Inferenzstatistik

BSc Psychologie SoSe 2021

Prof. Dr. Dirk Ostwald

(7) Statistische Modelle, Statistiken und Schätzer

Literaturhinweise

Die in dieser Einheit behandelten Konzepte sind grundlegend und werden unter anderem in Georgii (2009, Kapitel 7) diskutiert.

Statistische Modelle, Statistiken und Schätzer

- Statistische Modelle
- Statistiken und Schätzer
- Standardprobleme frequentistischer Inferenz
- Maximum-Likelihood Schätzer
- Selbstkontrollfragen

Statistische Modelle, Statistiken und Schätzer

- **Statistische Modelle**
- Statistiken und Schätzer
- Standardprobleme frequentistischer Inferenz
- Maximum-Likelihood Schätzer
- Selbstkontrollfragen

Definition (Statistisches Modell)

Ein *statistisches Modell* ist ein Tripel

$$\mathcal{M} := (\mathcal{X}, \mathcal{A}, \{\mathbb{P}_\theta | \theta \in \Theta\}) \tag{1}$$

bestehend aus einem *Stichprobenraum* $\mathcal{X}$, einer σ-Algebra $\mathcal{A}$ auf $\mathcal{X}$ und einer mindestens zweielementigen Menge $\{\mathbb{P}_\theta | \theta \in \Theta\}$ von Wahrscheinlichkeitsmaßen auf $(\mathcal{X}, \mathcal{A})$, die mit einer gewissen Indexmenge Θ indiziert sind.

Bemerkungen

- Im Gegensatz zum W-Raum betrachten wir zwei oder mehr W-Maße.
- Das jeweils zugrundeliegende Wahrscheinlichkeitsmaß ist mit $\theta \in \Theta$ indiziert.
- Für Erwartungswerte und (Ko)Varianzen bezüglich $\mathbb{P}_\theta$ schreiben wir $\mathbb{E}_\theta$, $\mathbb{V}_\theta$, $\mathbb{C}_\theta$.

Definition (Parametrische, Standard- und Produktmodelle)

- Ein statistisches Modell $\mathcal{M}$ heißt ein *parametrisches Modell*, wenn $\Theta \subset \mathbb{R}^d$ für ein $d \in \mathbb{N}$ ist. Für $d = 1$ heißt $\mathcal{M}$ ein *einparametriges Modell*.

- Ein statistisches Modell $\mathcal{M}$ heißt ein *diskretes Modell*, wenn $\mathcal{X}$ diskret ist. Jedes $\mathbb{P}_\theta$ besitzt dann eine WMF p_θ definiert durch $p_\theta := \mathbb{P}_\theta(x)$. Ein statistisches Modell $\mathcal{M}$ heißt ein *stetiges Modell*, wenn $\mathcal{X} \subset \mathbb{R}^n$ ist und jedes $\mathbb{P}_\theta$ eine WDF p_θ besitzt.

- Für ein statistisches Modell $\mathcal{M}_0 := (\mathcal{X}_0, \mathcal{A}_0, \{\mathbb{P}_\theta^0 | \theta \in \Theta\})$ und $n > 1$ heißt das statistische Modell $\mathcal{M} := (\mathcal{X}, \mathcal{A}, \{\mathbb{P}_\theta | \theta \in \Theta\})$, für welches $\mathcal{X} := \times_{i=1}^n \mathcal{X}_0$ das n-fache kartesische Produkt von $\mathcal{X}_0$ mit sich selbst ist, $\mathcal{A}$ die entsprechende Produkt-σ-Algebra ist, und $\{\mathbb{P}_\theta | \theta \in \Theta\}$ die entsprechende Menge an Produktmaßen $\prod_{i=1}^n \mathbb{P}_\theta^0$ ist, das zu $\mathcal{M}_0$ gehörige *Produktmodell*.

Bemerkungen

- Der Vorgang des Beobachtens, der ein zufälliges Ergebnis liefert, wird mit Zufallsvektoren X, die Werte in $\mathcal{X}$ annehmen, beschrieben.
- Im Kontext statistischer Modelle nennt man diese Zufallsvektoren *Beobachtung, Messung*, oder *Stichprobe*. Ihre Realisierungen $x \in \mathcal{X}$, also konkret vorliegende Datenwerte, werden *Beobachtungswert, Messwert, Stichprobenwert* oder *Datensatz* genannt.
- Produktmodelle modellieren die unabhängige Wiederholung eines Einzelexperiments. Der entsprechende Zufallsvektor $X := (X_1, ..., X_n)$ entspricht dann einer Menge von unabhängigen Zufallsvariablen/vektoren $X_1, ..., X_n \sim \mathbb{P}_\theta^0$.
- Wenn für ein Produktmodell die Menge $\mathcal{X}_0$ eindimensional ist, also z.B. $\mathcal{X}_0 = \mathbb{R}$ gilt, spricht man von einem *univariaten statistischen Modell*.

Bemerkungen

- Wenn für ein Produktmodell die Menge $\mathcal{X}_0$ mehrdimensional ist, also z.B. $\mathcal{X}_0 = \mathbb{R}^m, m > 1$ ist, spricht man auch von *multivariaten statistischen Modellen*. Wir betrachten in diesem Kurs nur univariate statistische Modelle.
- Die am häufigsten angewendeten statistischen Modelle sind parametrische stetige Produktmodelle.
- In einem konkreten Datenanalyseproblem nimmt man an, dass die beobachteten Stichprobenwerte $x = (x_1, ..., x_n)$ der Stichprobe $X = (X_1, ..., X_n)$ durch $\mathbb{P}_{\tilde{\theta}}$ generiert wurde, wobei $\tilde{\theta}$ hier den *wahren, aber unbekannten, Parameterwert* bezeichnet. Der wahre, aber unbekannten, Parameterwert $\tilde{\theta}$ bleibt auch nach der statistischen Analyse unbekannt.
- In der mathematischen Analyse von Inferenzmethoden betrachtet man alle möglichen wahren, aber unbekannten, Parameterwerte, schreibt also einfach $\{\mathbb{P}_\theta | \theta \in \Theta\}$.

Beispiele

Bernoullimodell

$\mathcal{M} := (\mathcal{X}, \mathcal{A}, \{\mathbb{P}_\theta | \theta \in \Theta\})$ mit $\mathcal{X} := \{0,1\}^n$, $\mathcal{A} := \mathcal{P}\left(\{0,1\}^n\right)$, $\theta := \mu$, $\Theta :=]0,1[$, also

$$\{\mathbb{P}_\theta | \theta \in \Theta\} := \left\{ \prod_{i=1}^{n} \text{Bern}(\mu) | \mu \in]0,1[\right\}, \tag{2}$$

und damit $X_1, ..., X_n \sim$ $\text{Bern}(\mu)$.

Normalverteilungsmodell

$\mathcal{M} := (\mathcal{X}, \mathcal{A}, \{\mathbb{P}_\theta | \theta \in \Theta\})$ mit $\mathcal{X} := \mathbb{R}^n$, $\mathcal{A} := \mathcal{B}(\mathbb{R}^n)$, $\theta := (\mu, \sigma^2)$, $\Theta := \mathbb{R} \times \mathbb{R}_{>0}$, also

$$\{\mathbb{P}_\theta | \theta \in \Theta\} := \left\{ \prod_{i=1}^{n} N(\mu, \sigma^2) | (\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}_{>0} \right\}, \tag{3}$$

und damit $X_1, ..., X_n \sim$ $N(\mu, \sigma^2)$.

Statistische Modelle, Statistiken und Schätzer

- Statistische Modelle
- **Statistiken und Schätzer**
- Standardprobleme frequentistischer Inferenz
- Maximum-Likelihood Schätzer
- Selbstkontrollfragen

Definition (Statistik)

$\mathcal{M} := (\mathcal{X}, \mathcal{A}, \{\mathbb{P}_\theta | \theta \in \Theta\})$ sei ein statistisches Modell und $(\Sigma, \mathcal{S})$ sei ein Messraum. Dann heißt eine Zufallsvariable $S : \mathcal{X} \to \Sigma$ eine *Statistik*.

Bemerkungen

- Beobachtungen und Statistiken werden durch Zufallsvariablen modelliert.
- Beobachtungen modellieren die stochastische Generation von Daten; Statistiken modellieren von Datenanalyst:innen konstruierte Funktionen, die aufgrund von Beobachtungen essentielle Größen extrahieren, aus denen sich sinnvolle Schlüsse ziehen lassen.

Beispiele (Statistiken)

$\mathcal{M}$ sei das Normalverteilungsmodell. Dann sind zum Beispiel

- das *Stichprobenmittel*

$$\bar{x}_n : \mathbb{R}^n \to \mathbb{R}, x \mapsto \bar{x}_n(x) := \frac{1}{n}\sum_{i=1}^{n} x_i, \tag{4}$$

- die *Stichprobenvarianz*

$$s_n^2 : \mathbb{R}^n \to \mathbb{R}_{\geq 0}, x \mapsto s_n^2(x) := \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x}_n)^2, \tag{5}$$

- und die T-*Statistik* für $\mu_0 \in \mathbb{R}$ und Stichprobenstandardabweichung s_n

$$t : \mathbb{R}^n \to \mathbb{R}, x \to t(x) := \sqrt{n}\left(\frac{\bar{x}_n - \mu_0}{s_n}\right) \tag{6}$$

Statistiken.

Definition (Schätzer)

$\mathcal{M} := (\mathcal{X}, \mathcal{A}, \{\mathbb{P}_\theta | \theta \in \Theta\})$, $(\Sigma, \mathcal{S})$ sei ein Messraum, und $\tau : \Theta \to \Sigma$ sei eine Abbildung, die jedem $\theta \in \Theta$ eine Kenngröße $\tau(\theta) \in \Sigma$ zuordnet. Eine Statistik $\hat{\tau} : \mathcal{X} \to \Sigma$ heißt dann ein *Schätzer* für τ.

Bemerkungen

- Typische Beispiele für τ sind
  - $\tau(\theta) := \theta$ für die Schätzung von θ,
  - $\tau(\theta) := \theta_i$ mit $\theta \in \mathbb{R}^d, d > 1$ für die Schätzung einer Komponente von θ,
  - $\tau(\theta) := \mathbb{E}_\theta(X_0)$ für die Schätzung des Erwartungswerts einer Beobachtung,
  - $\tau(\theta) := \mathbb{V}_\theta(X_0)$ für die Schätzung der Varianz einer Beobachtung.
- Schätzer nehmen Zahlwerte in Σ an und heißen deshalb auch *Punktschätzer*.
- Nicht jeder Schätzer ist ein guter Schätzer, man definiert deshalb *Schätzgütekriterien*.
- Für $\hat{\tau}$ bei $\tau(\theta) := \theta$ schreibt man auch $\hat{\theta}$

Beispiele (Schätzer)

$\mathcal{M}$ sei das Normalverteilungsmodell.

- Dann ist zum Beispiel das Stichprobenmittel $\bar{x}_n : \mathbb{R}^n \to \mathbb{R}$ ein Schätzer für

$$\tau : \mathbb{R} \times \mathbb{R}_{>0} \to \mathbb{R}, (\mu, \sigma_2) \mapsto \tau(\mu, \sigma^2) := \mu. \tag{7}$$

  Ebenso ist $\bar{x}_n$ ein Schätzer für

$$\tau : \mathbb{R} \times \mathbb{R}_{>0} \to \mathbb{R}, (\mu, \sigma_2) \mapsto \tau(\mu, \sigma^2) := \mathbb{E}_{\mu,\sigma^2}(X_0) = \mu. \tag{8}$$

- Weiterhin ist die konstante Funktion

$$\hat{\tau} : \mathbb{R}^n \to \mathbb{R}, x \mapsto \hat{\tau}(x) := 42 \tag{9}$$

  ein Schätzer für

$$\tau : \mathbb{R} \times \mathbb{R}_{>0} \to \mathbb{R}_{>0}, (\mu, \sigma_2) \mapsto \tau(\mu, \sigma^2) := \sigma^2. \tag{10}$$

Dass eine Funktion $\hat{\tau} : \mathcal{X} \to \Sigma$ ein Schätzer ist, heißt nicht, dass sie ein guter Schätzer ist!

Gütekriterien für Schätzer sind der Inhalt von Vorlesungseinheit (8) Schätzereigenschaften.

Statistische Modelle, Statistiken und Schätzer

- Statistische Modelle
- Statistiken und Schätzer
- **Standardprobleme frequentistischer Inferenz**
- Maximum-Likelihood Schätzer
- Selbstkontrollfragen

(1) **Parameterschätzung**

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

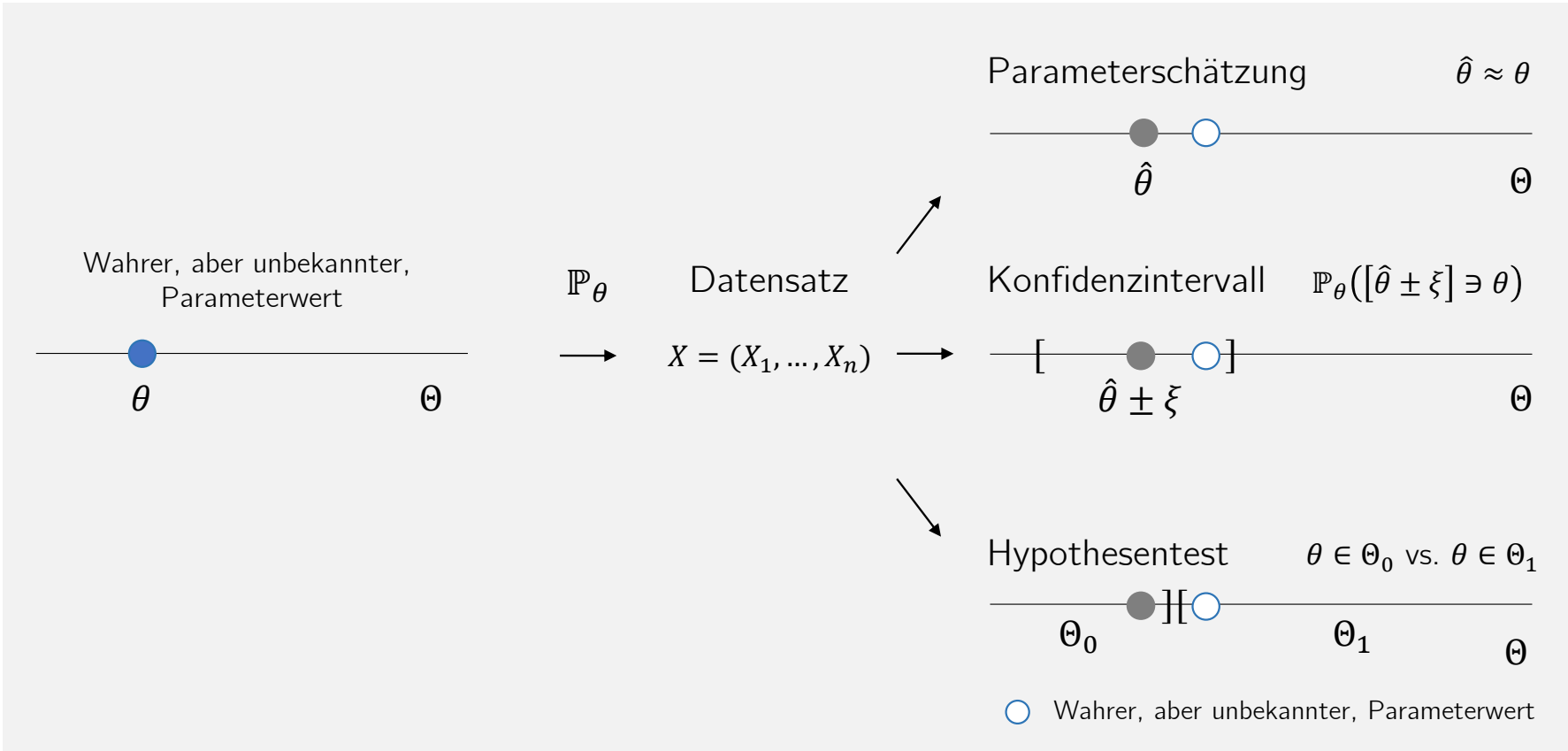
(2) **Konfidenzintervalle**

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

(3) **Hypothesentest**

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

Standardprobleme frequentistischer Inferenz



Die Standardannahmen frequentistischer Inferenz

- $\mathcal{M}$ sei ein statistisches Modell mit $X_1, ..., X_n \sim p_\theta$. Es wird angenommen, dass ein vorliegender Datensatz eine der möglichen Realisierungen von $X_1, ..., X_n \sim p_\theta$ ist.
- Aus frequentistischer Sicht kann man das Experiment unendlich oft wiederholen und zu jedem Datensatz Schätzer oder Statistiken auswerten, z.B. das Stichprobenmittel:

$$\begin{aligned}
&\text{Datensatz (1)} : x^{(1)} = \left(x_1^{(1)}, x_2^{(1)}, ..., x_n^{(1)}\right) \text{ mit } \bar{x}_n^{(1)} = \frac{1}{n}\sum_{i=1}^n x_i^{(1)} \\
&\text{Datensatz (2)} : x^{(2)} = \left(x_1^{(2)}, x_2^{(2)}, ..., x_n^{(2)}\right) \text{ mit } \bar{x}_n^{(2)} = \frac{1}{n}\sum_{i=1}^n x_i^{(2)} \\
&\text{Datensatz (3)} : x^{(3)} = \left(x_1^{(3)}, x_2^{(3)}, ..., x_n^{(3)}\right) \text{ mit } \bar{x}_n^{(3)} = \frac{1}{n}\sum_{i=1}^n x_i^{(3)} \\
&\text{Datensatz (4)} : x^{(4)} = \left(x_1^{(4)}, x_2^{(4)}, ..., x_n^{(4)}\right) \text{ mit } \bar{x}_n^{(4)} = \frac{1}{n}\sum_{i=1}^n x_i^{(4)} \\
&\text{Datensatz (5)} : x^{(5)} = ...
\end{aligned}$$

- Um die Qualität statischer Methoden zu beurteilen betrachtet die frequentistische Statistik deshalb die Wahrscheinlichkeitsverteilungen von Schätzern und Statistiken unter Annahme von $X_1, ..., X_n \sim p_\theta$. Was zum Beispiel ist die Verteilung von $\bar{x}_n^{(1)}$, $\bar{x}_n^{(2)}$, $\bar{x}_n^{(3)}$, $\bar{x}_n^{(4)}$, ... also die Verteilung der Zufallsvariable $\bar{X}_n := \frac{1}{n}\sum_{i=1}^n X_i$?
- Wenn eine statistische Methode im Sinne der frequentitischen Standardannahme gut ist, dann heißt das also, dass sie bei häufiger Anwendung "im Mittel" gut ist. Im Einzelfall, also bei nur einem Datensatz, kann sie auch recht schlecht sein.

Anwendungsbeispiel

Beck Depressions-Inventar II (BDI II)

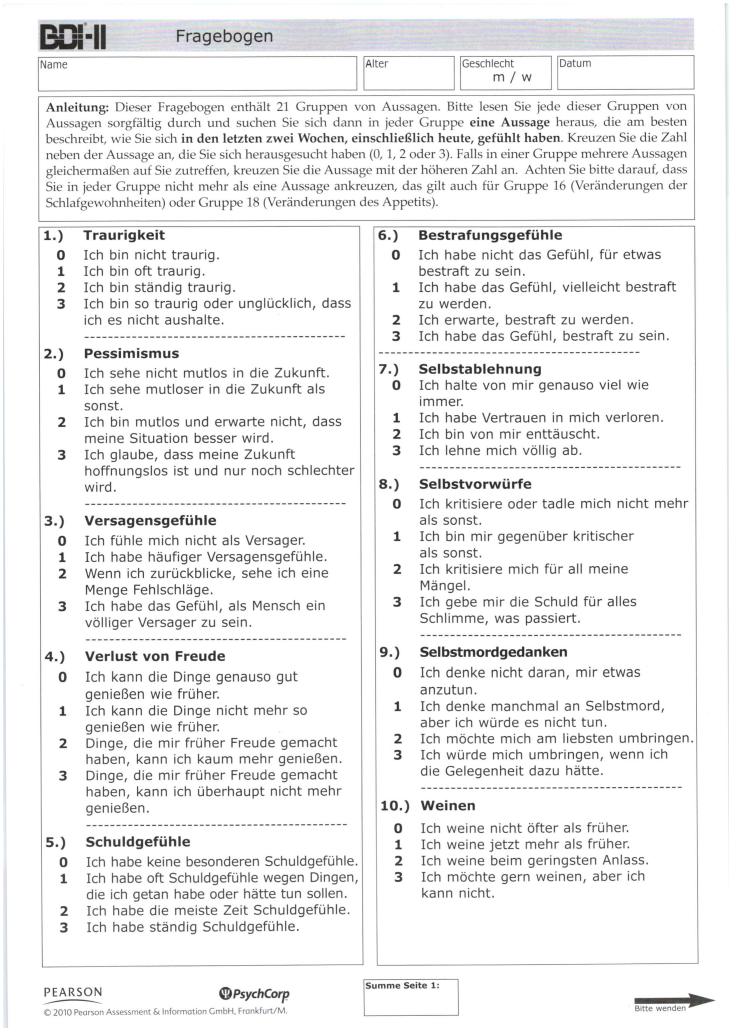
- Weltweit verbreitester Fragebogen zur Selbstbeurteilung des Depressionsschweregrades
- Konzipiert von Beck (1961), seitdem kleinere Revisionen
- 21 Fragen ("Items") mit inhaltlichen Aussagen aufsteigendem Schweregrads (0-3) zu diagnoserelevanten Themengebieten wie Traurigkeit, Versagensgefühle, Schlafverhalten, Selbstmordgedanken, ...
- Beispielitem (14) Wertlosigkeit

  0 Ich fühle mich nicht wertlos.
  1 Ich halte mich für weniger wertvoll und nützlich als sonst.
  2 Verglichen mit anderen Menschen fühle ich mich viel weniger wert.
  3 Ich fühle mich völlig wertlos.

- Auswertung durch Bestimmung des Summenwerts (BDI II Σ) über Items.
- Ergebnisklassifikation: 0-8 keine Depression, 9-13 minimale Depression, 14-19 leichte Depression, 20-28 mittelschwere Depression, 29-63 schwere Depression

Anwendungsbeispiel

Beck Depressions-Inventar II (BDI II)



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:

Bitte wenden

11.) Unruhe
- 0 Ich bin nicht unruhiger als sonst.
- 1 Ich bin unruhiger als sonst.
- 2 Ich bin so unruhig, dass es mir schwerfällt, still zu sitzen.
- 3 Ich bin so unruhig, dass ich mich ständig bewegen oder etwas tun muss.

12.) Interessenverlust
- 0 Ich habe das Interesse an anderen Menschen oder an Tätigkeiten nicht verloren.
- 1 Ich habe weniger Interesse an anderen Menschen oder an Dingen als sonst.
- 2 Ich habe das Interesse an anderen Menschen oder Dingen zum größten Teil verloren.
- 3 Es fällt mir schwer, mich überhaupt für irgend etwas zu interessieren.

13.) Entschlussunfähigkeit
- 0 Ich bin so entschlussfreudig wie immer.
- 1 Es fällt mir schwerer als sonst, Entscheidungen zu treffen.
- 2 Es fällt mir sehr viel schwerer als sonst, Entscheidungen zu treffen.
- 3 Ich habe Mühe, überhaupt Entscheidungen zu treffen.

14.) Wertlosigkeit
- 0 Ich fühle mich nicht wertlos.
- 1 Ich halte mich für weniger wertvoll und nützlich als sonst.
- 2 Verglichen mit anderen Menschen fühle ich mich viel weniger wert.
- 3 Ich fühle mich völlig wertlos.

15.) Energieverlust
- 0 Ich habe so viel Energie wie immer.
- 1 Ich habe weniger Energie als sonst.
- 2 Ich habe so wenig Energie, dass ich kaum noch etwas schaffe.
- 3 Ich habe keine Energie mehr, um überhaupt noch etwas zu tun.

16.) Veränderungen der Schlafgewohnheiten
- 0 Meine Schlafgewohnheiten haben sich nicht verändert.
- 1a Ich schlafe etwas mehr als sonst
- 1b Ich schlafe etwas weniger als sonst.
- 2a Ich schlafe viel mehr als sonst.
- 2b Ich schlafe viel weniger als sonst.
- 3a Ich schlafe fast den ganzen Tag.
- 3b Ich wache 1-2 Stunden früher auf als gewöhnlich und kann dann nicht mehr einschlafen.

17.) Reizbarkeit
- 0 Ich bin nicht reizbarer als sonst.
- 1 Ich bin reizbarer als sonst.
- 2 Ich bin viel reizbarer als sonst.
- 3 Ich fühle mich dauernd gereizt.

18.) Veränderungen des Appetits
- 0 Mein Appetit hat sich nicht verändert.
- 1a Mein Appetit ist etwas schlechter als sonst.
- 1b Mein Appetit ist etwas größer als sonst.
- 2a Mein Appetit ist viel schlechter als sonst.
- 2b Mein Appetit ist viel größer als sonst.
- 3a Ich habe überhaupt keinen Appetit.
- 3b Ich habe ständig Heißhunger.

19.) Konzentrationsschwierigkeiten
- 0 Ich kann mich so gut konzentrieren wie immer.
- 1 Ich kann mich nicht mehr so gut konzentrieren wie sonst.
- 2 Es fällt mir schwer, mich längere Zeit auf irgend etwas zu konzentrieren.
- 3 Ich kann mich überhaupt nicht mehr konzentrieren.

20.) Ermüdung oder Erschöpfung
- 0 Ich fühle mich nicht müder oder erschöpfter als sonst.
- 1 Ich werde schneller müde oder erschöpft als sonst.
- 2 Für viele Dinge, die ich üblicherweise tue, bin ich zu müde oder erschöpft.
- 3 Ich bin so müde oder erschöpft, dass ich fast nichts mehr tun kann.

21.) Verlust an sexuellem Interesse
- 0 Mein Interesse an Sexualität hat sich in letzter Zeit nicht verändert.
- 1 Ich interessiere mich weniger für Sexualität als früher.
- 2 Ich interessiere mich jetzt viel weniger für Sexualität.
- 3 Ich habe das Interesse an Sexualität völlig verloren.

Summe Seite 2:

Übertrag Seite 1: | Gesamt Seite 1+2:

Anwendungsbeispiel

Beck Depressions-Inventar II (BDI II)

VP	BDI II Σ
1	11
2	14
3	10
4	9
5	8
6	17
7	3
8	20
9	13
10	3
11	6
12	2

Modellannahmen

Wir legen das Normalverteilungsmodell zugrunde, d.h. wir nehmen an, dass die BDI II Σ Werte Realisierungen von u.i.v. Zufallsvariablen $X_1, ..., X_{12} \sim N(\mu, \sigma^2)$ sind. Dies entspricht der Annahme, dass sich der BDI II Σ Wert einer VP durch Addition einer normalverteilten Fehlervariable mit Erwartungswertparameter 0 und Varianzparameter σ^2 zu dem über VP identischen Wert μ ergibt.

Fragestellungen

(1) Was sind sinnvolle Tipps für die wahren, aber unbekannten, Parameterwerte μ und σ^2?

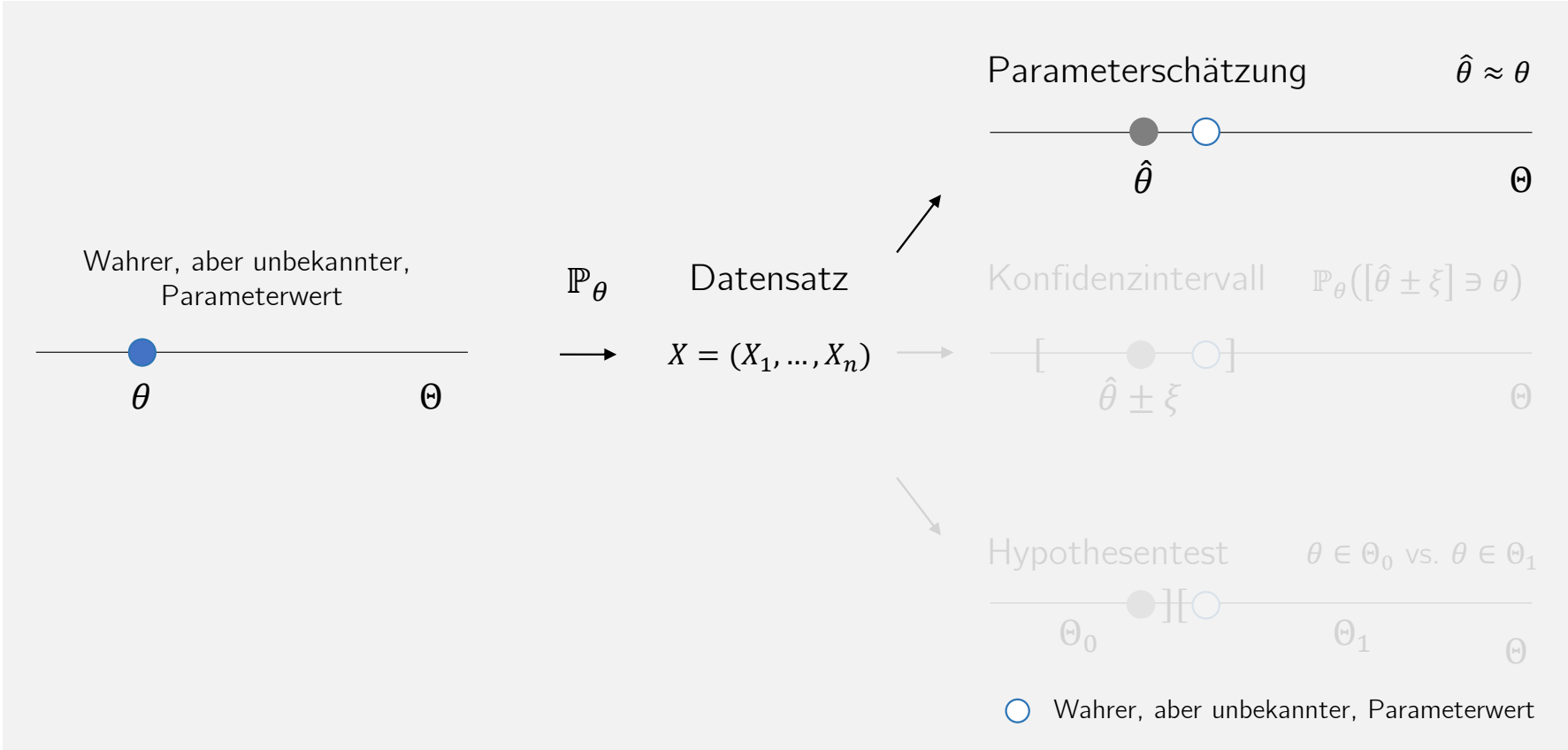
(2) Wie hoch ist im frequentistischen Sinn die mit diesen Tipps assoziierte Unsicherheit?

(3) Entscheiden wir uns sinnvollerweise für die Hypothese, dass $\mu \leq 8$ ist oder dass $\mu > 8$ ist?

Statistische Modelle, Statistiken und Schätzer

- Statistische Modelle
- Statistiken und Schätzer
- Standardprobleme frequentistischer Inferenz
- **Maximum-Likelihood Schätzer**
- Selbstkontrollfragen

Maximum-Likelihood Schätzer



Definition (Parameterpunktschätzer)

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

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

Alternative Visualisierung der Parameterpunktschätzung (Georgii, 2009, Kapitel 7)

- Man beachte, dass hier die Notation $\vartheta = \theta$, $P_\vartheta = \mathbb{P}_\theta$, $T(x) = \hat{\theta}(x)$ benutzt wird.

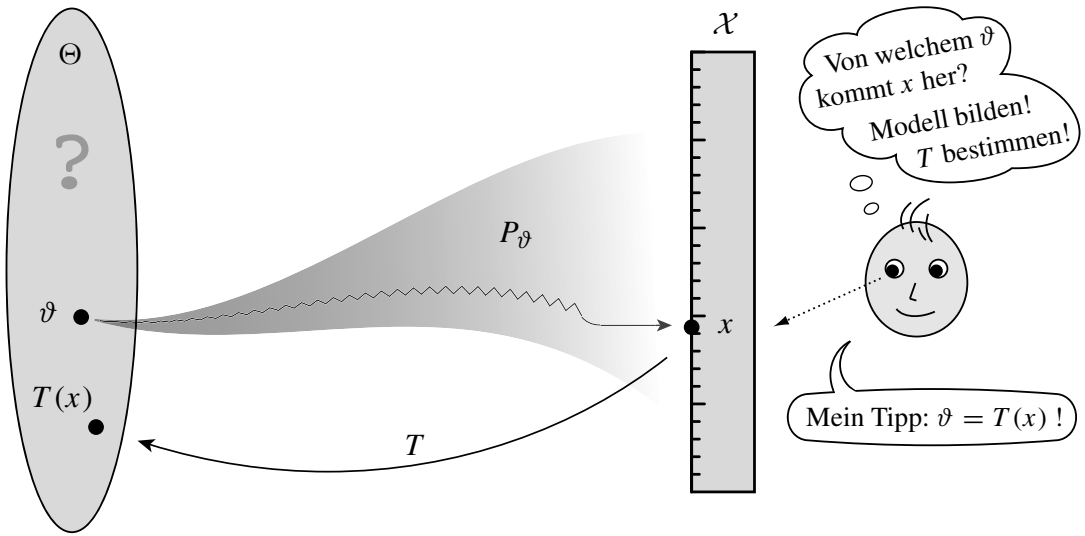


Abbildung 7.1: Das Prinzip des Schätzens: Eine unbekannte Größe ϑ soll ermittelt werden; die möglichen ϑ's bilden eine Menge Θ. Das tatsächlich vorliegende ϑ bestimmt ein Zufallsgesetz P_ϑ, welches das Zufallsexperiment steuert und im konkreten Fall zum Messwert x führt. Um auf ϑ zurückschließen zu können, muss der Statistiker zunächst für jedes ϑ das Zufallsgesetz P_ϑ analysieren – das geschieht bei der Modellbildung. Die ermittelte Struktur aller P_ϑ's muss er dann ausnutzen, um eine Abbildung $T : \mathcal{X} \to \Theta$ so geschickt zu konstruieren, dass der Wert $T(x)$ dem tatsächlich vorliegenden ϑ typischerweise möglichst nahe kommt – ganz egal, welches ϑ das nun wirklich ist.

Prinzipien zur Gewinnung von Parameterschätzern

- Die Definition eines Parameterschätzers macht keine Aussage darüber, wie man Parameterschätzer findet. Zur Gewinnung von Parameterschätzern haben sich deshalb verschiedene Prinzipien etabliert.

- Populäre Methoden 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*.

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

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

$$L_n : \Theta \to [0, \infty[, \theta \mapsto L_n(\theta) := \prod_{i=1}^{n} p_\theta(x_i). \tag{11}$$

Die *Log-Likelihood-Funktion* ist definiert als

$$\ell_n : \Theta \to \mathbb{R}, \theta \mapsto \ell_n(\theta) := \ln L_n(\theta). \tag{12}$$

Bemerkungen

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

Definition (Maximum-Likelihood Schätzer)

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

$$\hat{\theta}_n^{\mathsf{ML}} : \mathcal{X} \to \Theta, x \mapsto \hat{\theta}_n^{\mathsf{ML}}(x) := \arg\max_{\theta \in \Theta} L_n(\theta) = \arg\max_{\theta \in \Theta} \ell_n(\theta). \tag{13}$$

Bemerkungen

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

Maximum-Likelihood Schätzer

Maximum-Likelihood Schätzung für parametrische statistische Produktmodelle

(1) Formulierung der Log-Likelihood-Funktion.

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

(3) Auflösen nach potentiellen Maximumstellen.

Dabei nutzt man typischerweise

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

Beispiel (Bernoullimodell)

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

(1) Formulierung der Log-Likelihood-Funktion

Es gilt

$$L_n :]0,1[\to]0,1[, \mu \mapsto L_n(\mu) := \prod_{i=1}^{n} \mu^{x_i}(1-\mu)^{1-x_i} = \mu^{\sum_{i=1}^{n} x_i}(1-\mu)^{n-\sum_{i=1}^{n} x_i}. \quad (14)$$

Logarithmieren ergibt

$$\ell_n :]0,1[\to \mathbb{R}, \mu \mapsto \ell_n(\mu) = \ln \mu \sum_{i=1}^{n} x_i + \ln(1-\mu)\left(n - \sum_{i=1}^{n} x_i\right). \quad (15)$$

Beispiel (Bernoullimodell)

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

$$\begin{aligned}\frac{d}{d\mu}\ell_n(\mu) &= \frac{d}{d\mu}\left(\ln\mu\sum_{i=1}^{n}x_i + \ln(1-\mu)\left(n-\sum_{i=1}^{n}x_i\right)\right)\\ &= \frac{d}{d\mu}\ln\mu\sum_{i=1}^{n}x_i + \frac{d}{d\mu}\ln(1-\mu)\left(n-\sum_{i=1}^{n}x_i\right) \qquad (16)\\ &= \frac{1}{\mu}\sum_{i=1}^{n}x_i - \frac{1}{1-\mu}\left(n-\sum_{i=1}^{n}x_i\right).\end{aligned}$$

Die *Maximum-Likelihood Gleichung* ergibt sich also zu

$$\frac{1}{\hat{\mu}_n^{\mathsf{ML}}}\sum_{i=1}^{n}x_i - \frac{1}{1-\hat{\mu}_n^{\mathsf{ML}}}\left(n-\sum_{i=1}^{n}x_i\right) = 0. \qquad (17)$$

Beispiel (Bernoullimodell)

(3) Auflösen nach potentiellen Maximumstellen

Es gilt

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

$$\Leftrightarrow \hat{\mu}_n^{\mathsf{ML}}(1-\hat{\mu}_n^{\mathsf{ML}}) \left(\frac{1}{\hat{\mu}_n^{\mathsf{ML}}} \sum_{i=1}^{n} x_i - \frac{1}{1-\hat{\mu}_n^{\mathsf{ML}}} \left(n - \sum_{i=1}^{n} x_i \right) \right) = 0$$

$$\Leftrightarrow \sum_{i=1}^{n} x_i - \hat{\mu}_n^{\mathsf{ML}} \sum_{i=1}^{n} x_i - n\hat{\mu}_n^{\mathsf{ML}} + \hat{\mu}_n^{\mathsf{ML}} \sum_{i=1}^{n} x_i = 0 \tag{18}$$

$$\Leftrightarrow n\hat{\mu}_n^{\mathsf{ML}} = \sum_{i=1}^{n} x_i$$

$$\Leftrightarrow \hat{\mu}_n^{\mathsf{ML}} = \frac{1}{n} \sum_{i=1}^{n} x_i .$$

Beispiel (Bernoullimodell)

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

$$\hat{\mu}_n^{\mathsf{ML}} : \{0,1\}^n \to [0,1], x \mapsto \hat{\mu}_n^{\mathsf{ML}}(x) := \frac{1}{n}\sum_{i=1}^{n} x_i \tag{19}$$

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

Beispiel (Normalverteilungsmodell)

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

(1) Formulierung der Log-Likelihood-Funktion

Es gilt

$$\begin{aligned} L_n : \mathbb{R} \times \mathbb{R}_{>0} \to \mathbb{R}_{>0}, (\mu, \sigma^2) \mapsto L_n(\mu, \sigma^2) &:= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x_i - \mu)^2\right) \\ &= \left(2\pi\sigma^2\right)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2\right). \end{aligned} \tag{20}$$

Logarithmieren ergibt

$$\ell_n : \mathbb{R} \times \mathbb{R}_{>0} \to \mathbb{R}, (\mu, \sigma^2) \mapsto \ell_n(\mu, \sigma^2) = -\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2. \tag{21}$$

Beispiel (Normalverteilungsmodell)

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

$$\frac{d}{d\mu}\ell_n(\mu) = -\frac{d}{d\mu}\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2 = -\frac{1}{2\sigma^2}\sum_{i=1}^{n}\frac{d}{d\mu}(x_i-\mu)^2 = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu). \tag{22}$$

Weiterhin ergibt sich

$$\frac{d}{d\sigma^2}\ell_n(\sigma^2) = -\frac{n}{2}\frac{d}{d\sigma^2}\ln\sigma^2 - \frac{d}{d\sigma^2}\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2 = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i-\mu)^2. \tag{23}$$

Die Maximum-Likelihood Gleichungen haben also die Form

$$\begin{aligned}\sum_{i=1}^{n}(x_i-\hat{\mu}_n^{\mathsf{ML}}) &= 0\\ -\frac{n}{2\hat{\sigma}_n^{2\mathsf{ML}}} + \frac{1}{2\hat{\sigma}_n^{4\mathsf{ML}}}\sum_{i=1}^{n}(x_i-\mu)^2 &= 0\end{aligned} \tag{24}$$

Beispiel (Normalverteilungsmodell)

(3) Auflösen nach potentiellen Maximumstellen Es ergibt sich

$$\sum_{i=1}^{n}(x_i - \hat{\mu}_n^{\mathsf{ML}}) = 0 \Leftrightarrow \sum_{i=1}^{n} x_i = n\hat{\mu}_n^{\mathsf{ML}} \Leftrightarrow \hat{\mu}_n^{\mathsf{ML}} = \frac{1}{n}\sum_{i=1}^{n} x_i. \tag{25}$$

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

Einsetzen ergibt dann weiterhin

$$\begin{aligned} -\frac{n}{2\hat{\sigma}_n^{2\mathsf{ML}}} + \frac{1}{2\hat{\sigma}_n^{4\mathsf{ML}}}\sum_{i=1}^{n}(x_i - \hat{\mu}_n^{\mathsf{ML}})^2 &= 0 \\ \Leftrightarrow -n\hat{\sigma}_n^{2\mathsf{ML}} + \sum_{i=1}^{n}(x_i - \hat{\mu}_n^{\mathsf{ML}})^2 &= 0 \qquad (26) \\ \Leftrightarrow \hat{\sigma}_n^{2\mathsf{ML}} &= \frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{\mu}_n^{\mathsf{ML}})^2. \end{aligned}$$

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

Beispiel (Normalverteilungsmodell)

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

$$\hat{\mu}_n^{\mathsf{ML}} : \mathbb{R}^n \to \mathbb{R}, x \mapsto \hat{\mu}_n^{\mathsf{ML}}(x) := \frac{1}{n} \sum_{i=1}^{n} x_i \tag{27}$$

und

$$\hat{\sigma}_n^{2^{\mathsf{ML}}} : \mathbb{R}^n \to \mathbb{R}_{\geq 0}, x \mapsto \hat{\sigma}_n^{2^{\mathsf{ML}}}(x) := \frac{1}{n} \sum_{i=1}^{n} \left(x_i - \hat{\mu}_n^{\mathsf{ML}}\right)^2 . \tag{28}$$

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

Anwendungsbeispiel

Beck Depressions-Inventar II (BDI II)

VP	BDI II Σ
1	11
2	14
3	10
4	9
5	8
6	17
7	3
8	20
9	13
10	3
11	6
12	2

Modellannahmen

Wir legen das Normalverteilungsmodell zugrunde, d.h. wir nehmen an, dass die BDI II Σ Werte Realisierungen von u.i.v. Zufallsvariablen $X_1, ..., X_{12} \sim N(\mu, \sigma^2)$ sind. Dies entspricht der Annahme, dass sich der BDI II Σ Wert einer VP durch Addition einer normalverteilten Fehlervariable mit Erwartungswertparameter 0 und Varianzparameter σ^2 zu dem über VP identischen Wert μ ergibt.

Fragestellungen

(1) Was sind sinnvolle Tipps für die wahren, aber unbekannten, Parameterwerte μ und σ^2?

(2) Wie hoch ist im frequentistischen Sinne die mit diesen Tipps assoziierte Unsicherheit?

(3) Entscheiden wir uns sinnvollerweise für die Hypothese, dass $\mu \leq 8$ ist oder dass $\mu > 8$ ist?

Anwendungsbeispiel

Beck Depressions-Inventar II (BDI II)

VP	BDI II Σ
1	11
2	14
3	10
4	9
5	8
6	17
7	3
8	20
9	13
10	3
11	6
12	2

Modellannahmen

Wir legen das Normalverteilungsmodell zugrunde, d.h. wir nehmen an, dass die BDI II Σ Werte Realisierungen von u.i.v. Zufallsvariablen $X_1, ..., X_{12} \sim N(\mu, \sigma^2)$ sind. Dies entspricht der Annahme, dass sich der BDI II Σ Wert einer VP durch Addition einer normalverteilten Fehlervariable mit Erwartungswertparameter 0 und Varianzparameter σ^2 zu dem über VP identischen Wert μ ergibt.

Fragestellungen

(1) Was sind sinnvolle Tipps für die wahren, aber unbekannten, Parameterwerte μ und σ^2?

(2) Wie hoch ist im frequentistischen Sinne die mit diesen Tipps assoziierte Unsicherheit?

(3) Entscheiden wir uns sinnvollerweise für die Hypothese, dass $\mu \leq 8$ ist oder dass $\mu > 8$ ist?

Anwendungsbeispiel

Beck Depressions-Inventar II (BDI II)

VP	BDI II Σ
1	11
2	14
3	10
4	9
5	8
6	17
7	3
8	20
9	13
10	3
11	6
12	2

Modellannahmen

Wir legen das Normalverteilungsmodell zugrunde, d.h. wir nehmen an, dass die BDI II Σ Werte Realisierungen von u.i.v. Zufallsvariablen $X_1, ..., X_{12} \sim N(\mu, \sigma^2)$ sind. Dies entspricht der Annahme, dass sich der BDI II Σ Wert einer VP durch Addition einer normalverteilten Fehlervariable mit Erwartungswertparameter 0 und Varianzparameter σ^2 zu dem über VP identischen Wert μ ergibt.

Fragestellungen

(1) Was sind sinnvolle Tipps für die wahren, aber unbekannten, Parameterwerte μ und σ^2?

Basierend auf dem Prinzip der Maximum-Likelihood Schätzung sind

$$\hat{\mu}_n^{\text{ML}} = 9.\bar{6} \text{ und } \hat{\sigma}_n^{2\,\text{ML}} = 29.7 \tag{29}$$

sinnvolle Tipps für μ und σ^2.

Statistische Modelle, Statistiken und Schätzer

- Statistische Modelle
- Statistiken und Schätzer
- Standardprobleme frequentistischer Inferenz
- Maximum-Likelihood Schätzer
- **Selbstkontrollfragen**

Selbstkontrollfragen

1. Definieren und erläutern Sie den Begriff des statistischen Modells.
2. Definieren und erläutern Sie den Begriff eine parametrischen statistischen Produktmodells.
3. Erläutern Sie den Unterschied zwischen univariaten und multivariaten statistischen Modellen.
4. Formulieren Sie das Bernoullimodel.
5. Formulieren Sie das Normalverteilungsmodell.
6. Definieren und erläutern Sie den Begriff der Statistik.
7. Definieren und erläutern Sie den Begriff des Schätzers.
8. Nennen und erläutern Sie die Standardprobleme der frequentistischen Inferenz.
9. Erläutern Sie die Standardannahmen der frequentistischen Inferenz.
10. Definieren und erläutern Sie den Begriff des Parameterpunktschätzers.
11. Definieren Sie die Begriffe der Likelihood-Funktion und der Log-Likelihood Funktion.
12. Definieren Sie den Begriff des Maximum-Likelihood Schätzers.
13. Erläutern Sie das Vorgehen zur ML Schätzung für ein parametrisches statistisches Produktmodell.
14. Geben Sie den ML Schätzer für den Parameter μ des Bernoullimodells an.
15. Geben Sie den ML Schätzer für den Parameter μ des Normalverteilungmodels an.
16. Geben Sie den ML Schätzer für den Parameter σ^2 des Normalverteilungmodels an.

Literatur

Beck, A. T. (1961). An Inventory for Measuring Depression. *Archives of General Psychiatry*, 4(6):561.

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