsf{var}(X_i) + \mathsf{var}(\bar{X}_n) - 2\mathsf{cov}(X_i, \bar{X}_n) \\
&= \frac{1}{n-1}\sum_{i=1}^{n} \mathsf{var}(X_1) + \frac{\mathsf{var}(X_1)}{n} - 2\frac{\mathsf{var}(X_1)}{n} = \frac{(n-1)\mathsf{var}(X_1)}{n-1} = \mathsf{var}(X_1)
\end{aligned}$$

Konsistenz: Vgl. Übung 14.3 in [Dehling and Haupt, 2006].

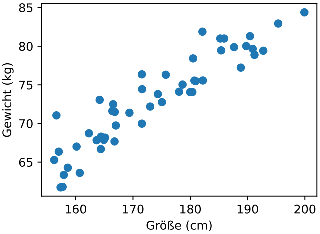
Lineare Regression

Oft interessiert uns nur θ, nicht die Verteilung $\mathbb{P}_\theta$

> **Beispiel:** Körpergröße X und -gewicht Y von Personen

> Wir vermuten einen linearen Zusammenhang: $Y = \alpha X + \beta$

> Gegeben sei eine Stichprobe *gepaarter Daten* $(x_1, y_1), \ldots, (x_n, y_n)$



> Der Zusammenhang ist nicht perfekt linear, deshalb modifizieren wir das Modell:

$$Y = \alpha X + \beta + \varepsilon$$

wobei ε ein zufälliger Fehler ist

> **Ziel:** Schätze α und β

> Wie finden wir geeignete Schätzer?

Lineare Regression

- Gegeben sei eine Stichprobe $(x_1, y_1), \ldots, (x_n, y_n)$
- **Ziel:** Schätze α und β des linearen Modells: $Y = \alpha X + \beta + \varepsilon$
- Ansatz: Minimiere die Vorhersagefehler $y_i - (\alpha x_i + \beta)$, für $i = 1, \ldots, n$ mit der *Methode der kleinsten Quadrate*

$$Q(\alpha, \beta) = \sum_{i=1}^{n} \big(y_i - (\alpha x_i + \beta)\big)^2$$

- Bestimme partielle Ableitungen

$$\frac{\partial Q}{\partial \alpha}(\alpha, \beta) = -\sum_{i=1}^{n} x_i \big(y_i - (\alpha x_i + \beta)\big) = \sum_{i=1}^{n} \alpha x_i^2 + \beta x_i - x_i y_i$$

$$\frac{\partial Q}{\partial \beta}(\alpha, \beta) = -\sum_{i=1}^{n} \big(y_i - (\alpha x_i + \beta)\big) = \sum_{i=1}^{n} \alpha x_i + \beta - y_i$$

Lineare Regression

- Gegeben sei eine Stichprobe $(x_1, y_1), \ldots, (x_n, y_n)$
- **Ziel:** Schätze α und β des linearen Modells: $Y = \alpha X + \beta + \varepsilon$
- Minimiere $Q(\alpha, \beta) = \sum_{i=1}^n \big(y_i - (\alpha x_i + \beta)\big)^2$
- Definiere $S = \sum_{i=1}^n x_i^2$ und $R = \sum_{i=1}^n x_i y_i$, dann gilt

$$\frac{\partial Q}{\partial \alpha}(\alpha, \beta) = \sum_{i=1}^n \alpha x_i^2 + \beta x_i - x_i y_i = \alpha S + \beta n \bar{X}_n - R$$

$$\frac{\partial Q}{\partial \beta}(\alpha, \beta) = \sum_{i=1}^n \alpha x_i + \beta - y_i = \alpha n \bar{X}_n + n\beta - n\bar{Y}_n$$

- Bestimme Hesse-Matrix

$$H_f = \begin{pmatrix} S & n\bar{X}_n \\ n\bar{X}_n & n \end{pmatrix}$$

- $\det(H_f) = n \sum_{i=1}^n x_i^2 - \big(\sum_{i=1}^n x_i\big)^2 > 0$ nach Cauchy-Schwarz Ungleichung, falls $(x_1, \ldots, x_n)^T, (1, \ldots, 1)^T$ linear unabhängig.

Lineare Regression

- Gegeben sei eine Stichprobe $(x_1, y_1), \ldots, (x_n, y_n)$
- **Ziel:** Schätze α und β des linearen Modells: $Y = \alpha X + \beta + \varepsilon$
- Minimiere $Q(\alpha, \beta) = \sum_{i=1}^{n} \big(y_i - (\alpha x_i + \beta)\big)^2$
- Ableitungen: $\frac{\partial Q}{\partial \alpha}(\alpha, \beta) = \alpha S + \beta n \bar{X}_n - R$ und $\frac{\partial Q}{\partial \beta}(\alpha, \beta) = \alpha n \bar{X}_n + n\beta - n\bar{Y}_n$

$$
\begin{aligned}
&\frac{\partial Q}{\partial \alpha}(\alpha, \beta) = 0 \text{ und } \frac{\partial Q}{\partial \beta}(\alpha, \beta) = 0 \\
&\iff \alpha S + \beta n \bar{X}_n = R \text{ und } \alpha n \bar{X}_n + n\beta = n\bar{Y}_n \\
&\iff \begin{pmatrix} S & n\bar{X}_n \\ n\bar{X}_n & n \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = H_f \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} R \\ n\bar{Y}_n \end{pmatrix} \\
&\iff \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = H_f^{-1} \begin{pmatrix} R \\ n\bar{Y}_n \end{pmatrix} = \frac{1}{nS - n^2 \bar{X}_n^2} \begin{pmatrix} n & -n\bar{X}_n \\ -n\bar{X}_n & S \end{pmatrix} \begin{pmatrix} R \\ n\bar{Y}_n \end{pmatrix}
\end{aligned}
$$

Lineare Regression

- Gegeben sei eine Stichprobe $(x_1, y_1), \ldots, (x_n, y_n)$
- **Ziel:** Schätze α und β des linearen Modells: $Y = \alpha X + \beta + \varepsilon$
- $Q(\alpha, \beta) = \sum_{i=1}^n \big(y_i - (\alpha x_i + \beta)\big)^2$ wird minimiert durch

$$\begin{pmatrix} \hat\alpha \\ \hat\beta \end{pmatrix} = \frac{1}{nS - n^2 \bar X_n^2} \begin{pmatrix} n & -n\bar X_n \\ -n\bar X_n & S \end{pmatrix} \begin{pmatrix} R \\ n\bar Y_n \end{pmatrix},$$

  da H_f positiv definit ist (genauer: $S > 0, \det(H_f) > 0$).
- Umformung ergibt: $\hat\beta = \frac{s_{XY}}{s_X^2}$ und $\hat\alpha = \bar y_n - \hat\beta$, wobei s_X^2 die Stichprobenvarianz und s_{XY} die *Stichprobenkovarianz* bezeichnet:

$$s_X^2 = \frac{1}{n-1}\sum_{i=1}^n (x_i - \bar x_n)^2, \quad s_{XY} = \frac{1}{n-1}\sum_{i=1}^n (x_i - \bar x_n)(y_i - \bar y_n).$$

Maximum Likelihood

- Bisher:
  - Mittelwert: Schätzer für den Erwartungswert
  - Stichprobenvarianz: Schätzer für die Varianz
  - Stichprobenkovarianz: Schätzer für die Kovarianz
- Die Schätzer sind gut (erwartungstreu und konsistent), aber wie finden wir im Allgemeinen gute Schätzer?
- Allgemeines Prinzip: Maximum-Likelihood-Schätzung (ML-Methode)
- **Idee:** Wähle Parameter, der die beobachtete Stichprobe am wahrscheinlichsten macht

Maximum Likelihood

Idee: Wähle Parameter, der die beobachtete Stichprobe am wahrscheinlichsten macht

Beispiel 141

- Wir werfen eine Münze n Mal
- Sei X_i der Indikator, dass der i-te Wurf "Kopf" ist, $i = 1, \ldots, n$
- Sei die Stichprobe $x_1, \ldots, x_n$ gegeben, wobei k Mal Kopf und $n-k$ Mal Zahl geworfen wurde
- Die Wahrscheinlichkeit für diese Stichprobe ist

$$\mathbb{P}(X_1 = x_1, \ldots, X_n = x_n) = \mathbb{P}(X_1 = x_1) \cdot \ldots \cdot \mathbb{P}(X_n = x_n) = p^k (1-p)^{n-k}$$

- Neue Perspektive: Für welches p ist die Wahrscheinlichkeit maximal?
- Definiere $L(p) = p^k (1-p)^{n-k}$
- Maximiere $L(p) \rightsquigarrow \hat{p} = \frac{k}{n}$

Maximum Likelihood

Beispiel 141: (Fortsetzung)

> Definiere $L(p) = p^k(1-p)^{n-k}$

> Es gilt

$$\begin{aligned} L'(p) &= kp^{k-1}(1-p)^{n-k} - (n-k)p^k(1-p)^{n-k-1} \\ &= \big[k(1-p) - (n-k)p\big]p^{k-1}(1-p)^{n-k-1} \end{aligned}$$

$$\begin{aligned} L'(p) = 0 &\iff k(1-p) - (n-k)p = 0 \\ &\iff k = np \iff p = \frac{k}{n} \end{aligned}$$

> $L''(p) > 0$: Übung

Maximum Likelihood

Die Maximum-Likelihood-Schätzung ist eine allgemeine Methode, um Schätzer zu finden:

Definition 64

Sei $X = (X_1, \ldots, X_n)^T : \Omega \to \mathcal{X}$ eine Zufallsvariable mit statistischem Modell $\mathcal{P} = \{\mathbb{P}_\theta^n : \theta \in \Theta\}$. Weiter sei $x_1, \ldots, x_n$ eine Stichprobe. Die Funktion $L : \Theta \to [0,1]$, definiert durch

$$L(\theta) = \begin{cases} \mathbb{P}_\theta(X_1 = x_1) \cdot \ \ldots \ \cdot \mathbb{P}_\theta(X_n = x_n) & \text{für diskrete ZV} \\ f_\theta(x_1) \cdot \ \ldots \ \cdot f_\theta(x_n) & \text{für stetige ZV} \end{cases}$$

heißt *Likelihood-Funktion*. Der Parameter $\hat{\theta}$, der $L(\theta)$ maximiert, heißt *Maximum-Likelihood-Schätzung*.
Die Abbildung $\hat{\theta}(x_1, \ldots, x_n)$ heißt *Maximum-Likelihood-Schätzer*.

Maximum Likelihood

Bemerkung 29

- Die ML-Methode liefert oft gute Schätzer
- Unter gewissen Annahmen, ist der ML-Schätzer
  - asymptotisch erwartungstreu, d.h. Bias $\xrightarrow{n\to\infty} 0$
  - konsistent
  - asymptotisch normalverteilt
- Oft ist es schwierig die Likelihood-Funktion $L(\theta)$ zu maximieren
- Stattdessen maximieren wir die *Log-Likelihood-Funktion* $\ell(\theta) := \log\big(L(\theta)\big)$

$$\ell(\theta) = \begin{cases} \sum_{i=1}^n \log\big(\mathbb{P}_\theta(X_1 = x_1)\log\big) & \text{für diskrete ZV} \\ \sum_{i=1}^n \log\big(f_\theta(x_i)\big) & \text{für stetige ZV} \end{cases}$$

Maximum Likelihood

Beispiel 142: (n-facher Münzwurf - Fortsetzung)

- Sei die Stichprobe $x_1, \ldots, x_n$ gegeben, wobei k Mal Kopf und $n-k$ Mal Zahl geworfen wurde
- Damit ist $L(p) = p^k(1-p)^{n-k}$ und

$$\ell(p) = \log\left(p^k(1-p)^{n-k}\right) = k\log(p) + (n-k)\log(1-p)$$

- Berechne Ableitungen

$$\ell'(p) = \tfrac{k}{p} - \tfrac{n-k}{1-p}$$
$$\ell''(p) = -\tfrac{k}{p^2} - \tfrac{n-k}{(1-p)^2}$$

- $\ell'(p) = 0 \iff \frac{k}{p} = \frac{n-k}{1-p} \iff k(1-p) = (n-k)p \iff k = np \iff p = \frac{k}{n}$
- $\ell''(p) < 0$, also maximiert $\hat{p} = \frac{k}{n}$ die (Log-)Likelihood-Funktion

Maximum Likelihood

Beispiel 143

Die Wartezeiten an einem Fahrkartenautomaten in Minuten seien unabhängig und Exponential-verteilt mit Dichte $f_\lambda(x) = \lambda e^{-\lambda x}$, für $x \geq 0$ und $\lambda > 0$. Es wurden folgende Wartezeiten gemessen (in Minuten): $2.4, 1.1, 0.9, 3.0, 1.6, 2.1, 0.7, 1.8$.

> Was ist der ML-Schätzer? Was ist die ML-Schätzung?

> Bestimme Log-Likelihood-Funktion und Ableitungen:

$$\ell(\lambda) = \sum_{i=1}^{n} \log\big(f_\lambda(x_i)\big) = \sum_{i=1}^{n} \log(\lambda e^{-\lambda x}) = n\log(\lambda) - \lambda \sum_{i=1}^{n} x_i$$

$$\ell'(\lambda) = \tfrac{n}{\lambda} - \sum_{i=1}^{n} x_i$$

$$\ell''(\lambda) = -\tfrac{n}{\lambda^2} < 0$$

Maximum Likelihood

Beispiel 143: (Fortsetzung)

Die Wartezeiten an einem Fahrkartenautomaten in Minuten seien unabhängig und Exponential-verteilt mit Dichte $f_\lambda(x) = \lambda e^{-\lambda x}$, für $x \geq 0$ und $\lambda > 0$. Es wurden folgende Wartezeiten gemessen (in Minuten): $2.4, 1.1, 0.9, 3.0, 1.6, 2.1, 0.7, 1.8$.

> Was ist der ML-Schätzer? Was ist die ML-Schätzung?

> Ableitung: $\ell'(\lambda) = \frac{n}{\lambda} - \sum_{i=1}^n x_i$

> Setze Ableitung gleich 0:

$$\ell'(\lambda) = 0 \iff \frac{n}{\lambda} = \sum_{i=1}^n x_i \iff \lambda = \frac{1}{\frac{1}{n}\sum_{i=1}^n x_i} = \frac{1}{\bar{X}_n}$$

> $\hat{\lambda} = \frac{1}{\bar{X}_n}$ ist der ML-Schätzer.

> Für die Werte ergibt sich $\bar{X}_n = 1.7$, also $\hat{\lambda} = \frac{10}{17}$.

Maximum Likelihood

Übung 87

Alemannia Aachen hat in den ersten 10 Spielen der Saison 2024/2025 jeweils $2, 1, 1, 1, 1, 0, 0, 0, 1, 2$ Tore geschossen. Die Anzahl der Tore je Spiel sei Poisson-verteilt mit Parameter λ, d.h.

$$\mathbb{P}(X = k) = e^{-\lambda}\frac{\lambda^k}{k!}.$$

Bestimmen Sie mit der ML-Methode einen Schätzer für λ. Berechnen Sie anschließend die ML-Schätzung für die gegebene Stichprobe.

λ entspricht dem Erwartungswert der Tore je Spiel. Was ist die geschätzte Anzahl an Toren je Spiel?

Maximum Likelihood

> Seien $X_1, \ldots, X_n$ unabhängige Zufallsvariablen mit Verteilung $\mathcal{N}(\mu, \sigma^2)$

> Sei $x_1, \ldots, x_n$ eine zugehörige Stichprobe

> Was sind die ML-Schätzer für μ bzw. σ^2, falls

1. $\sigma^2 = \sigma_0^2$ bekannt ist,
2. $\mu = \mu_0$ bekannt ist,
3. μ und σ^2 unbekannt sind

Maximum Likelihood

ML Schätzung für die Normalverteilung

- Seien $X_1, \dots, X_n$ unabhängige Zufallsvariablen mit Verteilung $\mathcal{N}(\mu, \sigma^2)$ mit **bekannter** Varianz $\sigma^2 = \sigma_0^2$
- Sei $x_1, \dots, x_n$ eine zugehörige Stichprobe
- Die Log-Likelihood-Funktion ist

$$\begin{aligned}\ell(\mu) &= \sum_{i=1}^{n} \log(f(x_i)) = \sum_{i=1}^{n} \log\left(\frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\left(-\frac{(x_i-\mu)^2}{2\sigma_0^2}\right)\right) \\ &= -\tfrac{n}{2}\log\left(2\pi\sigma_0^2\right) - \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma_0^2}\end{aligned}$$

- $\ell(\mu)$ ist maximal $\iff \sum_{i=1}^{n}(x_i-\mu)^2$ ist minimal $\iff \mu = \bar{x}_n$

Maximum Likelihood

ML Schätzung für die Normalverteilung

- Seien $X_1, \ldots, X_n$ unabhängige Zufallsvariablen mit Verteilung $\mathcal{N}(\mu, \sigma^2)$ mit **bekanntem** Erwartungswert $\mu = \mu_0$
- Sei $x_1, \ldots, x_n$ eine zugehörige Stichprobe
- Die Log-Likelihood-Funktion ist
$$\ell(\sigma^2) = -\tfrac{n}{2} \log\left(2\pi\sigma^2\right) - \frac{\sum_{i=1}^n (x_i - \mu_0)^2}{2\sigma^2}$$
- Ableiten ergibt
$$\ell'(\sigma^2) = -\frac{n}{2\sigma^2} + \frac{\sum_{i=1}^n (x_i - \mu_0)^2}{2\sigma^4}$$
$$\ell''(\sigma^2) = \frac{n}{2\sigma^4} - \frac{\sum_{i=1}^n (x_i - \mu_0)^2}{\sigma^6}$$
- $\ell'(\sigma^2) = 0 \iff \frac{n}{2\sigma^2} = \frac{\sum_{i=1}^n (x_i - \mu_0)^2}{2\sigma^4} \iff \sigma^2 = \frac{1}{n}\sum_{i=1}^n (x_i - \mu_0)^2$

Maximum Likelihood

ML Schätzung für die Normalverteilung

- Seien $X_1, \ldots, X_n$ unabhängige Zufallsvariablen mit Verteilung $\mathcal{N}(\mu, \sigma^2)$
- Sei $x_1, \ldots, x_n$ eine zugehörige Stichprobe
- Die Log-Likelihood-Funktion ist

$$\ell(\mu, \sigma^2) = -\tfrac{n}{2} \log\left(2\pi\sigma^2\right) - \frac{\sum_{i=1}^n (x_i - \mu)^2}{2\sigma^2}$$

- $\ell(\mu, \sigma^2)$ kann analog zu 1. und 2. minimiert werden
  - Dafür: Berechne Gradient (1. Ableitung) und Hesse-Matrix (2. Ableitung), setze Gradient gleich Null und löse Gleichungssystem
- Ergebnis: $\hat{\mu} = \bar{x}_n$ und $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \mu_0)^2$

Münzwurf

Beispiel 144

- Wir werfen eine Münze 100 Mal und erhalten 33 Mal Kopf und 67 Mal Zahl
- Wir können die Wahrscheinlichkeit $p = \mathbb{P}(\text{“Kopf”})$ schätzen:

$$\hat{p} = \frac{33}{100}$$

- $\hat{p}$ ist erwartungstreu und konsistent
- **Aber:** mit hoher Wahrscheinlichkeit ist $\hat{p} \neq p$ (“nur” $\hat{p} \approx p$)
- Wie können wir das Problem umgehen?
  - Angabe eines Intervalls $[p_0, p_1]$, sodass p mit hoher Wahrscheinlichkeit in diesem Intervall liegt
  - “Konfidenzintervall” statt “Punktschätzer”

Münzwurf

Beispiel 144: (Fortsetzung)

- Sei X_i der Indikator für “Kopf” im i-ten Münzwurf, d.h. $X_i \sim Ber(p)$
- Dann gilt $\hat{p} = \bar{X}_n$
- Weiter gilt $\mathbb{E}[X_i] = p$ und $\mathrm{var}(X_i) = p(1-p)$
- Aus dem zentralen Grenzwertsatz folgt

$$\mathbb{P}\left(\sqrt{n}\frac{\bar{X}_n - \mathbb{E}[X_i]}{\mathrm{var}(X_i)} \leq x\right) = \mathbb{P}\left(\sqrt{n}\frac{\hat{p}-p}{p(1-p)} \leq x\right) \xrightarrow{n\to\infty} \Phi(x)$$

- Aus der Konsistenz von $\hat{p}$ folgt, für großes n,

$$\mathbb{P}\left(\sqrt{n}\frac{\hat{p}-p}{\hat{p}(1-\hat{p})} \leq x\right) \approx \Phi(x)$$

Münzwurf

Beispiel 144: (Fortsetzung)

> Sei $\alpha \in (0,1)$ und $q = q_{1-\frac{\alpha}{2}}, -q = q_{\frac{\alpha}{2}}$ jeweils das $(1-\frac{\alpha}{2})$– bzw. $\frac{\alpha}{2}$-Quantil der Standardnormalverteilung

> Dann gilt

$$\mathbb{P}\left(\hat{p} - q\frac{\hat{p}(1-\hat{p})}{\sqrt{n}} \leq p \leq \hat{p} + q\frac{\hat{p}(1-\hat{p})}{\sqrt{n}}\right) = \Phi(q) - \Phi(-q) = 1 - \alpha$$

> Mit Wahrscheinlichkeit $1-\alpha$ liegt das wahre p in

$$\left[\hat{p} - q\frac{\hat{p}(1-\hat{p})}{\sqrt{n}}, \hat{p} + q\frac{\hat{p}(1-\hat{p})}{\sqrt{n}}\right]$$

> Konkret:
> > Für $\alpha = 0.05$ ist $-q_{\frac{\alpha}{2}} = q_{1-\frac{\alpha}{2}} = 1.96$
> > Falls $\hat{p} = \frac{33}{100}$ und $n = 100$, liegt p mit 95% Wahrscheinlichkeit in dem Intervall $[0.2867, 0.3733]$

Konfidenzintervall

Definition 65

Sei $X = (X_1, \dots, X_n)^T : \Omega \to \mathcal{X}$ eine Zufallsvariable mit statistischem Modell $\mathcal{P} = \{\mathbb{P}_\theta^n : \theta \in \Theta\}$. Weiter sei $x_1, \dots, x_n$ eine Stichprobe und $\alpha \in [0, 1]$. Eine Abbildung I, die jeder Stichprobe $x = (x_1, \dots, x_n)^T \in \mathcal{X}$ eine Teilmenge $I(x) \subset \Theta$ zuordnet, heißt $(1-\alpha)$-*Konfidenzbereich* (auch *Vertrauensbereich*) für θ, falls

$$\mathbb{P}_\theta(\theta \in I(X_1, \dots, X_n)) \geq 1 - \alpha$$

für alle $\theta \in \Theta$. Falls $\Theta \subset \mathbb{R}$ und $I(x)$ für alle $x \in \mathcal{X}$ ein Intervall ist, heißt I ein $(1-\alpha)$-*Konfidenzintervall*.

Konfidenzintervall

Bemerkung 30

> α ist die Fehlerwahrscheinlichkeit, mit der θ nicht in I liegt

> Je kleiner die Fehlertoleranz α, desto größer I

> Falls $\alpha = 1$, reicht ein Punktschätzer $\hat{\theta}$

> Falls $\alpha = 0$, ist $I = \Theta$ (oft ist $\Theta = \mathbb{R}$)

> Konfidenzintervalle sind nicht eindeutig
>
> > Zu jedem α kann es viele Konfidenzintervalle geben
> >
> > Wir erhalten Eindeutigkeit durch Zentrierung bzw. indem wir einseitige Intervalle betrachten

Konfidenzintervall

Beispiel 145

Seien $X_1, \dots, X_n \sim \mathcal{N}(\mu, \sigma^2)$ und σ^2 bekannt. Wie finden wir ein Konfidenzintervall für μ?

> Sei $q = q_{1-\frac{\alpha}{2}}$ das $(1 - \frac{\alpha}{2})$-Quantil von $\mathcal{N}(0,1)$, dann $q_{\frac{\alpha}{2}} = -q$

> Für $Z \sim \mathcal{N}(0,1)$ gilt $\mathbb{P}(-q \leq Z \leq q) = \Phi(q) - \Phi(-q) = 1 - \alpha$

> Wir wissen, dass $\bar{X}_n \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$, insbes. $\sqrt{n}\frac{\bar{X}_n - \mu}{\sigma} \sim \mathcal{N}(0,1)$

> Damit folgt

$$1 - \alpha = \mathbb{P}\Big(-q \leq \sqrt{n}\frac{\bar{X}_n - \mu}{\sigma} \leq q\Big) = \mathbb{P}\Big(-\frac{q\sigma}{\sqrt{n}} - \bar{X}_n \leq \mu \leq \frac{q\sigma}{\sqrt{n}} - \bar{X}_n\Big)$$
$$= \mathbb{P}\Big(\bar{X}_n + \frac{q\sigma}{\sqrt{n}} \geq \mu \geq \bar{X}_n - \frac{q\sigma}{\sqrt{n}}\Big)$$

> Wir erhalten ein *zweiseitiges* $(1 - \alpha)$-Konfidenzintervall für μ:

$$\Big[\bar{X}_n - \frac{q\sigma}{\sqrt{n}}, \bar{X}_n + \frac{q\sigma}{\sqrt{n}}\Big]$$

Konfidenzintervall

Beispiel 145: (Fortsetzung)

- Seien $X_1, \dots, X_n \sim \mathcal{N}(\mu, \sigma^2)$ und σ^2 bekannt
- *Zweiseitiges* $(1-\alpha)$-Konfidenzintervall für μ

$$\left[\bar{X}_n - \frac{q_{1-\alpha/2}\sigma}{\sqrt{n}}, \bar{X}_n + \frac{q_{1-\alpha/2}\sigma}{\sqrt{n}}\right]$$

- Analog erhalten wir das *nach oben beschränkte, einseitige Konfidenzintervall* (kurz: *unteres Konfidenzintervall*)

$$\left(-\infty, \bar{X}_n + \frac{q_{1-\alpha}\sigma}{\sqrt{n}}\right]$$

und das *nach unten beschränkte, einseitige Konfidenzintervall* (kurz: *oberes Konfidenzintervall*)

$$\left[\bar{X}_n - \frac{q_{1-\alpha}\sigma}{\sqrt{n}}, \infty\right)$$

Konfidenzintervall

Beispiel 146

Seien $X_1, \ldots, X_n \sim \mathcal{N}(\mu, \sigma^2)$ und σ^2 **nicht** bekannt. Wie finden wir ein Konfidenzintervall für μ?

> Idee: Nutze Konfidenzintervall für bekanntes σ^2 und ersetze σ^2 durch die Stichprobenvarianz $s_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$

> Achtung: Die Verteilung ändert sich

$$\sqrt{n} \frac{\bar{X}_n - \mu}{s_n} \sim t_{n-1}$$

t_{n-1} bezeichnet die t-*Verteilung mit* $n-1$ *Freiheitsgraden*

> Die t_n-Verteilung ist symmetrisch zur y-Achse und approximiert die Normalverteilung: $t_{n-1} \approx \mathcal{N}(0,1)$ für n groß

> Die Werte der t_n-Verteilung sind tabelliert

Konfidenzintervall

Beispiel 146: (Fortsetzung)

Seien $X_1, \dots, X_n \sim \mathcal{N}(\mu, \sigma^2)$ und σ^2 **nicht** bekannt. Wie finden wir ein Konfidenzintervall für μ?

> Sei $t_{n-1,1-\alpha/2}$ das $(1-\alpha/2)$-Quantil der t_{n-1}-Verteilung, dann ist

$$\left[\bar{X}_n - \frac{t_{n-1,1-\alpha/2} s_n}{\sqrt{n}}, \bar{X}_n + \frac{t_{n-1,1-\alpha/2} s_n}{\sqrt{n}}\right]$$

ein $(1-\alpha)$-Konfidenzintervall für μ

> Wir erhalten wieder die einseitigen Konfidenzintervalle

$$\left(-\infty, \bar{X}_n + \frac{t_{n-1,1-\alpha} s_n}{\sqrt{n}}\right] \quad \text{und} \quad \left[\bar{X}_n - \frac{t_{n-1,1-\alpha} s_n}{\sqrt{n}}, \infty\right)$$

Konfidenzintervall

Beispiel 147

Seien $X_1, \dots, X_n \sim \mathcal{N}(\mu, \sigma^2)$. Wie finden wir ein Konfidenzintervall für σ^2?

> Die Stichprobenvarianz ist χ^2-verteilt:

$$\frac{(n-1)s_n^2}{\sigma^2} \sim \chi^2_{n-1} \quad \text{und} \quad \frac{\sum_{i=1}^n (X_i - \mu)^2}{\sigma^2} \sim \chi^2_n$$

> Falls der Erwartungswert bekannt ist, nutze χ^2_n- ansonsten χ^2_{n-1}-Verteilung

> Die Werte der χ^2_n-Verteilung sind tabelliert

> Sei m die Anzahl der Freiheitsgrade ($m = n$ falls μ bekannt, $m = n - 1$ falls μ unbekannt) und

$$\hat{\sigma}_n^2 = \begin{cases} s_n^2 & \text{falls } \mu \text{ unbekannt} \\ \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2 & \text{falls } \mu \text{ bekannt} \end{cases}$$

Konfidenzintervall

Beispiel 147

$X_1, \ldots, X_n \sim \mathcal{N}(\mu, \sigma^2)$. Wie finden wir ein Konfidenzintervall für σ^2?

> Seien $\chi^2_{m,1-\alpha/2}$ und $\chi^2_{m,\alpha/2}$ das $(1-\frac{\alpha}{2})$- bzw. $\frac{\alpha}{2}$-Quantil der χ^2_m-Verteilung

> Dann gilt

$$\begin{aligned} 1-\alpha &= F_{\chi^2_m}(\chi^2_{m,1-\alpha/2}) - F_{\chi^2_m}(\chi^2_{m,\alpha/2}) \\ &= \mathbb{P}\left(\chi^2_{m,\alpha/2} \leq \frac{m\hat{\sigma}^2_n}{\sigma^2} \leq \chi^2_{m,1-\alpha/2}\right) = \mathbb{P}\left(\frac{m\hat{\sigma}^2_n}{\chi^2_{m,1-\alpha/2}} \leq \sigma^2 \leq \frac{m\hat{\sigma}^2_n}{\chi^2_{m,\alpha/2}}\right) \end{aligned}$$

> Bei unbekanntem bzw. bekanntem μ, erhalten wir das $(1-\alpha)$-Konfidenzintervall für σ^2

$$\left[\frac{(n-1)s^2_n}{\chi^2_{n-1,1-\alpha/2}}, \frac{(n-1)s^2_n}{\chi^2_{n-1,\alpha/2}}\right] \quad \text{bzw.} \quad \left[\frac{\sum_{i=1}^n (X_i-\mu)^2}{\chi^2_{n,1-\alpha/2}}, \frac{\sum_{i=1}^n (X_i-\mu)^2}{\chi^2_{n,\alpha/2}}\right]$$

t- und χ^2-Verteilung

Optional 9

Wie jede andere Verteilung auch, können t- und χ^2-Verteilung über ihre Dichtefunktion definiert werden. Da die Verteilungen jedoch häufig im Zusammenhang mit der Normalverteilung auftreten, werden sie oft anders definiert.

1. Seien $X_1, \dots, X_n \sim \mathcal{N}(0,1)$ unabhängige Zufallsvariablen. Dann ist
$$X := X_1^2 + \cdots + X_n^2 = \sum_{i=1}^{n} X_i^2$$
Chi-Quadrat-verteilt mit n Freiheitsgraden (kurz χ_n^2-verteilt).

2. Seien $X \sim \mathcal{N}(0,1)$ und $Y \sim \chi_n^2$ unabhängige Zufallsvariablen. Dann ist
$$Z = \frac{X}{\sqrt{Y/n}}$$
t-verteilt mit n Freiheitsgraden.

Konfidenzintervall

Übung 88

Angenommen die Körpergröße einer Population folgt einer Normalverteilung $\mathcal{N}(\mu, \sigma^2)$ mit bekannter Varianz $\sigma^2 = 100$. Eine zufällig ausgewählte Stichprobe von 40 Personen hat einen Mittelwert von $\bar{X}_n = 173\,$cm und eine empirische Standardabweichung von $s_n = 9\,$cm.

1. Bestimmen Sie das 95%- und 99%-Konfidenzintervall für μ.
2. Bestimmen Sie das 95%- und 99%-Konfidenzintervall für μ unter der Voraussetzung, dass die Varianz σ^2 unbekannt ist.

Literatur I

Bundesministerium für Gesundheit (2024).
Infektionsradar.
https://infektionsradar.gesund.bund.de/de/covid/inzidenz.
Abgerufen: 2024-10-15.

Dehling, H. and Haupt, B. (2006).
Einführung in die Wahrscheinlichkeitstheorie und Statistik.
Springer-Verlag.

FH AACHEN
UNIVERSITY OF APPLIED SCIENCES

Literatur II

Dinnes, J., Sharma, P., Berhane, S., van Wyk, S., Nyaaba, N., Domen, J., Taylor, M., Cunningham, J., Davenport, C., Dittrich, S., Emperador, D., Hooft, L., Leeflang, M., McInnes, M., Spijker, R., Verbakel, J., Takwoingi, Y., Taylor-Phillips, S., Van den Bruel, A., and Deeks, J. (2022).
Rapid, point-of-care antigen tests for diagnosis of sars-cov-2 infection.
Cochrane Database of Systematic Reviews, (7).

Henze, N. et al. (1997).
Stochastik für Einsteiger, volume 4.
Springer.

FH AACHEN UNIVERSITY OF APPLIED SCIENCES