Technische Universität Berlin

Skript zur Vorlesung und Übung

Statistik

gelesen von Prof. Dr. Michael Scheutzow im Sommersemester 2022
Übung von Dr. Isabell Vorkastner.

Inhaltsverzeichnis

LaTeX von Viktor Stein.

Vielen Dank an Paul, Gemma, Lena, Joris, Hagen und Le Nam für das Finden und Korrigieren von Fehlern.

Diese Mitschrift ist weder von der Universität noch von dem Professor gebilligt und enthält fast sicher Fehler.

Zuletzt verändert am 16. November 2022.

Diese Vorlesung orientiert sich vornehmlich an [2, Kap. 7 - 12].

Liste der noch zu erledigenden Punkte

1 Parameterschätzung

1.1 Grundlagen und Beispiele

19.04.2022

Definition 1.1.1 (Statistisches Modell) statistisches Modell
Ein statistisches Modell ist ein Tripel $\left(\mathfrak{X}, \mathcal{F}, (\mathbb{P}_\vartheta)_{\vartheta \in \Theta}\right)$, wobei

- $\mathfrak{X} \neq \varnothing$ der Beobachtungs- oder Stichprobenraum,
- $\Theta \neq \varnothing$ eine Indexmenge, der Parameterraum,
- $\mathcal{F}$ eine σ-Algebra auf $\mathfrak{X}$ (die Algebra der Beobachtungen),
- $\mathbb{P}_\vartheta$ für $\vartheta \in \Theta$ ein Wahrscheinlichkeitsmaß auf $(\mathfrak{X}, \mathcal{F})$

sind.

Bermerkung (Interpretation des statistischen Modells) Sei ϑ unbekannt. Man beobachtet eine Realisierung $x \in \mathfrak{X}$, die mittels $\mathbb{P}_\vartheta$ generiert wird, und will daraus Rückschlüsse auf ϑ oder Funktionen, welche von ϑ abhängen, ziehen. ○

Notation. Es bezeichne stets $\mathcal{B}$ die Borel-σ-Algebra. Ist $Y: (\mathfrak{X}, \mathcal{F}) \to \left(\mathbb{R}, \mathcal{B}(\mathbb{R})\right)$ messbar, so sind $\mathbb{E}_\vartheta[Y] := \int_{\mathfrak{X}} Y(\omega) \,\mathrm{d}\mathbb{P}_\vartheta(\omega) \in [-\infty, \infty]$ beziehungsweise $\mathbb{V}_\vartheta[Y] \in [0, \infty]$ der Erwartungswert beziehungsweise die Varianz bezüglich $\mathbb{P}_\vartheta$.

Definition 1.1.2 (Statistik, Schätzer, Kenngrösse)
Seien $\left(\mathfrak{X}, \mathcal{F}, (\mathbb{P}_\vartheta)_{\vartheta \in \Theta}\right)$ ein statistisches Modell und $(\Sigma, \mathscr{S})$ ein Messraum.

- Eine messbare Abbildung $S: (\mathfrak{X}, \mathcal{F}) \to (\Sigma, \mathscr{S})$ heißt Statistik. Statistik
- Ist $\tau: \Theta \to \Sigma$ eine Abbildung, so heißt τ Kenngröße für ϑ und eine Statistik $T: \mathfrak{X} \to \Sigma$ Schätzer für τ. Schätzer

Definition 1.1.3 (Erwartungstreue)
Ein Schätzer $T: \mathfrak{X} \to \mathbb{R}^n$ für $\tau: \Theta \to \mathbb{R}^n$ ist erwartungstreu, wenn erwartungstreu

$$\mathbb{E}_\vartheta[T] = \tau(\vartheta) \qquad \forall \vartheta \in \Theta.$$

Der Bias von T ist Bias

$$\mathbb{B}[T]: \Theta \to [-\infty, \infty]^n, \qquad \vartheta \mapsto \mathbb{E}_\vartheta[T] - \tau(\vartheta).$$

Das Risiko von T ist Risiko

$$\mathbb{F}_\vartheta[T] := \mathbb{E}_\vartheta\left[|T - \tau(\vartheta)|^2\right].$$

Definition 1.1.4 (Produktmodell)
Ist $\big(E, \mathfrak{E}, (\mathbb{Q}_\vartheta)_{\vartheta\in\Theta}\big)$ ein statistisches Modell, so ist $\big(E^n, \mathfrak{E}^{\otimes n}, (\mathbb{Q}_\vartheta^{\otimes n})_{\vartheta\in\Theta}\big)$ für $n \in \mathbb{N}_{>0}$ das zugehörige n-fache Produktmodell. Produktmodell

Notation. In der Situation von Definition 1.1.4 schreiben wir oft

$$X\colon E^n \to E^n, \qquad (x_1, \ldots, x_n) \mapsto (x_1, \ldots, x_n)$$

für die Identität auf $\mathfrak{X} := E^n$ und $X_k(e_1, \ldots, e_n) := e_k$ für $k \in \{1, \ldots, n\}$. Dann sind $X_1, \ldots, X_n$ bezüglich $\mathbb{Q}_\vartheta^{\otimes n}$ unabhängig und identisch verteilt (u.i.v.) mit Verteilung $\mathbb{Q}_\vartheta$.

Beispiele statistischer Modelle

Beispiel 1.1.5 (Strahlenbelastung von Waldpilzen)
Die Strahlenbelastung von Waldpilzen soll überprüft werden. Dazu wird bei n unabhängigen Pilzproben die Anzahl der Geigerzähler-Impulse jeweils während einer Zeiteinheit gemessen. Ein geeignetes statistisches Modell aufzustellen ist Hausaufgabe 1.1. ◇

Beispiel 1.1.6 (Bernoulli-Modell)
Um den Wurf einer verbogenen Münze mit den Seiten "0" und "1" zu beschreiben, wählen wir $\mathfrak{X} = \{0, 1\}$, $\mathcal{F} := \mathcal{P}(\mathfrak{X})$ sowie $\Theta := [0, 1]$ und $\mathbb{P}_\vartheta(\{1\}) = \vartheta$ und $\mathbb{P}_\vartheta(\{0\}) = 1 - \vartheta$, sodass $(\mathbb{P}_\vartheta)_{\vartheta\in\Theta}$ alle Wahrscheinlichkeitsmaße auf $(\mathfrak{X}, \mathcal{F})$ umfasst. Es ist auch möglich, stattdessen z.B. $\Theta = [\frac{1}{4}, \frac{3}{4}]$ zu wählen.

Um ϑ zu schätzen, wählen wir $\tau\colon \Theta \to \Theta$ als die Identität. Auf $\Sigma := \Theta$ *wählen* wir die σ-Algebra $\mathcal{B}(\Theta)$. Ein naheliegender Schätzer für τ ist

Wäre stattdessen z.B. $\Theta = (0, 1)$, ist T kein Schätzer, da $T(0) = 0 \notin \Theta$.

$$T\colon \mathfrak{X} \to \Theta, \qquad x \mapsto x.$$

welcher erwartungstreu ist, denn für $\vartheta \in \Theta$ gilt

$$\mathbb{E}_\vartheta[T] = 1 \cdot \mathbb{P}_\vartheta(T = 1) + 0 \cdot \mathbb{P}_\vartheta(T = 0) = 1 \cdot \vartheta + 0 \cdot (1 - \vartheta) = \vartheta = \tau(\vartheta).$$ ◇

Beispiel 1.1.7 (Bernoulli-Produktmodell)
Um $n \in \mathbb{N}_{>0}$ unabhängige Münzwürfe zu modellieren, wählen wir das n-fache Produktmodell des Modells aus Beispiel 1.1.6: für $x \in \mathfrak{X} := \{0, 1\}^n$ und $\vartheta \in \Theta := [0, 1]$ ist

$$\mathbb{P}_\vartheta^{\otimes n}(\{x\}) = \vartheta^{\sum_{k=1}^n x_k}(1 - \vartheta)^{n - \sum_{k=1}^n x_k}.$$

Sei die Kenngröße τ wieder die Identität auf Θ. Ein natürlicher Schätzer für τ ist der empirische Mittelwert $T\colon \mathfrak{X} \to \Theta$, $x \mapsto T(x_1, \ldots, x_n) := \frac{1}{n}\sum_{k=1}^n x_k$, welcher erwartungstreu ist:

$$\mathbb{E}_\vartheta[T] = \mathbb{E}_\vartheta\left[\frac{1}{n}\sum_{k=1}^n X_k\right] = \frac{1}{n}\sum_{k=1}^n \underbrace{\mathbb{E}_\vartheta[X_k]}_{=\vartheta} = \vartheta.$$ ◇

Beispiel 1.1.8 (Binomialmodell: Anzahl der Einsen bei n-fachem Münzwurf)
Um den n-fachen Münzwurf, bei welchem lediglich die Anzahl der Einsen und nicht die Reihenfolge beachtet wird, zu modellieren, wählen wir $\mathfrak{X} := \{0, \ldots, n\}$ (viel kleiner als $\{0,1\}^n!$), $\mathcal{F} := \mathcal{P}(\mathfrak{X})$, $\Theta := [0,1]$ und die Binomialverteilungen $\mathbb{P}_\vartheta(\{k\}) := \binom{n}{k}\vartheta^k(1-\vartheta)^{n-k}$ für $k \in \mathfrak{X}$ und $\vartheta \in \Theta$.

Um ϑ zu schätzen, wählen wir wieder τ als die Identität auf Θ. Die Statistik $T\colon \mathfrak{X} \to \Theta$, $x \mapsto \frac{1}{n}x$ ist ein natürlicher und erwartungstreuer Schätzer für τ, denn für $\vartheta \in \Theta$ gilt

$$\mathbb{E}_\vartheta[T] = \sum_{k=0}^{n} T(k)\mathbb{P}_\vartheta(\{k\}) = \frac{1}{n}\sum_{k=0}^{n} k \cdot \binom{n}{k}\vartheta^k(1-\vartheta)^{n-k} = \frac{1}{n} \cdot n\vartheta = \vartheta.$$

◇

Beispiel 1.1.9 (Lebensdauer von Glühbirnen)
Die Lebensdauer von Glühbirnen sei $\exp(\vartheta)$-verteilt mit $\vartheta \in (0,\infty)$. Wir wählen das statistische Modell mit $\mathfrak{X} = (0,\infty)$, $\mathcal{F} = \mathcal{B}(\mathfrak{X})$, $\Theta = \mathfrak{X}$ und als $\mathbb{P}_\vartheta$ die Exponentialverteilungen mit Dichte $x \mapsto \vartheta e^{-\vartheta x}$ bezüglich des Lebesgue-Maßes für $\vartheta \in \Theta$. Betrachte nun das zugehörige n-fache Produktmodell.

Ein natürlicher (da $\mathbb{E}[\exp(\lambda)] = \lambda^{-1}$ ist) Schätzer für ϑ ist

$$T(x) := T(x_1, \ldots, x_n) := \frac{1}{\frac{1}{n}\sum_{k=1}^{n} x_k},$$

welcher schon für $n = 1$ nicht erwartungstreu ist: für $\vartheta \in \Theta$ gilt

$$\mathbb{E}_\vartheta[T] = \int_0^\infty T(x)\vartheta e^{-\vartheta x}\,\mathrm{d}x = \vartheta \underbrace{\int_0^\infty \frac{1}{x}e^{-\vartheta x}\,\mathrm{d}x}_{=\infty} = \infty \neq \vartheta.$$

Man kann aber fast analog zeigen, dass $\frac{1}{T}$ ein erwartungstreuer Schätzer für $\frac{1}{\vartheta}$ ist. ◇

Beispiel 1.1.10 (Gauss-Produktmodell)
Zu einem gegebenen Mittelwert $m \in \mathbb{R}$ ist $\big(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), (\mathcal{N}_{m,\vartheta}^{\otimes n})_{\vartheta>0}\big)$ das n-fache Gausssche Produktmodell. Die Statistik

$$T_m\colon \mathbb{R}^n \to [0,\infty), \qquad x \mapsto \sqrt{\frac{\pi}{2}}\frac{1}{n}\sum_{k=1}^{n}|x_k - m|$$

ist ein erwartungstreuer Schätzer für $\tau\colon (0,\infty) \to (0,\infty)$ $\vartheta \mapsto \sqrt{\vartheta}$ (Hausaufgabe 1.2). ◇

Beispiel 1.1.11 (Nichtparametrisches Modell)
Seien $\mathfrak{X} = \mathbb{R}$, $\mathcal{F} = \mathcal{B}(\mathbb{R})$ sowie $(\mathbb{P}_\vartheta)_{\vartheta\in\Theta}$ die Familie aller Wahrscheinlichkeitsmaße auf $(\mathfrak{X}, \mathcal{F})$. Wir können Θ als die (unendlichdimensionale) Menge der Verteilungsfunktionen F wählen[1]. In diesem Fall nennt man $\big(\mathfrak{X}, \mathcal{F}, (\mathbb{P}_\vartheta)_{\vartheta\in\Theta}\big)$ unparametrisches statistisches Modell.

Im dazugehörigen n-fachen Produktmodell wollen wir die Verteilungsfunktion $\vartheta = F$ schätzen. Eine natürlicher Schätzerfür die Identität τ auf Θ ist

$$T_n\colon \mathbb{R}^n \to \Theta, \qquad (x_1,\ldots,x_n) \mapsto \left(F_n\colon \mathbb{R} \to [0,1], \qquad s \mapsto \frac{1}{n}\sum_{k=1}^{n}\mathbb{1}_{(-\infty,s]}(x_k)\right).$$

Wir nennen F_n die empirische Verteilungsfunktion. ◇

empirische Verteilungsfunktion

[1] Eine sinnvolle σ-Algebra auf Θ wird durch die Punktauswertungen $F \mapsto F(t)$ für $t \in \mathbb{R}$ erzeugt, cf. WT II: Konstruktion des Wiener-Maßen / Brownsche Bewegung.

Beispiel 1.1.12 (Produktmodell mit Gleichverteilung)
Für das Modell $\left(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), (\mathcal{U}_\vartheta^{\otimes n})_{\vartheta\in\mathbb{R}}\right)$, wobei $n \geqslant 1$ und $\mathcal{U}_\vartheta$ die Gleichverteilung auf dem Intervall $[\vartheta - \frac{1}{2}, \vartheta + \frac{1}{2}]$ sind, sind

$$M(x) := \frac{1}{n}\sum_{k=1}^{n} x_k \qquad \text{und} \qquad T(x) := \frac{1}{2}\left(\max_{1\leqslant k\leqslant n} x_k + \min_{1\leqslant j\leqslant n} x_j\right)$$

zwei erwartungstreue Schätzer für den Parameter ϑ. Die Verifizierung dieser Aussage und die Berechnung der Varianzen $\mathbb{V}[M]$ und $\mathbb{V}[T]$ ist Hausaufgabe 1.3. Der Wert $M(x)$ heißt empirischer Mittelwert der Beobachtung x. ◇ empirischer Mittelwert

20.04.2022

Beispiel 1.1.13 (Qualitätskontrolle - Verschiedene Ansätze der Statistik)
Wir bekommen eine Lieferung von N Orangen, von denen $\vartheta \in \Theta := \{0, \ldots, N\}$ Stück faul sind. Anhand einer Stichprobe von $n \ll N$ Orangen, von denen $x \in \{0, \ldots, n\}$ faul sind, schätzen wir die Anzahl der faulen Orangen in der gesamten Lieferung.

Ansatz 1: Naive Schätzung / Hochrechnung. Unter den Annahme, dass die Stichprobe ungefähr repräsentativ ist, gilt $\frac{x}{n} \approx \frac{\vartheta}{N}$. Ein natürlicher Schätzer für $\tau = \mathrm{id}_\Theta$ ist also

$$T\colon \mathfrak{X} \to \Sigma, \qquad x \mapsto x \cdot \frac{N}{n},$$

wobei $\mathfrak{X} = \{0, \ldots, n\}$, $\mathcal{F} = \mathcal{P}(\mathfrak{X})$ und $\Sigma = \Theta$ sind. Für $\vartheta \in \Theta$ wählen wir $\mathbb{P}_\vartheta$ als die hypergeometrische Verteilung

$$\mathrm{Hyp}_{n,\vartheta,N-\vartheta}(k) = \frac{\binom{\vartheta}{k}\binom{N-\vartheta}{n-k}}{\binom{N}{n}} \mathbb{1}_{\{0,\ldots,\vartheta\}}(k).$$

Der Schätzer T ist erwartungstreu, denn für $\vartheta \in \Theta$ gilt

$$\begin{aligned}
\mathbb{E}_\vartheta[T] &= \sum_{k=0}^{n} T(k)\mathbb{P}_\vartheta(\{k\}) = \sum_{k=0}^{n\wedge\vartheta} k\frac{N}{n}\frac{\binom{\vartheta}{k}\binom{N-\vartheta}{n-k}}{\binom{N}{n}} = \vartheta \sum_{k=1}^{n\wedge\vartheta} \frac{\binom{\vartheta-1}{k-1}\binom{N-\vartheta}{n-k}}{\binom{N-1}{n-1}} \\
&= \vartheta \sum_{k=0}^{(n-1)\wedge(\vartheta-1)} \underbrace{\frac{\binom{\vartheta-1}{k}\binom{N-\vartheta}{n-1-k}}{\binom{N-1}{n-1}}}_{=\mathrm{Hyp}_{n-1,\vartheta-1,N-\vartheta}(k)} = \vartheta,
\end{aligned}$$

wobei $a \wedge b := \min(a, b)$ eine Kurzschreibweise ist.

Ansatz 2: Schätzung mit Fehlerangabe. Wir wählen ein Konfidenzintervall $C(x)$ (cf. Abschnitt 2), sodass mit hinreichender Wahrscheinlichkeit $\vartheta \in C(x)$ gilt. Zu einem Irrtumsniveau $0 < \alpha \ll 1$ können wir z.B. ein $m \in \mathbb{N}$ wählen, sodass

$$\mathbb{P}_\vartheta\left(\{x \in \mathfrak{X} : \vartheta \in [T(x) - m, T(x) + m]\}\right) \geqslant 1 - \alpha \qquad \forall \vartheta \in \Theta$$

oder ein $\tilde{m} \in \mathbb{N}$, sodass

$$\mathbb{P}_\vartheta\left(\{x \in \mathfrak{X} : \vartheta \in [0, T(x) + \tilde{m}]\}\right) \geqslant 1 - \alpha \qquad \forall \vartheta \in \Theta.$$

Ansatz 3: Entscheidungsfindung (cf. Abschnitt 4.) Die Orangen sollen zurück geschickt werden, wenn mehr als 5% aller Orangen faul sind.[2]

[2]Wir nehmen an, dass N durch 20 teilbar ist.

Die Nullhypothese H_0 ist $\vartheta \in \{0, \dots, \frac{N}{20}\}$ und die Alternative H_1 ist $\vartheta \in \{\frac{N}{20}+1, \dots, N\}$.

Wir benötigen ein Entscheidungsverfahren bzw. einen Schwellenparameter $c \geqslant 0$ sodass $x \leqslant c$ eine Entscheidung für H_0 und $x > c$ eine Entscheidung für H_1 bedingt und müssen dafür ein geeignetes $c > 0$ wählen. ◇

Beispiel 1.1.14 (Einziger erwartungstreuer Schätzer ist unsinnig)
Im folgenden Modell existiert genau ein erwartungstreuer Schätzer für eine bestimmte Kenngröße, doch dieser ist Unsinn.

Die wöchentliche Anzahl an Unfällen in einer Stadt sei Poiss(ϑ)-verteilt und das statistische Modell $\left(\mathbb{N}_0, \mathcal{P}(\mathbb{N}_0), (\mathbb{P}_\vartheta)_{\vartheta\in\Theta}\right)$ mit $\mathbb{P}_\vartheta \sim \mathrm{Poiss}(\vartheta)$ für $\vartheta \in \Theta := (0,\infty)$. Es gilt $\mathbb{P}_\vartheta(\{k\}) = e^{-\vartheta}\frac{\vartheta^k}{k!}$. Die Wahrscheinlichkeit, dass in drei unabhängigen Wochen keine Unfälle passieren ist $\tau\colon \Theta \to \mathbb{R}$, $\vartheta \mapsto \mathbb{P}_\vartheta(\{0\})^3 = e^{-3\vartheta}$. Ein Schätzer $T\colon \mathbb{N}_0 \to \mathbb{R}$ ist erwartungstreu genau dann wenn für alle $\vartheta \in \Theta$ gilt

$$e^{-3\vartheta} \overset{!}{=} \mathbb{E}_\vartheta[T] = \sum_{k=0}^{\infty} T(k) e^{-\vartheta}\frac{\vartheta^k}{k!},$$

also genau dann wenn

$$\sum_{k=0}^{\infty}(-2)^k\frac{\vartheta^k}{k!} = e^{-2\vartheta} \overset{!}{=} \sum_{k=0}^{\infty} T(k)\frac{\vartheta^k}{k!},$$

für alle $\vartheta \in \Theta$ gilt. Nach dem Identitätssatz sind die beiden Potenzreihen genau dann gleich, wenn ihre Koeffizienten übereinstimmen, das heißt, wenn $T(k) = (-2)^k$ für alle $k \in \mathbb{N}_0$ gilt.

Dieser Schätzer ist jedoch unsinnig, da $[0,1] \ni \tau(\vartheta) \neq (-2)^k$ für alle $k \in \mathbb{N}_0$ gilt. ◇

Beispiel 1.1.15 (Gameshow) **21.04.2022**
Der Moderator wählt ein beliebiges $\vartheta > 0$. Dann produziert ein Apparat Realisierungen von unabhängig auf $[0,\vartheta]$ gleichmäßig verteilten Zufallsvariablen $X_1, \dots, X_n$. Die Spieler sollen ϑ anhand der Realisierungen schätzen. Wir wählen $\Theta := (0,\infty)$ sowie das n-fache Produktmodell des Modells mit $\mathfrak{X} = [0,\infty)$ (ein andere Möglichkeit wäre $\mathbb{R}$), $\mathcal{F} := \mathcal{B}(\mathfrak{X})$ und $\mathbb{P}_\vartheta := \mathcal{U}_{[0,\vartheta]}$ für $\vartheta \in \Theta$, wobei $\mathcal{U}_{[0,\vartheta]}$ die Gleichverteilung auf $[0,\vartheta]$ ist, das heißt

$$\mathcal{U}_{[0,\vartheta]}([0,t]) = \frac{t}{\vartheta} \wedge 1 \qquad \forall t \geqslant 0.$$

Wir wählen $\Sigma := \Theta$, $\mathscr{S} := \mathcal{F}$ und wollen einen guten Schätzer für die Identität τ finden.

1. Möglichkeit. Wir wählen

$$T_n\colon \mathfrak{X}^n \to \mathfrak{X}, \qquad x = (x_1, \dots, x_n) \mapsto \frac{2}{n}\sum_{k=1}^{n} x_k,$$

da $\mathbb{E}[X] = \frac{1}{2}\vartheta$ für $X \sim \mathcal{U}_{[0,\vartheta]}$ ist.

2. Möglichkeit. Wir wählen

$$\tilde{T}_n\colon \mathfrak{X}^n \to \mathfrak{X}, \qquad x = (x_1, \dots, x_n) \mapsto \max(x_1, \dots, x_n).$$

Wir zeigen, dass die Schätzfolgen $(T_n)_{n\in\mathbb{N}}$ und $(\tilde{T}_n)_{n\in\mathbb{N}}$ konsistent sind, also, dass $\mathbb{P}_\vartheta(|T_n - \vartheta| > \varepsilon) \xrightarrow{n\to\infty} 0$ (und analog für $\tilde{T}_n$) für alle $\vartheta \in \Theta$ und $\varepsilon > 0$ gilt.

Für $\vartheta \in \Theta$ und $\varepsilon > 0$ gilt

$$\begin{aligned}\mathbb{P}_\vartheta(|T_n - \vartheta| > \varepsilon) &= \mathbb{P}_\vartheta\left(\left|\frac{2}{n}\sum_{k=1}^n X_k - \vartheta\right| > \varepsilon\right) \\ &= \mathbb{P}_\vartheta\left(\left|\frac{1}{n}\sum_{k=1}^n X_k - \mathbb{E}_\vartheta[X_k]\right| > \frac{\varepsilon}{2}\right) \xrightarrow[\substack{\text{Gesetz der}\\ \text{großen Zahlen}}]{n\to\infty} 0.\end{aligned}$$

Wir können ohne Beschränkung der Allgemeinheit zu $\vartheta \in \Theta$ ein $0 < \varepsilon < \vartheta$ wählen. Dann gilt, weil $\tilde{T}_n(x)$ für alle $x \in \mathfrak{X}$ kleiner gleich ϑ ist, dass

$$\begin{aligned}\mathbb{P}_\vartheta(|\tilde{T}_n - \vartheta| > \varepsilon) &= \mathbb{P}_\vartheta(\tilde{T}_n < \vartheta - \varepsilon) = \mathbb{P}_\vartheta(X_1 \vee \ldots \vee X_n < \vartheta - \varepsilon) \\ &= \mathbb{P}_\vartheta(X_1 < \vartheta - \varepsilon, \ldots, X_n < \vartheta - \varepsilon) \overset{(\star)}{=} \mathbb{P}_\vartheta(X_1 < \vartheta - \varepsilon)^n \\ &= \left(\frac{\vartheta - \varepsilon}{\vartheta}\right)^n \xrightarrow[0<\varepsilon<\vartheta]{n\to\infty} 0,\end{aligned}$$

wobei wir in $(\star)$ die Unabhängigkeit der $X_1, \ldots, X_n$ und deren identische Verteilung ausnutzen.

Nun untersuchen wir für jedes $n \in \mathbb{N}$ die Erwartungstreue der beiden Schätzer. Für $\vartheta \in \Theta$ gilt

$$\mathbb{E}_\vartheta[T_n] = \frac{2}{n}\sum_{k=1}^n \mathbb{E}_\vartheta[X_k] = \frac{2}{n}\cdot n \cdot \frac{\vartheta}{2} = \vartheta.$$

Die Dichte von $\tilde{T}_n$ bezüglich $\mathbb{P}_\vartheta$ ist $n\left(\frac{\cdot}{\vartheta}\right)^{n-1}\frac{1}{\vartheta}\mathbb{1}_{[0,\vartheta]}$, die Ableitung von der Verteilungsfunktion $\mathbb{P}_\vartheta(\tilde{T}_n \leqslant y) = \left(\frac{y}{\vartheta} \wedge 1\right)^n \mathbb{1}_{[0,\infty)}(y)$. Somit gilt

$$\mathbb{E}_\vartheta[\tilde{T}_n] = n\int_0^\vartheta y \cdot \left(\frac{y}{\vartheta}\right)^{n-1}\frac{1}{\vartheta}\,\mathrm{d}y = n\vartheta^{-n}\int_0^\vartheta y^n\,\mathrm{d}y = n\vartheta^{-n}\frac{\vartheta^{n+1}}{n+1} = \frac{n}{n+1}\vartheta, \tag{1}$$

also ist $\tilde{T}_n$ nicht erwartungstreu, aber die Folge $(\tilde{T}_n)_{n\in\mathbb{N}}$ asymptotisch erwartungstreu, das heißt

asymptotisch erwartungstreu

$$\mathbb{B}_\vartheta(\tilde{T}_n) = \frac{n}{n+1}\vartheta - \vartheta \xrightarrow{n\to\infty} 0 \qquad \forall \vartheta \in \Theta.$$

3. Möglichkeit. Wir wählen den Schätzer $T_n^* := \frac{n+1}{n}\tilde{T}_n$. Nach den vorangegangenen Berechnung ist T_n^* erwartungstreu für jedes $n \in \mathbb{N}$ und die Schätzfolge $(T_n^*)_{n\in\mathbb{N}}$ ist konsistent.

Vergleich der Schätzer. Wir berechnen nun die Varianz der Schätzer, um sie zu vergleichen. Die Varianz einer Zufallsvariable $X \sim \mathcal{U}_{[0,\vartheta]}$ ist $\mathbb{V}_\vartheta[X] = \frac{\vartheta^2}{12}$. Die $X_1, \ldots, X_n$ sind unabhängig identisch verteilt. Aufgrund der Rechenregeln für die Varianz gilt

$$\mathbb{V}_\vartheta[T_n] = \mathbb{V}_\vartheta\left[\frac{2}{n}\sum_{k=1}^n X_k\right] = \frac{4}{n^2}\mathbb{V}_\vartheta\left[\sum_{k=1}^n X_k\right] = \frac{4}{n}\mathbb{V}_\vartheta[X_1] = \frac{4}{n}\frac{\vartheta^2}{12} = \frac{1}{3n}\vartheta^2. \tag{2}$$

Ferner gilt

$$\begin{aligned}\mathbb{V}_\vartheta[\tilde{T}_n] &= \int_0^\vartheta x^2 n\left(\frac{x}{\vartheta}\right)^{n-1}\frac{1}{\vartheta}\,\mathrm{d}x - \mathbb{E}_\vartheta[\tilde{T}_n]^2 \overset{(1)}{=} n\vartheta^{-n}\int_0^\vartheta x^{n+1}\,\mathrm{d}x - \frac{n^2}{(n+1)^2}\vartheta^2 \\ &= \frac{n}{n+2}\vartheta^2 - \frac{n^2}{(n+1)^2}\vartheta^2 = \frac{n}{(n+1)^2(n+2)}\vartheta^2.\end{aligned} \tag{3}$$

Beachte, dass dieser Schätzer sich eventuell nicht um ϑ zentriert. Es folgt

$$\mathbb{V}_\vartheta[T_n^*] = \left(\frac{n+1}{n}\right)^2 \mathbb{V}_\vartheta[\tilde{T}_n] \overset{(3)}{=} \frac{1}{n(n+2)}\vartheta^2. \tag{4}$$

Daher ist T_n^* ein besserer Schätzer: er zentriert sich schneller um ϑ.

Wichtiger ist die mittlere quadratische Abweichung von ϑ $\mathbb{F}_\vartheta[T]$. Aufgrund der Erwartungstreue von T_n gilt

$$\mathbb{F}_\vartheta[T_n] = \mathbb{V}_\vartheta[T_n] \overset{(2)}{=} \frac{1}{3n}\vartheta^2$$

und ferner

$$\begin{aligned}\mathbb{F}_\vartheta[\tilde{T}_n] &= \mathbb{E}_\vartheta\left[|\tilde{T}_n - \vartheta|^2\right] = \mathbb{V}_\vartheta[\tilde{T}_n - \vartheta] + \left(\mathbb{E}_\vartheta[\tilde{T}_n - \vartheta]\right)^2 \\ &= \mathbb{V}_\vartheta[\tilde{T}_n] + \mathbb{B}_\vartheta[\tilde{T}_n]^2 \overset{(3)}{=} \frac{n}{(n+1)^2(n+2)}\vartheta^2 + \frac{1}{(n+1)^2}\vartheta^2 = \frac{2}{(n+1)(n+2)}\vartheta^2.\end{aligned}$$

Schließlich gilt aufgrund der Erwartungstreue von T_n^*

$$\mathbb{F}_\vartheta[T_n^*] = \mathbb{V}_\vartheta[T_n^*] \overset{(4)}{=} \frac{1}{n(n+2)}\vartheta^2.$$

Für jedes $n \in \mathbb{N}$ gilt also

$$\mathbb{F}_\vartheta[T_n^*] \leqslant \mathbb{F}_\vartheta[\tilde{T}_n] \leqslant \mathbb{F}_\vartheta[T_n] \qquad \forall \vartheta \in (0, \infty).$$

Ein Schätzer mit noch geringerem Risiko wäre $\hat{T}_n := \frac{n+2}{n+1}\tilde{T}_n$, denn dann gilt $\mathbb{F}_\vartheta[\hat{T}_n] \leqslant \mathbb{F}_\vartheta[T_n^*]$. Dieser Schätzer ist aber nicht erwartungstreu!

In diesem Beispiel existiert kein risikominimierender Schätzer T_n^{**} mit $\mathbb{F}_\vartheta[T_n^{**}] \leqslant \mathbb{F}_\vartheta[\check{T}_n]$ für alle $\vartheta \in \Theta$ und alle Schätzer $\check{T}_n$.

Beweis. Angenommen, so ein Schätzer T_n^{**} existiert. Seien $\vartheta_0 \in \Theta$ und $T_n^{(\vartheta_0)} \equiv \vartheta_0$ ein Schätzer auf $\mathfrak{X}$. Dann gilt $\mathbb{F}_{\vartheta_0}[T_n^{**}] \leqslant \mathbb{F}_\vartheta[T_n^{(\vartheta_0)}] = 0$ und somit $\mathbb{F}_{\vartheta_0}[T_n^{**}] = 0$ für alle $\vartheta_0 \in \Theta$, was einen Widerspruch darstellt. □

Es kann hingegen risikominimierende erwartungstreue Schätzer geben. ◇

1.2 Maximum Likelihood Schätzung

Definition 1.2.1 (Absolut stetig [4, Def. 3.5])
Seien μ und ν Maße auf (Ω, Q). Dann ist ν absolut stetig bezüglich μ und wir schreiben $\nu \ll \mu$ wenn aus $A \in Q$ und $\mu(A) = 0$ folgt, dass $\nu(A) = 0$. absolut stetig

Satz 1.2.1: Radon-Nikodym (1913, 1930)

Seien μ und ν σ-endliche Maße auf (Ω, Q). Dann hat ν eine Dichte f bezüglich μ, das heißt $\nu(A) = \int_A f \,\mathrm{d}\mu$ für alle $A \in Q$, und wir schreiben $f = \frac{\mathrm{d}\nu}{\mathrm{d}\mu}$, genau dann wenn $\nu \ll \mu$ gilt. Dann ist f bis auf Modifikation auf μ-Nullmengen eindeutig und reellwertig.

Definition 1.2.2 (Standardmodell, Likelihoodfunktion)
Ein Modell $\big(\mathfrak{X}, \mathcal{F}, (\mathbb{P}_\vartheta)_{\vartheta\in\Theta}\big)$ ist ein Standardmodell, wenn ein dominierendes σ-endliches Maß μ_0 auf $(\mathfrak{X}, \mathcal{F})$ existiert, das heißt $\mathbb{P}_\vartheta \ll \mu_0$ für alle $\vartheta \in \Theta$. Die Likelihoodfunktion ist Standardmodell Likelihoodfunktion

$$\rho\colon \mathfrak{X} \times \Theta \to \mathbb{R}, \qquad (x, \theta) \mapsto \frac{\mathrm{d}\mathbb{P}_\vartheta}{\mathrm{d}\mu_0}(x).$$

Definition 1.2.3 (Maximum-Likelihood-Schätzer)
Sei $\big(\mathfrak{X}, \mathcal{F}, (\mathbb{P}_\vartheta)_{\vartheta\in\Theta}\big)$ ein Standardmodell mit Likelihoodfunktion ρ. Ein Schätzer $T\colon \mathfrak{X} \to \Sigma \supset \Theta$ ist ein Maximum-Likelihood-Schätzer für $\tau\colon \Theta \to \Theta$, $\vartheta \mapsto \vartheta$, wenn Maximum-Likelihood-Schätzer

$$\rho(x, T(x)) = \sup_{\vartheta\in\Theta} \rho(x, \vartheta) \qquad \text{für } \mathbb{P}_\vartheta - \text{fast alle } x \in \mathfrak{X} \;\; \forall \vartheta \in \Theta.$$

Beispiel 1.2.4 (Likelihoodfunktion für abzählbares $\mathfrak{X}$)
Seien $\mathfrak{X}$ höchstens abzählbar unendlich und μ_0 das Zählmaß auf $\mathfrak{X}$, welches σ-endlich ist. Dann gilt für jede Familie $(\mathbb{P}_\vartheta)_{\vartheta\in\Theta}$ von Wahrscheinlichkeitsmaßen auf $(\mathfrak{X}, \mathcal{P}(\mathfrak{X}))$ das $\mathbb{P}_\vartheta \ll \mu_0$ für alle $\vartheta \in \Theta$ gilt, denn aus $\mu_0(A) = 0$ für $A \in \mathcal{P}(\mathfrak{X})$ folgt $A = \varnothing$ und somit $\mathbb{P}_\vartheta(A) = 0$, also ist $\big(\mathfrak{X}, \mathcal{P}(\mathfrak{X}), (\mathbb{P}_\vartheta)_{\vartheta\in\Theta}\big)$ ein Standardmodell.

26.04.2022

Da nur $\varnothing$ eine μ_0-Nullmenge ist, ist die eindeutige (nicht fast sicher, sondern sicher!) Likelihoodfunktion

$$\rho(x, \vartheta) = \frac{\mathrm{d}\mathbb{P}_\vartheta}{\mathrm{d}\mu_0}(x) = \mathbb{P}_\vartheta(\{x\})$$

für $\vartheta \in \Theta$ und $x \in \mathfrak{X}$, denn es gilt

$$\mathbb{P}_\vartheta(\{x\}) = \int_{\{x\}} \rho(y, \vartheta) \,\mathrm{d}\mu_0(y) = \rho(x, \vartheta) \cdot \mu_0(\{x\}) = \rho(x, \vartheta).$$

◇

Beispiel 1.2.5 (Kontinuierliches $\mathfrak{X}$)
Seien $\mathfrak{X} = \mathbb{R}^n$ und $\mathcal{F} = \mathcal{B}(\mathbb{R}^n) = \mathcal{B}(\mathbb{R})^{\otimes n}$. Das Lebesgue-Maß auf $(\mathfrak{X}, \mathcal{F})$ ist σ-endlich. Sei $(\mathbb{P}_\vartheta)_{\vartheta\in\Theta}$ eine Familie von Wahrscheinlichkeitsmaßen auf $(\mathfrak{X}, \mathcal{F})$, welche eine Dichte $\rho(\cdot, \vartheta)$ bezüglich μ_0 haben. Dann ist $\big(\mathfrak{X}, \mathcal{F}, (\mathbb{P}_\vartheta)_{\vartheta\in\Theta}\big)$ ein Standardmodell. ◇

Gegenbeispiel 1.2.6 (Kein dominierendes σ-endliches Maß)
Seien $\mathfrak{X} = \mathbb{R}$ (man könnte auch $\mathbb{R}^n$ wählen), $\mathcal{F} = \mathcal{B}(\mathfrak{X})$ und $(\mathbb{P}_\vartheta)_{\vartheta\in\Theta}$ die Menge aller Wahrscheinlichkeitsmaße auf $(\mathfrak{X}, \mathcal{F})$. Dann gibt es kein σ-endliches Maß μ_0 mit $\mathbb{P}_\vartheta \ll \mu_0$ für alle $\vartheta \in \Theta$.

Beweis. Angenommen, so ein σ-endliches Maß μ_0 existiert. Für $x \in \mathfrak{X}$ sei $\delta_x(A) := \mathbb{1}_A(x)$ für $A \in \mathcal{F}$. Da nach Annahme $\delta_x \ll \mu_0$ für alle $x \in \mathfrak{X}$ gilt, folgt wegen $\delta_x(\{x\}) = 1 > 0$, dass $\mu_0(\{x\}) > 0$ ist. Dies widerspricht der σ-Endlichkeit von μ_0: Da μ_0 σ-endlich ist, existiert eine Folge $(A_n)_{n\in\mathbb{N}} \subset \mathcal{F}$ mit $\mu_0(A_k) < \infty$ für alle $k \in \mathbb{N}$ und $\bigcup_{k\in\mathbb{N}} A_k = \mathfrak{X}$. Da die abzählbare Vereinigung abzählbarer Menge abzählbar ist, aber $\mathfrak{X}$ aber überabzählbar, existiert ein $j \in \mathbb{N}$ sodass A_j überabzählbar ist. Für jedes $x \in A_j$ existiert ein $k_x \in \mathbb{N}$ sodass $\mu(\{x\}) > \frac{1}{k_x}$ ist. Für $k \in \mathbb{N}$ definieren wir $B_k := \left\{x \in A_j : \mu_0(\{x\}) > \frac{1}{k}\right\}$. Dann ist $A_j = \bigcup_{k\in\mathbb{N}} B_k$. Da A_j überabzählbar ist, existiert ein $m \in \mathbb{N}$ sodass B_m überabzählbar ist. Dann ist $\mu_0(B_m) = \sum_{x\in B_m} \mu_0(\{x\}) > \sum_{x\in B_m} \frac{1}{k} = \infty$. □

Beispiel 1.2.7 (Maximum-Likelihood-Schätzer für das Binomialmodell)
Betrachte $\mathfrak{X} = \{0, \dots, n\}$, $\mathcal{F} = \mathcal{P}(\mathfrak{X})$ und μ_0 das Zählmaß auf $(\mathfrak{X}, \mathcal{F})$ (oder jedes andere Maß, dessen einzige Nullmenge $\varnothing$ ist) sowie $\Theta = [0, 1]$. Nach Beispiel 1.2.4 ist die Dichte $\rho(x, \vartheta) = \mathbb{P}_\vartheta(\{x\}) = \binom{n}{x}\vartheta^x(1-\vartheta)^{n-x}$ für alle $x \in \mathfrak{X}$ und $\vartheta \in \Theta$.

Wir wollen für beobachtes $x \in \mathfrak{X}$ das $\vartheta \in \Theta$ wählen, für das die Likelihoodfunktion $\rho(x, \cdot)$ maximal ist. Da ln streng monoton wachsend ist, ist es äquivalent die Loglikelihoodfunktion

Loglikelihood-funktion

$$(0, 1) \ni \vartheta \mapsto \ln\big(\rho(\vartheta, x)\big) = \ln\left(\binom{n}{x}\right) + x\ln(\vartheta) + (n-x)\ln(1-\vartheta)$$

zu maximieren, da das Maximum nicht am Rand auftreten kann, da $\rho(0, \cdot) = \rho(1, \cdot) \equiv 0$ ist. Für $\vartheta \in (0, 1)$ ist

$$\frac{\partial}{\partial\vartheta}\ln\big(\rho(\vartheta, x)\big) = \frac{x}{\vartheta} - \frac{n-x}{1-\vartheta}$$

fallend in ϑ und Nullsetzen ergibt den Maximierer - den Maximum-Likelihood-Schätzer -

$$T_{\mathrm{ML}}(x) := \vartheta_{\mathrm{ML}}(x) = \frac{x}{n},$$

welchen wir bereits in Beispiel 1.1.8 gefunden hatten. ◇

Beispiel 1.2.8 (Erfolg beim Münzwurf)
Wir wollen schätzen mit welcher Wahrscheinlichkeit ein Münzwurf erfolgreich ist. Dafür werfen wir die Münze solange bis der erste Erfolg eintritt, stoppen dann und notieren uns die Anzahl der Würfe. Ein geeignetes statisches Modell anzugeben und den Maximum-Likelihood-Schätzer für den unbekannten Erfolgsparameter der Münze anzugeben und auf Erwartungstreue zu prüfen ist Hausaufgabe 2.2. ◇

Wir sehen nun, dass ein Maximum-Likelihood-Schätzer nicht immer gut sein muss.

Beispiel 1.2.9 (Maximum-Likelihood-Schätzer für die Gameshow)
Das Modell $\big(\mathfrak{X}, \mathcal{F}, (\mathbb{P}_\vartheta)_{\vartheta\in\Theta}\big)$ der Gameshow ist ein Standardmodell bezüglich des Lebesgue-Maßes μ_0 auf $(\mathfrak{X}, \mathcal{F})$. Es gilt $\mathbb{P}_\vartheta \ll \mu_0$ z.B. mit Dichte $\rho(x, \vartheta) = \vartheta^{-n}\mathbb{1}_{[0,\vartheta]^n}$. Demnach ist der Maximum-Likelihood-Schätzer für ϑ

$$T_{\mathrm{ML}}(x) = \vartheta_{\mathrm{ML}}(x) = \max(\{x_1, \dots, x_n\}),$$

welcher systematisch unterschätzt, da stets $\vartheta_{\mathrm{ML}} \leqslant \vartheta$ ist. ◇

Beispiel 1.2.10 (Maximum-Likelihood-Schätzer für Laplace-Produktmodell)
Betrachte das statistische Produktmodell $\big(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), (\mathcal{Q}_\vartheta^{\otimes n})_{\vartheta\in\mathbb{R}}\big)$, wobei $\mathcal{Q}_\vartheta$ für jedes $\vartheta \in \mathbb{R}$ ein Wahrscheinlichkeitsmaß auf $\big(\mathbb{R}, \mathcal{B}(\mathbb{R})\big)$ mit Dichte

$$\rho(\vartheta, x) := \frac{1}{2} e^{-|x-\vartheta|}, \qquad x \in \mathbb{R}$$

bezüglich des Lebesgue-Maßes ist. Das Bestimmen des Maximum-Likelihood-Schätzers für ϑ, der nur für ungerades n eindeutig bestimmt ist, ist Hausaufgabe 2.3. ◇

Beispiel 1.2.11 (Maximum-Likelihood-Schätzer für Glühbirnen)
Das Glühbirnenmodell $\big(\mathfrak{X}, \mathcal{F}, (\mathbb{P}_\vartheta)_{\vartheta\in\Theta}\big)$ ist ein Standardmodell bezüglich des Lebesgue-Maßes mit Dichte $\rho(x, \vartheta) = \prod_{k=1}^n \vartheta e^{-\vartheta x_k}$. Für $x \in \mathfrak{X}$ und $\vartheta \in (0, \infty)$ gilt

$$\ln\big(\rho(x,\vartheta)\big) = \sum_{k=1}^n \ln(\vartheta) - \vartheta x_k = n\ln(\vartheta) - \vartheta \sum_{k=1}^n x_k.$$

und somit ist

$$\frac{\partial}{\partial\vartheta} \ln\big(\rho(x,\vartheta)\big) = \frac{n}{\vartheta} - \sum_{k=1}^n x_k$$

fallend in ϑ. Der Maximum-Likelihood-Schätzer für ϑ ist daher

$$\vartheta_{\mathrm{ML}}(x) = \frac{1}{\frac{1}{n}\sum_{k=1}^n x_k}.$$

Bemerkung 1.2.12 (Andere Dichte gibt andere Maximum-Likelihood-Schätzer)
Wählen wir anstatt die Dichte

$$\tilde{\rho}\colon\ \mathfrak{X}\times\Theta \to \mathfrak{X}, \qquad (x,\vartheta) \mapsto \begin{cases} \vartheta e^{-\vartheta x}, & \text{für } x \neq \vartheta, \\ h(x), & \text{sonst,} \end{cases}$$

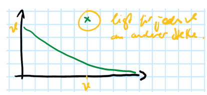


Abb. 1: Der Graph der Abbildung $\tilde{\rho}(x_0, \cdot)$ für ein $x_0 \in \mathfrak{X}$.

wobei $h\colon \mathfrak{X} \to \mathfrak{X}$ eine messbare Funktion mit $h(x) > \sup_{\vartheta\in\Theta} \vartheta e^{-\vartheta x} = \frac{1}{e\cdot x}$ ist, so ist auch $\tilde{\rho}(\cdot, \vartheta)$ eine Dichte von $\mathbb{P}_\vartheta$ bezüglich μ_0.

Der Maximum-Likelihood-Schätzer für $\tilde{\rho}$ ist $\tilde{T}_{\mathrm{ML}}(x) = x$, da $\max_{\vartheta\in\Theta} \tilde{\rho}(x,\vartheta) = h(x)$ für alle $x \in \mathfrak{X}$ ist, während der Maximum-Likelihood-Schätzer für die „kanonische“ Dichte $T_{\mathrm{ML}}(x) = \frac{1}{x}$ und somit völlig verschieden ist. ○

Beispiel 1.2.13 (Maximum-Likelihood-Schätzer für Mittelwert im Gauss-Modell)
Betrachte das Gauss I Modell mit $\mathfrak{X} = \mathbb{R}^n$, $\mathcal{F} = \mathcal{B}(\mathfrak{X})$, $\mathbb{P}_\vartheta := \mathcal{N}_{\vartheta,\sigma^2}^{\otimes n}$ für $\vartheta \in \Theta := \mathbb{R}$ und $\sigma > 0$. Dies ist ein Standardmodell bezüglich des Lebesgue-Maßes mit den Dichten

$$\rho(x,\vartheta) = \prod_{k=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_k-\vartheta)^2}{2\sigma^2}\right).$$

Somit gilt

$$\ln\big(\rho(x,\vartheta)\big) = \sum_{k=1}^n -\frac{1}{2}\ln(2\pi\sigma^2) - \frac{(x_k-\vartheta)^2}{2\sigma^2}$$

und

$$\frac{\partial}{\partial \vartheta} \ln\big(\rho(x,\vartheta)\big) = \frac{1}{\sigma^2} \sum_{k=1}^{n} (x_k - \vartheta).$$

Nullsetzen ergibt den Maximum-Likelihood-Schätzer

$$\vartheta_{\mathrm{ML}} = \frac{1}{n} \sum_{k=1}^{n} x_k.$$

◇

Beispiel 1.2.14 (ML-Schätzer für Varianz und Mittelwert des Gauss-Modells) Betrachte das vorangegangene Modell, jetzt aber mit $\Theta = \mathbb{R} \times (0,\infty)$ und $\mathbb{P}_{\vartheta=(\mu,\sigma^2)} = \mathcal{N}_{\mu,\sigma^2}^{\otimes n}$. Dies ist ein Standardmodell bezüglich des Lebesgue-Maßes mit den Dichten

$$\rho\big(x,(\mu,\sigma^2)\big) = \prod_{k=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_k-\mu)^2}{2\sigma^2}\right).$$

Somit gilt

$$\ln \rho\big(x,(\mu,\sigma^2)\big) = \sum_{k=1}^{n} -\frac{1}{2}\ln(2\pi\sigma^2) - \frac{(x_k-\mu)^2}{2\sigma^2}$$

und

$$\frac{\partial}{\partial \mu} \ln\big(\rho(x,(\mu,\sigma^2))\big) = \frac{1}{\sigma^2} \sum_{k=1}^{n} (x_k - \vartheta)$$

sowie (cf. [4, Bsp. 10.16])

$$\frac{\partial}{\partial \sigma^2} \ln\big(\rho(x,(\mu,\sigma^2))\big) = \frac{\partial}{\partial \sigma^2}\left(-\frac{n}{2}\ln(2\pi\sigma^2) - \sum_{k=1}^{n} \frac{(x_k-\mu)^2}{2\sigma^2}\right) = -\frac{n}{2\sigma^2} + \sum_{k=1}^{n} \frac{(x_k-\mu)^2}{2\sigma^4}.$$

Nullsetzen ergibt den Maximum-Likelihood-Schätzer

$$T_{\mathrm{ML}}(x) := (\hat{\mu},\hat{\sigma}) := \left(\frac{1}{n}\sum_{k=1}^{n} x_k, \frac{1}{n}\sum_{k=1}^{n} (x_k - \hat{\mu})^2\right)$$

als Komponenten den empirischen Mittelwert $\hat{\mu}$ und die empirische Varianz $\hat{\sigma}$ von x haben.◇ empirische Varianz

Bemerkung 1.2.15 (Der Fall $n = 1$) Im Fall $n = 1$ ist $(\hat{\mu},\hat{\sigma}) = (\hat{\mu}, 0) \notin \Theta$, somit ist T_{ML} kein Schätzer. Wir sollten deshalb den Fall $n = 1$ "verbieten", denn für $\tilde{\Theta} := \mathbb{R} \times [0,\infty)$ haben die Maße $\mathbb{P}_\vartheta$ nicht mehr alle Dichten bezüglich des Lebesgue-Maßes und somit ist $\big(\mathfrak{X}, \mathcal{F}, (\mathbb{P}_\vartheta)_{\vartheta\in\tilde{\Theta}}\big)$ kein Standardmodell mehr, da es kein dominierendes σ-endliches Maß mehr gibt (cf. Gegenbeispiel 1.2.6). ○

27.04.2022

Lemma 1.2.16 (Erwartungstreue der Stichprobenmittel und -varianz) *Sei $\big(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), (\mathbb{Q}_\vartheta^{\otimes n})_{\vartheta\in\Theta}\big)$ ein Produktmodell mit Erwartungswert $m(\vartheta) := \mathbb{E}_\vartheta[X] := \mathbb{E}[\mathbb{Q}_\vartheta] < \infty$ und Varianz $v(\vartheta) := \mathbb{V}_\vartheta[X] := \mathbb{V}[\mathbb{Q}_\vartheta] < \infty$ für $X \sim \mathbb{P}_\vartheta$.*

Das Stichprobenmittel $M(x) := \frac{1}{n}\sum_{k=1}^{n} x_k$ ist ein erwartungstreuer Schätzer für $m(\vartheta)$, jedoch ist die Stichprobenvarianz $V(x) := \frac{1}{n}\sum_{k=1}^{n} \big(x_k - M(x)\big)^2$ kein erwartungstreuer (aber ein asymptotisch erwartungstreuer) Schätzer für $v(\vartheta)$. Die korrigierte Stichprobenvarianz $V^*(x) := \frac{1}{n-1}\sum_{k=1}^{n} \big(x_k - M(x)\big)^2$ hingegen ist ein erwartungstreuer Schätzer für $v(\vartheta)$.

Beweis. Aufgrund der Linearität des Erwartungswerts gilt für alle $\vartheta \in \Theta$

$$\mathbb{E}_\vartheta[M] = \frac{1}{n}\sum_{k=1}^n \mathbb{E}_\vartheta[X_k] = \frac{1}{n}\sum_{k=1}^n m(\vartheta) = m(\vartheta)$$

und

$$\begin{aligned}\mathbb{E}_\vartheta[V] &= \frac{1}{n}\sum_{k=1}^n \mathbb{E}_\vartheta\left[(X_k - M)^2\right] = \frac{1}{n}\sum_{k=1}^n \mathbb{V}_\vartheta[X_k - M] \overset{\text{u.i.v}}{=} \mathbb{V}_\vartheta[X_1 - M] \\ &= \mathbb{V}_\vartheta\left[\frac{n-1}{n}X_1 - \frac{1}{n}\sum_{k=2}^n X_k\right] \\ &\overset{\text{u.i.v}}{=} \left(\frac{n-1}{n}\right)^2 \mathbb{V}_\vartheta[X_1] + \frac{n-1}{n^2}\mathbb{V}_\vartheta[X_1] = \frac{n-1}{n}v(\vartheta).\end{aligned}$$

□

Beispiel 1.2.17 (Maximum-Likelihood-Schätzer für Qualitätskontrolle)
Betrachte wieder das Modell zur Qualitätskontrolle gelieferter Orangen aus Beispiel 1.1.13. Die Likelihoodfunktion ist

$$\rho(x,\vartheta) = \mathbb{P}_\vartheta(\{x\}) = \frac{\binom{\vartheta}{x}\binom{N-\vartheta}{n-x}}{\binom{N}{n}} \mathbb{1}_{x\leqslant\vartheta}(x,\vartheta).$$

Um die Maxima von $\rho(x,\cdot)$ zu finden, betrachten wir für $x \leqslant \vartheta$ den Quotienten

$$\frac{\rho(x,\vartheta)}{\rho(x,\vartheta-1)} = \frac{\binom{\vartheta}{x}\binom{N-\vartheta}{n-x}}{\binom{\vartheta-1}{x}\binom{N-\vartheta+1}{n-x}} = \frac{\vartheta(N-\vartheta+1-n+x)}{(\vartheta-x)(N-\vartheta+1)},$$

wobei wir im letzten Schritt zwei mal die Identität $\binom{n-1}{k} = \frac{n-k}{n}\binom{n}{k}$ benutzen. Dann gilt für $x \leqslant \vartheta$

$$\begin{aligned}\frac{\rho(x,\vartheta)}{\rho(x,\vartheta-1)} \geqslant 1 &\iff \vartheta(x-n) \geqslant -x(N-\vartheta+1) \iff \vartheta(\cancel{x} - n\cancel{x}) \geqslant -x(N+1) \\ &\iff \vartheta \leqslant x\frac{N+1}{n}.\end{aligned}$$

Somit ist $\rho(x,\cdot)$ wachsend auf $\left\{0,\ldots,\min\left(\left\lfloor x\frac{N+1}{n}\right\rfloor, N\right)\right\}$ und fallend für größere Werte.

Daher ist der Maximum-Likelihood-Schätzer

$$T(x) = \begin{cases}\left\lfloor x\frac{N+1}{n}\right\rfloor, & \text{für } x < n,\\ N, & \text{für } x = n\end{cases}$$

◇

Beispiel 1.2.18 (Fischpopulation)
Ein Teich enthält eine unbekannte Anzahl Fische. Um die Anzahl ϑ zu schätzen, markieren wir r Fische. Später fangen wir n Fische, von denen x markiert sind.

Das zugehörige statistische Modell hat $\mathfrak{X} = \{0,\ldots,n\}$, $\mathcal{F} = \mathcal{P}(\mathfrak{X})$ und $\mathbb{P}_\vartheta := \mathrm{Hyp}_{n,r,\vartheta-r}$ für $\Theta := \mathbb{N}_{\geqslant r}$. Die Likelihoodfunktion ist

$$\rho(x,\vartheta) = \frac{\binom{r}{x}\binom{\vartheta-r}{n-x}}{\binom{\vartheta}{n}} \mathbb{1}_{\{x\leqslant\vartheta\}}(x,\vartheta).$$

Analog zu Beispiel 1.2.17 können wir für $x \neq 0$ zeigen, dass $\rho(x, \cdot)$ auf $\{r, \dots, \lfloor \frac{nr}{x} \rfloor\}$ wachsend ist und für größere Werte fallend. Dann ist $T(x) := \lfloor \frac{nr}{x} \rfloor$ ein Maximum-Likelihood-Schätzer für ϑ, wenn $x \neq 0$ ist.

Die Funktion $\rho(0, \cdot)$ ist wachsend auf ganz Θ. Man könnte $T(0) = \infty$ setzen, jedoch ist $\infty \notin \Theta$. Man könnte also Θ um $\{\infty\}$ erweitern und $\mathbb{P}_\infty := \delta_0$ setzen.

Ein anderes Problem ist, dass für kleine x der Schätzer $T(x)$ stark davon abhängt, ob man einen Fischer mehr oder weniger fängt und daher ist der Schätzer unzuverlässlich. Ist x klein, ist es besser, das Experiment mit größerem r zu wiederholen. $\diamond$

Beispiel 1.2.19 (Hardy-Weinberg Gleichgewicht)
In einer (unendlich großen) Population gibt es drei genetische Typen; aa, aA und AA mit den Häufigkeiten

$$p_{aa}(\vartheta) := \vartheta^2, \qquad p_{aA}(\vartheta) := 2\vartheta(1-\vartheta), \qquad p_{AA}(\vartheta) := (1-\vartheta)^2$$

für $\vartheta \in [0, 1]$. Beachte, dass mit der binomischen Formel $p_{aa} + p_{aA} + p_{AA} \equiv 1$ gilt.

Sei $E := \{aa, aA, AA\}$. In einer Stichprobe der Größe n beobachten wir n_t Individuen vom Typ $t \in E$.

Wir könnten als statistisches Modell eine Multinomialverteilung auf E wählen oder (einfacheres Modell, aber größerer Beobachtungsraum und die unnötige Information der Reihenfolge wird vermerkt) das n-fache Produktmodell von $\big(E, \mathcal{P}(E), (\mathbb{P}_\vartheta)_{\vartheta\in[0,1]}\big)$, wobei $\mathbb{P}_\vartheta(\{k\}) = p_k(\vartheta)$ für $k \in E$ und $\vartheta \in \Theta := [0, 1]$.

Sei $n_k(x) := |\{i \in \{1, \dots, n\} : x_i = k\}|$ die Anzahl der Individuen von Typ $k \in E$ in der Stichprobe.

Die Likelihoodfunktion ist

$$\rho(x, \vartheta) = \prod_{k=1}^{n} p_{x_k}(\vartheta) = \prod_{k\in E} p_k(\vartheta)^{n_k(x)} = \vartheta^{2n_{aa}(x)}\big(2\vartheta(1-\vartheta)\big)^{n_{aA}(x)}(1-\vartheta)^{2n_{AA}(x)}$$

und somit die Log-Likelihoodfunktion für $\vartheta \in (0, 1)$

$$\ln\big(\rho(x, \vartheta)\big) = 2n_{aa}(x)\ln(\vartheta) + n_{aA}(x)\ln\big(2\vartheta(1-\vartheta)\big) + 2n_{AA}(x)\ln(1-\vartheta).$$

Es folgt

$$\begin{aligned}\frac{\partial}{\partial\vartheta}\ln\big(\rho(x, \vartheta)\big) &= \frac{2n_{aa}(x)}{\vartheta} + \frac{n_{aA}(x)}{2\vartheta(1-\vartheta)}(2-4\vartheta) - \frac{2n_{AA}(x)}{1-\vartheta} \\ &= \frac{2(1-\vartheta)n_{aa}(x) + (1-2\vartheta)n_{aA}(x) - 2\vartheta n_{AA}(x)}{\vartheta(1-\vartheta)}\end{aligned}$$

und somit unter Ausnutzung von $n_{aa} + n_{aA} + n_{AA} \equiv n$

$$\frac{\partial}{\partial\vartheta}\ln\big(\rho(x, \vartheta)\big) = 0 \iff \vartheta = \frac{2n_{aa}(x) + n_{aA}(x)}{2\big(n_{aa}(x) + n_{aA}(x) + n_{AA}(x)\big)} = \frac{2n_{aa}(x) + n_{aA}(x)}{2n}.$$

Ist $n_{aa}(x) = n$, so ist $\rho(x, \vartheta) = \vartheta^{2n}$ und somit $\arg\max_{\vartheta\in[0,1]} \rho(x, \vartheta) = 1$. Ist $n_{AA}(x) = n$, so ist $\rho(x, \vartheta) = (1-\vartheta)^{2n}$ und somit $\arg\max_{\vartheta\in[0,1]} \rho(x, \vartheta) = 0$.

Gilt $n_{aa} < n$ und $n_{AA} < n$, so gilt $\rho(\cdot, 0) = \rho(\cdot, 1) \equiv 0$ und $\rho(x, \vartheta) > 0$ für $\vartheta \in (0, 1)$. Daher wird das Maximum in

$$\vartheta = \frac{2n_{aa}(x) + n_{aA}(x)}{2n}$$

angenommen, und dieser Ausdruck ist der Maximum-Likelihood-Schätzer $T(x)$ für ϑ. ◇

28.04.2022

Gegenbeispiel 1.2.20 (Entrauschen von Bildern)
Sei $\vartheta \in \mathbb{R}^{m\times n}$ ein aus logarithmierten[3] Grauwerten bestehendes „Bild", welches zufällig verrauscht ist. Wir wählen als Modell $\mathfrak{X} = \mathbb{R}^{m\times n}$, $\mathcal{F} = \mathcal{B}(\mathfrak{X})$ und $\mathbb{P}_\vartheta$ als die Verteilung von $(\vartheta_{i,j} + Y_{i,j})_{i,j=1}^{m,n}$, wobei $Y_{i,j} \overset{\text{u.i.v}}{\sim} \mathcal{N}(0,1)$. Dies ist ein Standardmodell bezüglich des Lebesgue-Maßes auf $\mathbb{R}^{m\times n}$ und der (unsinnige!) Maximum-Likelihood-Schätzer ist $T_{\mathrm{ML}}(x) = x$ für $x \in \mathfrak{X}$, weil $\mathbb{E}[Y_{i,j}] = 0$ gilt: die Likelihoodfunktion des (i,j)-ten Eintrages ist $\mathcal{N}(\vartheta_{i,j}, 1)$ verteilt und somit ist $\arg\max_{\vartheta\in\mathbb{R}} (2\pi)^{-\frac{1}{2}} \exp\left(\frac{(x-\vartheta)^2}{2}\right) = x$.

Später werden wir sehen, dass ein Bayesscher Ansatz sinnvoller ist, bei welchem man eine a-priori Verteilung der ϑ wählt. Dies ist realistischer, da Bilder strukturiert sind und nicht nur zufällige Ansammlungen von Grauwerten. ◇

[3]Damit die alle Werte in $\mathbb{R}$ angenommen werden können. Den Grauwert 0 müssen wir ignorieren.

1.3 Beste Schätzer und die Cramér-Rao Ungleichung

Wir haben in Beispiel 1.1.15 gesehen, dass die minimale quadratische Abweichung („Risiko“) überall kein gutes Kriterium für die Güte von Schätzern ist: man kann *nicht* erwarten, dass in einem Modell ein Schätzer T für reellwertige Kenngrößen τ existiert mit $\mathbb{E}_\vartheta[|T-\tau(\vartheta)|^2] \leqslant \mathbb{E}_\vartheta[|S-\tau(\vartheta)|^2]$ für alle $\vartheta \in \Theta$ und alle Schätzer S. Wenn wir nur erwartungstreue T und S zulassen, können wir aber sinnvolle Aussagen treffen.

Definition 1.3.1 (Bester Schätzer)
Sei $\left(\mathfrak{X}, \mathcal{F}, (\mathbb{P}_\vartheta)_{\vartheta\in\Theta}\right)$ ein statistisches Modell. Ein erwartungstreuer Schätzer T für die reelle Kenngröße τ heißt varianzminimierend oder bester Schätzer für τ (oder UMVU - *uniformly minimal variance unbiased*), wenn $\mathbb{V}_\vartheta[T] < \infty$ und bester Schätzer

$$\mathbb{V}_\vartheta[T] \leqslant \mathbb{V}_\vartheta[S] \qquad \forall \vartheta \in \Theta, \forall \text{ erwartungstreue Schätzer } S \text{ für } \tau.$$

Satz 1.3.1: Eindeutigkeit bester Schätzer

Der beste Schätzer ist $\mathbb{P}_\vartheta$-fast sicher eindeutig.

Beweis. Angenommen, S und T sind beste Schätzer. Da aufgrund der Linearität des Erwartungswerts auch der Schätzer $\frac{1}{2}(S+T)$ erwartungstreu ist, gilt für alle $\vartheta \in \Theta$

$$\begin{aligned}\mathbb{V}_\vartheta[T] \overset{T \text{ UMVU}}{\leqslant} \mathbb{V}_\vartheta\left[\frac{1}{2}(S+T)\right] &= \frac{1}{4}\big(\mathbb{V}_\vartheta[S] + \mathbb{V}_\vartheta[T] + 2\operatorname{Cov}_\vartheta[S,T]\big)\\ &\overset{\mathbb{V}_\vartheta[S]=\mathbb{V}_\vartheta[T]}{=} \frac{1}{2}\left(\mathbb{V}_\vartheta[T] + \operatorname{Cov}_\vartheta[S,T]\right).\end{aligned}$$

Es folgt

$$\operatorname{Cov}_\vartheta[S,T] \geqslant \mathbb{V}_\vartheta[S] \qquad \forall \vartheta \in \Theta. \tag{5}$$

Daher gilt

$$\mathbb{V}_\vartheta[T-S] = \mathbb{V}_\vartheta[T] + \mathbb{V}_\vartheta[S] - 2\operatorname{Cov}_\vartheta[S,T] = 2\,\mathbb{V}_\vartheta[T] - 2\operatorname{Cov}_\vartheta[S,T] \overset{(5)}{\leqslant} 0$$

und somit $\mathbb{V}_\vartheta[T-S] = 0$ für alle $\vartheta \in \Theta$, also ist $T-S$ $\mathbb{P}_\vartheta$-fast sicher konstant. Aufgrund der erwartungstreue von T und S gilt $\mathbb{E}_\vartheta[T-S] = 0$, und daher ist $T = S$ $\mathbb{P}_\vartheta$-fast sicher für alle $\vartheta \in \Theta$. □

Bemerkung 1.3.2 (Intuition: Existenz bester Schätzer) Nehmen wir an, dass es genau drei Schätzer T_1, T_2 und T_3 gibt, deren Varianzen von der folgenden Form sind

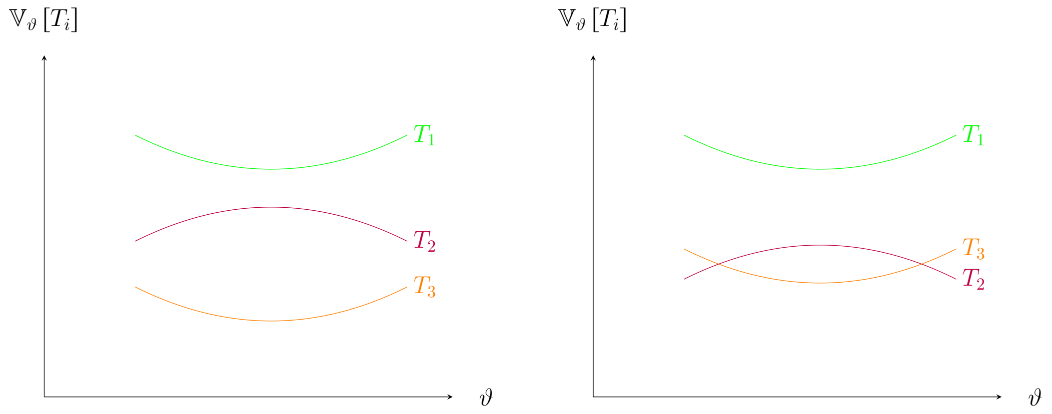


Abb. 2: Im linken Fall ist der beste Schätzer T_3, im rechten Fall existiert kein bester Schätzer. [Bild getexed von Thomas.]

Der Nachweis, dass kein bester Schätzer existiert, kann aufwendig sein.

Satz 1.3.2: Charakterisierung bester Schätzer

Sei U_0 die Menge aller erwartungstreuen Schätzer von 0 mit endlicher Varianz, das heißt

$$U_0 = \{S\colon\ \mathfrak{X} \to \Theta \text{ messbar} : \mathbb{E}_\vartheta[S] = 0,\ \mathbb{V}_\vartheta[S] < \infty\ \forall \vartheta \in \Theta\}.$$

Sei T ein erwartungstreuer Schätzer von τ mit $\mathbb{V}_\vartheta[T] < \infty$ für alle $\vartheta \in \Theta$. Dann sind die folgenden Aussagen äquivalent

1. T ist ein bester Schätzer für τ.
2. $\mathbb{E}_\vartheta[TU] = 0$ für alle $U \in U_0$ und alle $\vartheta \in \Theta$.

Beweis. "1 $\Longrightarrow$ 2": Seien $U \in U_0$ und $\vartheta \in \Theta$ beliebig. Für $c \in \mathbb{R}$ sei $T_c := T + cU$. Dann ist auch auch T_c erwartungstreu und es gilt

$$\mathbb{V}_\vartheta[T_c] \geqslant \mathbb{V}_\vartheta[T] \qquad \forall c \in \mathbb{R}\,.$$

Dies ist genau dann der Fall, wenn für alle $c \in \mathbb{R}$

$$c^2\, \mathbb{V}_\vartheta[U] + 2c\, \mathrm{Cov}_\vartheta[U, T] \geqslant 0$$

gilt, was genau dann der Fall ist, wenn $\mathrm{Cov}_\vartheta[U, T] = 0$ ist. Somit ist

$$\mathbb{E}_\vartheta[TU] = \mathbb{E}_\vartheta[T]\, \mathbb{E}_\vartheta[U] + \mathrm{Cov}_\vartheta[T, U] = 0$$

für alle $\vartheta \in \Theta$.

"2 $\Longrightarrow$ 1": Sei S ein erwartungstreuer Schätzer von τ mit $\mathbb{V}_\vartheta[S] < \infty$. Dann gilt $T - S \in U_0$, da aufgrund von Cauchy-Schwarz $\mathbb{V}_\vartheta[T - S] \leqslant \mathbb{V}_\vartheta[T] + \mathbb{V}_\vartheta[S] + 2\sqrt{\mathbb{V}_\vartheta[T]\, \mathbb{V}_\vartheta[S]}$ gilt. Für alle $\vartheta \in \Theta$ folgt aufgrund von $\mathbb{E}_\vartheta[T] = \mathbb{E}_\vartheta[S] = \tau(\vartheta)$ $(\star)$, dass

$$\begin{aligned} 0 &= \mathbb{E}_\vartheta[T(T - S)] = \mathbb{E}_\vartheta[T^2] - \mathbb{E}_\vartheta[ST] \\ &\overset{(\star)}{=} \mathbb{E}_\vartheta[T^2] - \mathbb{E}_\vartheta[T]^2 - \mathbb{E}_\vartheta[ST] + \mathbb{E}_\vartheta[T]\, \mathbb{E}_\vartheta[S] = \mathbb{V}_\vartheta[T] - \mathrm{Cov}_\vartheta[T, S]. \end{aligned}$$

Aufgrund der CAUCHY-SCHWARZ-Ungleichung $\mathrm{Cov}_\vartheta[T,S]^2 \leqslant \mathbb{V}_\vartheta[T]\,\mathbb{V}_\vartheta[S]$ folgt

$$\mathbb{V}_\vartheta[T] \leqslant \mathbb{V}_\vartheta[S] \qquad \forall \vartheta \in \Theta.$$

□

Beispiel 1.3.3 (Modell ohne besten Schätzer: Gleichverteilung)
Wir betrachten $\big(\mathbb{R}, \mathcal{B}(\mathbb{R}), (\mathbb{P}_\vartheta)_{\vartheta\in\mathbb{R}}\big)$, wobei $\mathbb{P}_\vartheta$ die Gleichverteilung auf $[\vartheta - \frac{1}{2}, \vartheta + \frac{1}{2}]$ ist, das heißt die Likelihoodfunktion ist

$$\rho(x,\vartheta) = \mathbb{1}_{[\vartheta-\frac{1}{2},\vartheta+\frac{1}{2}]}(x).$$

In diesem Modell existiert kein bester Schätzer für ϑ.

Beweis. Angenommen, T ist ein bester Schätzer für ϑ. Für $U \in U_0$ gilt

Nur, wenn U stetig ist (Hauptsatz)!

$$\int_{\vartheta-\frac{1}{2}}^{\vartheta+\frac{1}{2}} U(x)\,\mathrm{d}x = \mathbb{E}_\vartheta[U] = 0$$

für alle $\vartheta \in \Theta$. Differenzieren beider Seiten nach ϑ ergibt

$$U\left(\vartheta + \frac{1}{2}\right) - U\left(\vartheta - \frac{1}{2}\right) = 0.$$

Für alle $U \in U_0$ gilt also $U(x) = U(x+1)$ für alle $x \in \mathbb{R}$.

Analog folgt aus $\mathbb{E}_\vartheta[UT] = 0$ für alle $U \in U_0$ und alle $\vartheta \in \Theta$, dass für alle $U \in U_0$

Angenommen, U sei stetig. Können wir die Stetigkeit von TU garantieren?

$$T(x)U(x) = T(x+1)U(x+1) \qquad \forall x \in \mathbb{R}$$

gilt. Mit $U(x) = U(x+1)$ für alle $x \in \mathbb{R}$ folgt daraus

$$T(x) = T(x+1) \qquad \forall x \in \mathbb{R}\,. \tag{6}$$

Da T erwartungstreu ist, gilt

$$\int_{\vartheta-\frac{1}{2}}^{\vartheta+\frac{1}{2}} T(x)\,\mathrm{d}x = \mathbb{E}_\vartheta[T] = \vartheta \qquad \forall \vartheta \in \Theta.$$

Differenzieren nach ϑ dieser Gleichung ergibt

$$T\left(\vartheta + \frac{1}{2}\right) - T\left(\vartheta - \frac{1}{2}\right) = 1,$$

was (6) widerspricht. □

Unsere Aufgabe ist es nun, einen besten Schätzer T zu berechnen, sofern dies möglich ist. Dafür benötigen wir jedoch einige Regularitätsbedingungen: wir betrachten nur spezielle statistische Modelle.

DEFINITION 1.3.4 (REGULÄRES MODELL, SCOREFUNKTION, FISHER-INFO)
Ein Standardmodell $\big(\mathfrak{X}, \mathcal{F}, (\mathbb{P}_\vartheta)_{\vartheta\in\Theta}\big)$ ist regulär, wenn

regulär

- $\Theta \subset \mathbb{R}$ ein offenes Intervall ist,

- die Likelihoodfunktion ρ strikt positiv auf $\mathfrak{X} \times \Theta$ und nach ϑ stetig differenzierbar ist (man kann diese Eigenschaften auch nur fast überall fordern).

Dann ist die Scorefunktion

Scorefunktion

$$U_\vartheta(x) := \frac{\partial}{\partial \vartheta} \ln\big(\rho(x,\vartheta)\big) = \frac{\frac{\partial}{\partial \vartheta}\rho(x,\vartheta)}{\rho(x,\vartheta)}$$

wohldefiniert. Wir benutzen die Kurzschreibweise $\rho'(x,\vartheta) := \frac{\partial}{\partial \vartheta}\rho(x,\vartheta)$.

Zuletzt fordern wir, dass die Fisher-Information des Modells $I(\vartheta) := \mathbb{V}_\vartheta[U_\vartheta] \in (0,\infty)$ ist und die Vertauschungsrelation

Fisher-Information

$$\int_{\mathfrak{X}} \frac{\partial}{\partial \vartheta}\rho(x,\vartheta)\,\mathrm{d}\mu_0(x) \overset{!}{=} \frac{\partial}{\partial \vartheta}\underbrace{\int_{\mathfrak{X}} \rho(x,\vartheta)\,\mathrm{d}\mu_0(x)}_{=1} = 0 \tag{7}$$

gilt, wobei wir fordern, dass die linke Seite wohldefiniert ist, das heißt, dass $\frac{\partial}{\partial \vartheta}\rho(\cdot,\vartheta)$ bezüglich μ_0 integrierbar ist.

Es ist Hausaufgabe 3.1 (a), zu zeigen, dass die Fisher-Information unabhängig von dem dominierenden Maß ist.

Bemerkung 1.3.5 (Hinreichende Bedingungen für die Vertauschungsrelation)
Die Vertauschungsrelation (7) ist erfüllt, wenn $\mathfrak{X}$ endlich und μ_0 das Zählmaß auf $\big(\mathfrak{X}, \mathcal{P}(\mathfrak{X})\big)$ ist. Allgemein ist hinreichend, dass für alle $\vartheta_0 \in \Theta$ eine Umgebung $N(\vartheta_0) \subset \Theta$ existiert mit

$$\int_{\mathfrak{X}} \sup_{\vartheta \in N(\vartheta_0)} \left|\frac{\partial}{\partial \vartheta}\rho(x,\vartheta)\right| \mathrm{d}\mu_0(x) < \infty.$$

Dies folgt aus dem Satz über majorisierte Konvergenz, denn für $h \neq 0$ mit $\vartheta + h \in N(\vartheta_0)$ gilt

$$\int_{\mathfrak{X}} \underbrace{\frac{\rho(x,\vartheta+h) - \rho(x,\vartheta)}{h}}_{\xrightarrow{h \to 0} \frac{\partial}{\partial \vartheta}\rho(x,\vartheta)} \mathrm{d}\mu_0(x) < \infty$$

und $\frac{\partial}{\partial \vartheta}\rho(\cdot,\vartheta)$ ist als punktweise Grenzwert messbarer Funktionen eine messbare Funktion. Nach dem Mittelwertsatz existiert ein $\tilde{\vartheta}(x)$ zwischen ϑ und $\vartheta + h$ mit $\frac{\rho(x,\vartheta+h)-\rho(x,\vartheta)}{h} = \frac{\partial}{\partial \vartheta}\rho\big(x, \tilde{\vartheta}(x)\big)$. ○

Bemerkung 1.3.6 ($\mathbb{E}_\vartheta[U_\vartheta] = 0$ im regulären Modell)
In einem regulären Modell gilt

$$\mathbb{E}_\vartheta[U_\vartheta] = \int_{\mathfrak{X}} U_\vartheta(x)\,\mathrm{d}\mathbb{P}_\vartheta(x) = \int_{\mathfrak{X}} \frac{\rho'(x,\vartheta)}{\cancel{\rho(x,\vartheta)}} \cancel{\rho(x,\vartheta)}\,\mathrm{d}\mu_0(x) = \int_{\mathfrak{X}} \rho'(x,\vartheta)\,\mathrm{d}\mu_0(x) \overset{(7)}{=} 0. \tag{8}$$

Es folgt

$$I(\vartheta) = \mathbb{V}_\vartheta[U_\vartheta] = \mathbb{E}_\vartheta[U_\vartheta^2]. \tag{9}$$

○

Lemma 1.3.7 (I unterhalbstetig)
Im regulären Modell ist I unterhalbstetig, das heißt

$$I(\vartheta) \leqslant \liminf_{h\to 0} I(\vartheta + h) \qquad \textit{für hinreichend kleine } h.$$

Beweis. Da $\vartheta \mapsto U_\vartheta$ als Quotient stetiger Funktionen stetig ist, gilt

$$\liminf_{h\to 0} U_{\vartheta+h}(x) = \lim_{h\to 0} U_{\vartheta+h}(x) = U_\vartheta(x) \qquad \forall x \in \mathfrak{X}.$$

Nach dem Lemma von Fatou folgt daraus

$$\mathbb{E}_\vartheta[U_\vartheta^2] \leqslant \liminf_{h\to 0} \mathbb{E}_\vartheta[U_{\vartheta+h}^2].$$

□

Frage: Ist I im regulären Modell sogar stetig?

Beispiel 1.3.8 (Formel für die Fisher-Information im Produktmodell)
Gegeben sei das reguläre statistische Modell $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \mathbb{P}_\vartheta^{\otimes n} : \vartheta \in \mathbb{R}_{>0})$, wobei $\mathbb{P}_\vartheta$ die Dichte

$$\rho(\vartheta, \cdot) = \frac{1}{\vartheta} f\left(\frac{\cdot}{\vartheta}\right),$$

besitzt und $f(x) > 0$ ist sowie $f'(x)$ existiert für alle $x \in \mathbb{R}$. Die Fisher-Information ist (Hausaufgabe 3.1 (b))

$$I(\vartheta) = \frac{n}{\vartheta^2} \int_{\mathbb{R}} \frac{\big(xf'(x) + f(x)\big)^2}{f(x)}.$$

◇

Definition 1.3.9 (Regulärer Schätzer)
Ein erwartungstreuer Schätzer T für eine reelle Kenngröße τ in einem regulären Modell ist regulär, wenn

regulär

$$\int_{\mathfrak{X}} T(x)\rho'(x,\vartheta)\,\mathrm{d}\mu_0(x) = \frac{\partial}{\partial\vartheta} \int_{\mathfrak{X}} T(x)\rho(x,\vartheta)\,\mathrm{d}\mu_0(x) \qquad \forall \vartheta \in \Theta$$

und beide Seiten definiert sind.

Bemerkung 1.3.10 Für einen erwartungstreuen Schätzer T für die Kenngröße τ gilt

$$\frac{\partial}{\partial\vartheta} \int_{\mathfrak{X}} T(x)\rho(x,\vartheta)\,\mathrm{d}\mu_0(x) = \frac{\partial}{\partial\vartheta} \int_{\mathfrak{X}} T(x)\,\mathrm{d}\mathbb{P}_\vartheta(x) = \frac{\partial}{\partial\vartheta} \mathbb{E}_\vartheta[T] = \frac{\partial}{\partial\vartheta} \tau(\vartheta) = \tau'(\vartheta).$$

○

Satz 1.3.3: Cramér-Rao Informationsungleichung I

Seien $\big(\mathfrak{X}, \mathcal{F}, (\mathbb{P}_\vartheta)_{\vartheta\in\Theta}\big)$ regulär, $\tau\colon \Theta \to \mathbb{R}$ stetig differenzierbar (mit $\tau' \neq 0$) und T ein regulärer Schätzer für τ. Dann gilt

$$\mathbb{V}_\vartheta[T] \geqslant \frac{\tau'(\vartheta)^2}{I(\vartheta)} \qquad \forall \vartheta \in \Theta. \tag{10}$$

Bemerkung 1.3.11 Gilt für T Gleichheit in (10), so existiert kein besserer regulärer Schätzer für τ. ○

Beweis. (von Satz 1.3.3) Sei $\vartheta \in \Theta$. Ohne Beschränkung der Allgemeinheit sei $\mathbb{V}_\vartheta[T] < \infty$. Es gilt

$$\begin{aligned} \operatorname{Cov}_\vartheta[T, U_\vartheta] &= \mathbb{E}_\vartheta[TU_\vartheta] - \mathbb{E}_\vartheta[T]\underbrace{\mathbb{E}_\vartheta[U_\vartheta]}_{\overset{(8)}{=}0} = \mathbb{E}_\vartheta[TU_\vartheta] = \int_{\mathfrak{X}} T(x)U_\vartheta(x)\,\mathrm{d}\mathbb{P}_\vartheta(x) \\ &= \int_{\mathfrak{X}} T(x)\frac{\rho'(x,\vartheta)}{\cancel{\rho(x,\vartheta)}}\cancel{\rho(x,\vartheta)}\,\mathrm{d}\mu_0(x) = \int_{\mathfrak{X}} T(x)\rho'(x,\vartheta)\,\mathrm{d}\mu_0(x) \\ &\overset{T \text{ regulär}}{=} \frac{\partial}{\partial\vartheta}\int_{\mathfrak{X}} T(x)\rho(x,\vartheta)\,\mathrm{d}\mu_0(x) = \tau'(\vartheta) \end{aligned} \tag{11}$$

und somit folgt mit der Cauchy-Schwarz-Ungleichung

$$\tau'(\vartheta)^2 \overset{(11)}{=} \operatorname{Cov}_\vartheta[T, U_\vartheta]^2 \leqslant \mathbb{V}_\vartheta[T]\,\mathbb{V}_\vartheta[U_\vartheta] \overset{(9)}{=} \mathbb{V}_\vartheta[T]I(\vartheta).$$ □

Bemerkung 1.3.12 (Bedeutung der Fisher-Information)
Gilt $I(\vartheta) = 0$, so ist fast sicher $\rho'(x,\vartheta) = 0$; $I(\vartheta)$ gibt uns die Information, die uns die Beobachtung von x über ϑ gibt. Ist $\rho(x,\vartheta)$ konstant auf einem Intervall, sind die zugehörigen $\mathbb{P}_\vartheta$ alle gleich, also hilft die Beobachtung x uns auch nicht, sie zu unterscheiden. ○

Beispiel 1.3.13 (Kein Widerspruch zu Satz 1.3.3)
Betrachte das statistische Modell $\left(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), (\mathbb{P}_\vartheta^{\otimes n})_{\vartheta\in\mathbb{R}}\right)$, wobei $\mathbb{P}_\vartheta$ für $\vartheta \in \Theta$ die Dichte

$$\rho(\vartheta, x) := e^{\vartheta - x}\,\mathbb{1}_{[\vartheta,\infty)}(x)$$

bezüglich des Lebesgue-Maßes hat.

Es ist Hausaufgabe 3.2 zu zeigen, dass für $\tau(\vartheta) := \vartheta$ der Maximum-Likelihood Schätzer $T(x) = x_{(1)}$ ist und den Erwartungswert $\vartheta + \frac{1}{n}$ hat, sowie dass $I(\vartheta) < \frac{1}{\mathbb{V}_\vartheta[T]}$ für alle $\vartheta \in \mathbb{R}$ gilt, was aber keinen Widerspruch zu Satz 1.3.3 darstellt. ◇

03.05.2022

Satz 1.3.4: Cramér-Rao Informationsungleichung II

Unter den Voraussetzungen von Satz 1.3.3 gilt Gleichheit in (10) für $\vartheta \in \Theta$ genau dann wenn

$$T(x) - \tau(\vartheta) = \tau'(\vartheta)\frac{U_\vartheta(x)}{I(\vartheta)} \qquad \text{für alle } \mathbb{P}_\vartheta\text{-fast alle } x \in \mathfrak{X}\,. \tag{12}$$

Dann ist I stetig und die Likelihoodfunktion ρ hat die Darstellung

$$\rho(x,\vartheta) = h(x)\exp\left(a(\vartheta)T(x) - b(\vartheta)\right), \qquad x \in \mathfrak{X},\ \vartheta \in \Theta,$$

(insb. ist das Modell exponentiell) wobei

$$a(\vartheta) := \int_\Delta^\vartheta \frac{I(\kappa)}{\tau'(\kappa)}\,\mathrm{d}\kappa$$

für ein beliebiges $\Delta \in \Theta$, $h\colon (\mathfrak{X}, \mathcal{F}) \to \big((0,\infty), \mathcal{B}\,((0,\infty))\big)$ messbar und

$$b(\vartheta) := \ln\left(\int_{\mathfrak{X}} h(x) \exp\left(a(\vartheta)T(x)\right) \mathrm{d}\mu_0(x)\right)$$

die Normalisierungskonstante bezüglich $\mathfrak{X}$ sind.

Beweis. 1 **Die Gleichung** (12)**.** Sowohl unter (10) sowie unter (12) gilt $\mathbb{V}_\vartheta[T] < \infty$. Sei $c\colon \Theta \to \mathbb{R}$, $\vartheta \mapsto \frac{\tau'(\vartheta)}{I(\vartheta)}$. Dann gilt

$$\begin{aligned} 0 \leqslant \mathbb{V}_\vartheta[T - c(\vartheta)U_\vartheta] &= \mathbb{V}_\vartheta[T] + c(\vartheta)^2 \underbrace{\mathbb{V}_\vartheta[U_\vartheta]}_{\overset{(9)}{=} I(\vartheta)} - 2c(\vartheta) \underbrace{\mathrm{Cov}_\vartheta[T, U_\vartheta]}_{\overset{(11)}{=} \tau'(\vartheta)} \\ &= \mathbb{V}_\vartheta[T] + \frac{\tau'(\vartheta)^2}{I(\vartheta)} - 2\frac{\tau'(\vartheta)^2}{I(\vartheta)} = \mathbb{V}_\vartheta[T] - \frac{\tau'(\vartheta)^2}{I(\vartheta)}. \end{aligned}$$

Daraus folgt wieder die Crámer-Rao Ungleichung. Ferner folgt, dass genau dann Gleichheit in (10) gilt, wenn

$$\mathbb{V}_\vartheta[T - c(\vartheta)U_\vartheta] = 0 \qquad \forall \vartheta \in \Theta,$$

beziehungsweise, wenn $x \mapsto T(x) - c(\vartheta)U_\vartheta(x)$ $\mathbb{P}_\vartheta$-fast sicher konstant ist. Diese (von ϑ abhängige) Konstante muss für jedes ϑ mit dem Erwartungswert

$$\mathbb{E}_\vartheta[T - c(\vartheta)U_\vartheta] = \mathbb{E}_\vartheta[T] - c(\vartheta) \cdot 0 = \tau(\vartheta)$$

übereinstimmen. Somit gilt Gleichheit in (10) für alle $\vartheta \in \Theta$ genau dann, wenn

$$T(x) - c(\vartheta)U_\vartheta(x) = \tau(\vartheta)$$

für $\mathbb{P}_\vartheta$-fast alle $x \in \mathfrak{X}$, und das ist (12).

2 **Stetigkeit von** I**.** Angenommen, I ist unstetig in $\vartheta_0 \in \Theta$. Dann existiert ein $\varepsilon > 0$ und eine Folge $(\vartheta_n)_{n\in\mathbb{N}} \subset \Theta$ mit $\vartheta_n \xrightarrow{n\to\infty} \vartheta_0$, sodass $|I(\vartheta_n) - I(\vartheta_0)| > \varepsilon$ für alle $n \in \mathbb{N}$ gilt.

Sei $N_n := \{x \in \mathfrak{X} : T(x) - \tau(\vartheta_n) \neq c(\vartheta_n)U_{\vartheta_n}(x)\}$ für $n \in \mathbb{N}_0$. Nach der Voraussetzung (12) ist N_n eine $\mathbb{P}_{\vartheta_n}$-Nullmenge. Da ρ auf $\mathfrak{X} \times \Theta$ positiv ist, gilt auch $\mu_0(N_n) = 0$ für alle $n \in \mathbb{N}_0$ (cf. Lemma 1.5.10). Man sagt $\mathbb{P}_{\vartheta_n}$ und μ_0 sind äquivalent - sie haben dieselben Nullmengen). Da die abzählbare Vereinigung von μ_0-Nullmengen eine μ_0-Nullmenge ist, folgt $\mu_0\left(\bigcup_{n\in\mathbb{N}_0} N_n\right) = 0$.

Sei $x \notin N := \bigcup_{n\in\mathbb{N}_0} N_n$. Dann gilt für alle $n \in \mathbb{N}_0$

$$T(x) - \tau(\vartheta_n) = c(\vartheta_n)U_{\vartheta_n(x)} = \frac{\tau'(\vartheta_n)U_{\vartheta_n}(x)}{I(\vartheta_n)}. \tag{13}$$

Um die Stetigkeit von I zu zeigen, genügt es nun, nach I umzustellen.

Da $\mathbb{V}_{\vartheta_0}[U_{\vartheta_0}] = I(\vartheta_0) > 0$ gilt, existiert ein $x \notin N$ mit $U_{\vartheta_0}(x) \neq 0$. Wegen der Stetigkeit von $\vartheta \mapsto U_\vartheta(x)$ können wir annehmen, dass (möglicherweise erst durch

Wahl einer Teilfolge) $U_{\vartheta_n}(x) \neq 0$ für alle $n \in \mathbb{N}_0$ gilt. Aus (13) folgt somit jeweils $T(x) - \tau(\vartheta_n) \neq 0$ und damit

$$I(\vartheta_n) = \frac{\tau'(\vartheta_n)U_{\vartheta_n}(x)}{T(x) - \tau(\vartheta_n)} \xrightarrow{n\to\infty} \frac{\tau'(\vartheta_0)U_{\vartheta_0}(x)}{T(x) - \tau(\vartheta_0)} = I(\vartheta_0),$$

im Widerspruch zur Annahme $|I(\vartheta_n) - I(\vartheta_0)| > \varepsilon$ für alle $n \in \mathbb{N}$.

3 **Darstellung von ρ.** Sei nun $N_\vartheta := \{x \in \mathfrak{X} : T(x) - c(\vartheta)U_\vartheta(x) \neq \tau(\vartheta)\}$ für $\vartheta \in \Theta$. Dann gilt wie zuvor $\mu_0\left(\bigcup_{\vartheta\in\Theta\cap\mathbb{Q}} N_\vartheta\right) = 0$. Für $x \notin N := \bigcup_{\vartheta\in\Theta\cap\mathbb{Q}} N_\vartheta$ gilt

$$T(x) - c(\vartheta)U_\vartheta(x) - \tau(\vartheta) = 0 \qquad \forall \vartheta \in \Theta \cap \mathbb{Q}\,.$$

Wegen der Stetigkeit der linken Seite in $\vartheta \in \Theta$ folgt sogar

$$T(x) - c(\vartheta)U_\vartheta(x) - \tau(\vartheta) = 0 \qquad \forall \vartheta \in \Theta$$

und somit

$$\frac{1}{c(\vartheta)}\big(T(x) - \tau(\vartheta)\big) = \big(T(x) - \tau(\vartheta)\big)\frac{I(\vartheta)}{\tau'(\vartheta)} = U_\vartheta(x) = \frac{\partial}{\partial\vartheta}\ln\left(\rho(x,\vartheta)\right).$$

Integrieren dieser Gleichung bezüglich ϑ von $\Delta \in \Theta$ bis ϑ ergibt

$$\ln\left(\rho(x,\vartheta)\right) = \ln\left(\rho(x,\Delta)\right) + T(x)\int_\Delta^\vartheta \frac{I(\kappa)}{\tau'(\kappa)}\,\mathrm{d}\kappa - \underbrace{\int_\Delta^\vartheta \frac{I(\kappa)}{\tau'(\kappa)}\tau(\kappa)\,\mathrm{d}\kappa}_{=:\gamma(\vartheta)}.$$

Es folgt

$$\rho(x,\vartheta) = \underbrace{\rho(x,\Delta)}_{=:h(x)}\exp\left(T(x)a(\vartheta) - \gamma(\vartheta)\right).$$

Da $\int_\mathfrak{X} \rho(x,\vartheta)\,\mathrm{d}\mu_0(x) = 1$ für alle $\vartheta \in \Theta$ gilt, folgt

$$\int_\mathfrak{X} h(x)\exp\left(T(x)a(\vartheta)\right)\mathrm{d}\mu_0(x) = e^{\gamma(\vartheta)}$$

und somit $\gamma = b$. Für $x \in N$ definieren wir $h(x) \equiv 1 > 0$. □

Bemerkung 1.3.14 (Anschauung für c) Sei $\langle\cdot,\cdot\rangle_\vartheta$ das $L^2(\mathbb{P}_\vartheta)$-Skalarprodukt. Dann gilt (weil die Cauchy-Schwarz-Ungleichung genau eine Gleichung ist, wenn die Terme linear abhängig ist)

$$T - \tau(\vartheta) = \frac{\tau'(\vartheta)}{I(\vartheta)}U_\vartheta = \frac{\langle T - \tau(\vartheta), T - \tau(\vartheta)\rangle_\vartheta}{\langle U_\vartheta, U_\vartheta\rangle_\vartheta}U_\vartheta,$$

also ist $T - \tau(\vartheta)$ die orthogonale Projektion auf die Scorefunktion bzgl. des $L^2(\mathbb{P}_\vartheta)$-Skalarprodukts.∘

Wir geben Modellen, bei denen die Likelihoodfunktion diese Form hat, einen Namen.

Definition 1.3.15 (Exponentielles Modell (Pitman, Darmois, Koopman)) Ein Standardmodell $\left(\mathfrak{X},\mathcal{F},(\mathbb{P}_\vartheta)_{\vartheta\in\Theta}\right)$, wobei Θ ein offenes Intervall ist (es gibt auch eine technischere Theorie, wobei $\Theta \subset \mathbb{R}^k$ offen ist), heißt exponentielles Modell und $(\mathbb{P}_\vartheta)_{\vartheta\in\Theta}$ eine exponentielle Familie bezüglich der Statistik $T\colon (\mathfrak{X},\mathcal{F}) \to \left(\mathbb{R},\mathcal{B}(\mathbb{R})\right)$, wenn die Likelihoodfunktion die Gestalt exponentielles Modell

$$\rho(x,\vartheta) = h(x)\exp\left(a(\vartheta)T(x) - b(\vartheta)\right)$$

für $x \in \mathfrak{X}$ und $\vartheta \in \Theta$ hat, wobei $a \in \mathcal{C}^1(\Theta)$ mit $a'(\vartheta) \neq 0$ für alle $\vartheta \in \Theta$ und $h\colon (\mathfrak{X},\mathcal{F}) \to \left((0,\infty),\mathcal{B}\left((0,\infty)\right)\right)$ messbar ist. Weiter sei vorausgesetzt, dass für jedes $\vartheta \in \Theta$ die Statistik T *nicht* $\mathbb{P}_\vartheta$-fast überall konstant ist.

Bemerkung 1.3.16 (Gestalt von b) Da $\rho(\cdot,\vartheta)$ für jedes $\vartheta \in \Theta$ eine Dichte bezüglich μ_0 ist, gilt wie im dritten Schritt des Beweises von Satz 1.3.4

$$b(\vartheta) = \log\left(\int_{\mathfrak{X}} h(x)\exp\left(a(\vartheta)T(x)\right)\mathrm{d}\mu_0(x)\right).$$

○

Satz 1.3.5: Eigenschaften exponentieller Modelle

Für ein exponentielles Modell $\left(\mathfrak{X},\mathcal{F},(\mathbb{P}_\vartheta)_{\vartheta\in\Theta}\right)$ bezüglich der Statistik T mit der Likelihoodfunktion

$$\rho(x,\vartheta) = h(x)\exp\left(a(\vartheta)T(x) - b(\vartheta)\right) \tag{14}$$

gilt

1. $\mathbb{E}_\vartheta[T^2] < \infty$ für alle $\vartheta \in \Theta$,
2. $b \in \mathcal{C}^1(\Theta)$ und $b'(\vartheta) = a'(\vartheta)\,\mathbb{E}_\vartheta[T]$ für alle $\vartheta \in \Theta$. Insbesondere ist $\rho(x,\cdot) \in \mathcal{C}^1(\Theta)$ für alle $x \in \mathfrak{X}$,
3. die Kenngröße $\tau(\vartheta) := \mathbb{E}_\vartheta[T]$ ist stetig differenzierbar mit $\tau'(\vartheta) = a'(\vartheta)\,\mathbb{V}_\vartheta[T] \neq 0$ (merke, dass $\mathbb{V}_\vartheta[T] > 0$ da T nach Definition 1.3.15 $\mathbb{P}_\vartheta$-fast sicher nicht konstant ist),
4. $I(\vartheta) = a'(\vartheta)\tau'(\vartheta)$ $(= a'(\vartheta)^2\,\mathbb{V}_\vartheta[T] > 0)$,
5. Jede Statistik $S\colon (\mathfrak{X},\mathcal{F}) \to \left(\mathbb{R},\mathcal{B}(\mathbb{R})\right)$ mit $\mathbb{E}_\vartheta[|S|] < \infty$ für alle $\vartheta \in \Theta$ erfüllt die Vertauschungsrelation.

Insbesondere gilt $\mathbb{V}_\vartheta[T] \overset{3}{=} \frac{\tau'(\vartheta)}{a'(\vartheta)} \overset{4}{=} \frac{\tau'(\vartheta)^2}{I(\vartheta)}$ für alle $\vartheta \in \Theta$, somit gilt die Cramér-Rao-Ungleichung mit Gleichheit und das Modell und der Schätzer T sind regulär. Weiter ist T bester Schätzer für τ.

Beweis. Wir können durch Reparametrisierung annehmen, dass a die Identität auf Θ ist, da a streng monoton ist. (Warum sich dadurch die Likelihoodfunktion nicht ändert, ist nicht prüfungsrelevant.)

1. Seien $S\colon (\mathfrak{X},\mathcal{F}) \to \left(\mathbb{R},\mathcal{B}(\mathbb{R})\right)$ eine Statistik mit $\mathbb{E}_\vartheta[|S|] < \infty$ für alle $\vartheta \in \Theta$ sowie

$$u_S\colon \Theta \to \mathbb{R}, \qquad \vartheta \mapsto e^{b(\vartheta)}\,\mathbb{E}_\vartheta[S] = \int_{\mathfrak{X}} S(x)e^{\vartheta T(x)}h(x)\,\mathrm{d}\mu_0(x). \tag{15}$$

Für $\vartheta \in \Theta$ und $t \in \mathbb{R}$ mit $\vartheta \pm t \in \Theta$ (existiert da Θ ein offenes Intervall ist) gilt aufgrund des Satzes von Beppo-Levi (BL)

$$\begin{aligned}\sum_{k=0}^{\infty}\frac{|t|^k}{k!}\int_{\mathfrak{X}}|S(x)||T(x)|^k e^{\vartheta T(x)}h(x)\,\mathrm{d}\mu_0(x) &\overset{\text{BL}}{=} \int_{\mathfrak{X}}|S(x)|e^{|tT(x)|}e^{\vartheta T(x)}h(x)\,\mathrm{d}\mu_0(x)\\ &\overset{(\star)}{\leqslant} \int_{\mathfrak{X}}|S(x)|\left(e^{tT(x)}+e^{-tT(x)}\right)e^{\vartheta T(x)}h(x)\,\mathrm{d}\mu_0(x)\\ &= \mathbb{E}_{\vartheta+t}[|S|] + \mathbb{E}_{\vartheta-t}[|S|] < \infty,\end{aligned}$$

wobei $(\star)$ die Ungleichung $e^{|x|} \leqslant e^x + e^{-x}$ für alle $x \in \mathbb{R}$ ist[4]. Also ist $\mathbb{E}_\vartheta[|S|\cdot|T|^k] < \infty$ für alle $k \in \mathbb{N}$ und alle $\vartheta \in \Theta$. Insbesondere $\mathbb{E}_\vartheta[T^2] < \infty$ folgt für $S \equiv 1$ und $k = 2$.

[4] Da beide Funktionen gerade sind, können wir ohne Beschränkung der Allgemeinheit $x > 0$ annehmen (für $x = 0$ ist die Ungleichung offensichtlich strikt). Für diese x ist die Ungleichung einfach $e^x \leqslant e^x + e^{-x} \iff 0 \leqslant e^{-x}$.

2 und 3. Die Reihe
$$\sum_{k=0}^{\infty} \frac{t^k}{k!} \int_{\mathfrak{X}} S(x)T(x)^k e^{\vartheta T(x)} h(x)\,\mathrm{d}\mu_0(x)$$
ist absolut konvergent für alle $t \in \mathbb{R}$ mit $\vartheta \pm t \in \Theta$. Die Summe und das Integral sind vertauschbar nach dem Satz von FUBINI, das heißt die Reihe hat den Wert
$$\int_{\mathfrak{X}} S(x)e^{(\vartheta+t)T(x)} h(x)\,\mathrm{d}\mu_0(x) = u_S(\vartheta+t).$$
Somit ist u_S in einer entsprechenden Umgebung analytisch und es gilt
$$\begin{aligned} u_S'(\vartheta) &= \lim_{t\to 0} \frac{1}{t}\left(u_S(\vartheta+t) - u_S(\vartheta)\right) \\ &= \lim_{t\to 0} \int_{\mathfrak{X}} S(x)\left(\frac{e^{(\vartheta+t)T(x)} - e^{\vartheta T(x)}}{t}\right) h(x)\,\mathrm{d}\mu_0(x) \\ &= \int_{\mathfrak{X}} S(x)\left(\lim_{t\to 0}\frac{e^{(\vartheta+t)T(x)} - e^{\vartheta T(x)}}{t}\right) h(x)\,\mathrm{d}\mu_0(x) \\ &= \int_{\mathfrak{X}} S(x)T(x)e^{\vartheta T(x)} h(x)\,\mathrm{d}\mu_0(x) = e^{b(\vartheta)}\,\mathbb{E}_\vartheta[S\cdot T]. \end{aligned} \tag{16}$$
Für $S \equiv 1$ folgt
$$u_1'(\vartheta) = e^{b(\vartheta)}\,\mathbb{E}_\vartheta[T] \qquad \text{und} \qquad u_1''(\vartheta) = e^{b(\vartheta)}\,\mathbb{E}_\vartheta[T^2]. \tag{17}$$
Wegen $b(\vartheta) = \ln\big(u_1(\vartheta)\big)$ folgt
$$b'(\vartheta) = \frac{u_1'(\vartheta)}{u_1(\vartheta)} \overset{(15)}{=} \mathbb{E}_\vartheta[T] = \tau(\vartheta) \tag{18}$$
für alle $\vartheta \in \Theta$ und somit
$$\begin{aligned} \tau'(\vartheta) = b''(\vartheta) &= \left(\frac{u_1'(\vartheta)}{u_1(\vartheta)}\right)' = \frac{u_1(\vartheta)u_1''(\vartheta) - u_1'(\vartheta)^2}{u_1(\vartheta)^2} \\ &\overset{(17)}{=} \frac{e^{2b(\vartheta)}\,\mathbb{E}_\vartheta[T^2] - e^{2b(\vartheta)}\,\mathbb{E}_\vartheta[T]^2}{e^{2b(\vartheta)}} = \mathbb{E}_\vartheta[T^2] - \mathbb{E}_\vartheta[T]^2 = \mathbb{V}_\vartheta[T]. \end{aligned} \tag{19}$$
Mit $a'(\vartheta) = 1$ folgen die Aussagen.

4 Für alle $\vartheta \in \Theta$ gilt
$$U_\vartheta(x) = \frac{\partial}{\partial\vartheta}\ln\big(\rho(x,\vartheta)\big) \overset{(14)}{=} T(x) - b'(\vartheta), \tag{20}$$
und somit gilt
$$\mathbb{E}_\vartheta[U_\vartheta] = \mathbb{E}_\vartheta[T] - b'(\vartheta) \overset{(18)}{=} \tau(\vartheta) - \tau(\vartheta) = 0$$
und
$$I(\vartheta) = \mathbb{V}_\vartheta[U_\vartheta] \overset{(20)}{=} \mathbb{V}_\vartheta[T - b'(\vartheta)] = \mathbb{V}_\vartheta[T] \overset{(19)}{=} \tau'(\vartheta).$$

5 Es gilt
$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}\vartheta}\,\mathbb{E}_\vartheta[S] &\overset{(15)}{=} \frac{\mathrm{d}}{\mathrm{d}\vartheta} u_S(\vartheta)e^{-b(\vartheta)} = \left[u_S'(\vartheta) - b'(\vartheta)u_S(\vartheta)\right]e^{-b(\vartheta)} \\ &\overset{(16)}{=} \mathbb{E}_\vartheta[S\cdot T] - b'(\vartheta)\,\mathbb{E}_\vartheta[S] \overset{(18)}{=} \mathbb{E}_\vartheta[ST] - \mathbb{E}_\vartheta[S]\,\mathbb{E}_\vartheta[T]. \end{aligned} \tag{21}$$

Also folgt

$$\begin{aligned}\int_{\mathfrak{X}} S(x)\frac{\partial}{\partial\vartheta}\rho(x,\vartheta)\,\mathrm{d}\mu_0(x) &= \mathbb{E}_\vartheta[S\cdot U_\vartheta] \overset{(20)}{=} \mathbb{E}_\vartheta\left[S\cdot\big(T-\tau(\vartheta)\big)\right]\\ &= \mathbb{E}_\vartheta[ST]-\mathbb{E}_\vartheta[S]\underbrace{\mathbb{E}_\vartheta[T]}_{=\tau(\vartheta)}\\ &\overset{(21)}{=} \frac{\mathrm{d}}{\mathrm{d}\vartheta}\mathbb{E}_\vartheta[S] = \frac{\mathrm{d}}{\mathrm{d}\vartheta}\int_{\mathfrak{X}} S(x)\rho(x,\vartheta)\,\mathrm{d}\mu_0(x).\end{aligned}$$

□

Das Produkt regulärer Modelle ist regulär.

Lemma 1.3.17 (Produkt exponentieller Modelle ist exponentiell)
Ist $\big(\mathfrak{X},\mathcal{F},(\mathbb{P}_\vartheta)_{\vartheta\in\Theta}\big)$ ein exponentielles Modell bezüglich T und $n\in\mathbb{N}$, so ist das n-fache Produktmodell exponentiell bezüglich $T_n(x) := \frac{1}{n}\sum_{k=1}^n T(x_k)$. Insbesondere ist T_n der beste Schätzer für $\tau := \mathbb{E}_\cdot[T]$ (nach Satz 1.3.5).

Beweis. Ist $\rho(x,\vartheta) = h(x)e^{a(\vartheta)T(x)-b(\vartheta)}$ die Likelihoodfunktion von $\big(\mathfrak{X},\mathcal{F},(\mathbb{P}_\vartheta)_{\vartheta\in\Theta}\big)$, so ist die Likelihoodfunktion des Produktmodells

$$\begin{aligned}(x,\vartheta)\mapsto\prod_{k=1}^n\rho(x_k,\vartheta) &= \left(\prod_{k=1}^n h(x_k)\right)\exp\left(a(\vartheta)\cdot\sum_{k=1}^n T(x_k)-nb(\vartheta)\right)\\ &= \underbrace{\left(\prod_{k=1}^n h(x_k)\right)}_{=:h_n(x)}\exp\Big(\underbrace{n\cdot a(\vartheta)}_{=:a_n(\vartheta)}\cdot T_n(x)-\underbrace{nb(\vartheta)}_{=:b_n(\vartheta)}\Big)\end{aligned}$$

□

Bemerkung 1.3.18 (Fisher-Information des Produktmodells)
Ist I die Fisher-Information des Modells, so ist die Fisher-Information des Produktmodells

$$I_n(\vartheta) \overset{1.3.5}{=} a_n'(\vartheta)\tau'(\vartheta) = na'(\vartheta)\tau'(\vartheta) = nI(\vartheta).$$

Die Fisher-Information ist also exakt additiv - mehr Versuche geben uns mehr Information. ○

Bemerkung 1.3.19 Die Formel $I_n = nI$ gilt allgemein für reguläre Produktmodelle (cf. [2, Bem. 7.17]). Unter den Voraussetzungen von Satz 1.3.3 gilt für jedes $n\in\mathbb{N}$ und jeden regulären erwartungstreuen Schätzer T_n für τ des Produktmodells

$$\mathbb{V}_\vartheta[T_n] \geqslant \frac{\tau'(\vartheta)^2}{I_n(\vartheta)} = \frac{\tau'(\vartheta)^2}{nI(\vartheta)} \in O(n^{-1}) \text{ für } n\to\infty \qquad \forall\vartheta\in\Theta.$$

○

Bemerkung 1.3.20 Im Gameshow-Modell sahen wir, dass $T_n^*(x) := \frac{n+1}{n}x_{(n)}$ ein erwartungstreuer Schätzer für ϑ ist und $\mathbb{V}_\vartheta[T_n^*] = \frac{\vartheta^2}{n(n+2)}$. Dies ist *kein* Widerspruch, da dieses Modell nicht regulär ist, da die Likelihoodfunktion für $n=1$

$$\rho(x,\vartheta) = \frac{1}{\vartheta}\mathbb{1}_{[0,\vartheta]}(x)$$

und somit nicht stetig (und schon gar nicht differenzierbar) ist. ○

Beispiel 1.3.21 (Binomialmodell ist exponentiell)
Für $n \in \mathbb{N}$ bilden die Binomialverteilung $(B_{n,\vartheta})_{\vartheta\in\Theta:=(0,1)}$ eine exponentielle Familie auf $\big(\mathfrak{X} := \{0,\dots,n\}, \mathcal{F} := \mathcal{P}(\mathfrak{X})\big)$ mit dem Zählmaß μ_0 und $T(x) = \frac{1}{n}x$: für $x \in \mathfrak{X}$ und $\vartheta \in \Theta$ gilt

$$\begin{aligned}\rho(x,\vartheta) &= \binom{n}{x}\vartheta^x(1-\vartheta)^{n-x} = \binom{n}{x}\exp\left(x\ln(\vartheta) + (n-x)\ln(1-\vartheta)\right)\\ &= \underbrace{\binom{n}{x}}_{=:h(x)>0}\exp\left(\underbrace{\frac{1}{n}x}_{=T(x)}\cdot\underbrace{n\ln\left(\frac{\vartheta}{1-\vartheta}\right)}_{=:a(\vartheta)} + \underbrace{n\ln(1-\vartheta)}_{=:b(\vartheta)}\right).\end{aligned}$$

Nach Satz 1.3.5 ist T bester Schätzer für $\tau(\vartheta) := \mathbb{E}_\vartheta[T] = \vartheta$. ◇

Beispiel 1.3.22 (Gauß-I-Modell ist exponentiell)
Bei gegebener Varianz $\sigma^2 > 0$ hat die Familie $(\mathcal{N}_{\vartheta,\sigma^2})_{\vartheta\in\mathbb{R}}$ die Likelihoodfunktion

$$\begin{aligned}\rho(x,\vartheta) &= \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{1}{2\sigma^2}(x-\vartheta)^2\right)\\ &= \underbrace{\exp\left(-\frac{1}{2\sigma^2}x^2\right)}_{=:h(x)}\exp\left(\underbrace{x}_{=T(x)}\underbrace{\frac{\vartheta}{\sigma^2}}_{=:a(\vartheta)}\underbrace{-\frac{1}{2\sigma^2}\vartheta^2 - \frac{1}{2}\ln(2\pi\sigma^2)}_{=:-b(\vartheta)}\right).\end{aligned}$$

Nach Lemma 1.3.17 ist $T(x) := \frac{1}{n}\sum_{k=1}^n x_k$ der beste Schätzer für $\tau = \mathrm{id}_{(0,\infty)}$ im n-fachen Produktmodell. ◇

Beispiel 1.3.23 (Gauß-III-Modell ist exponentiell)
Bei gegebenen Mittelwert $m \in \mathbb{R}$ hat die Familie $(\mathcal{N}_{m,\vartheta})_{\vartheta\in(0,\infty)}$ auf $\big(\mathbb{R}, \mathcal{B}(\mathbb{R})\big)$ die Likelihoodfunktion

$$\begin{aligned}\rho(x,\vartheta) &= \frac{1}{\sqrt{2\pi\vartheta}}\exp\left(-\frac{1}{2\vartheta}(x-m)^2\right)\\ &= \exp\left(-\frac{1}{2\vartheta}(x-m)^2 - \frac{1}{2}\ln(2\pi\vartheta)\right),\end{aligned}$$

welche die Form (14) mit $T(x) := (x-m)^2$, $a(\vartheta) := -\frac{1}{2\vartheta}$, $b(\vartheta) := \frac{1}{2}\ln(2\pi\vartheta)$ und $h(x) = 1$. Nach Satz 1.3.5 ist T ein bester Schätzer für $\tau(\vartheta) := \frac{b'(\vartheta)}{a'(\vartheta)} = \vartheta$ mit Varianz $\mathbb{V}_\vartheta[T] = \frac{\tau'(\vartheta)}{a'(\vartheta)} = 2\vartheta^2$. Nach Lemma 1.3.17 ist $T(x) := \frac{1}{n}\sum_{k=1}^n (x_k - m)^2$ der beste Schätzer für $\tau = \mathrm{id}_{(0,\infty)}$ im n-fachen Produktmodell. ◇

(Das Gauss-Modell mit unbekanntem Mittelwert und unbekannter Varianz ist exponentiell im erweitertem Sinne.)

Beispiel 1.3.24 (Poisson-Modell ist exponentiell)
Das Poisson-Modell ist ein exponentielles Modell bezüglich $T(x) = x$, da

$$\rho(x,\vartheta) = e^{-\vartheta}\frac{\vartheta^x}{x!} = \frac{1}{x!}\exp\left(x\ln(\vartheta) - \vartheta\right)$$

die Form (14) mit $a(\vartheta) = \ln(\vartheta)$, $b(\vartheta) = \vartheta$ und $h(x) = \frac{1}{x!}$ hat. Also ist T ein bester Schätzer für $\tau(\vartheta) = \frac{b'(\vartheta)}{a'(\vartheta)} = \vartheta$ mit Varianz $\mathbb{V}_\vartheta[T] = \frac{\tau'(\vartheta)}{a'(\vartheta)} = \vartheta$.

Nach Lemma 1.3.17 ist $T(x) := \frac{1}{n}\sum_{k=1}^n x_k$ der beste Schätzer für $\tau = \mathrm{id}_{(0,\infty)}$ im n-fachen Produktmodell. ◇

1.4 Suffizienz und Vollständigkeit

Motivation. Was sind die für τ relevanten Information, die wir aus einer Beobachtung erfahren können? 10.05.2022

Bei n unabhängigen Münzwürfen (n-faches BERNOULLI-Produktmodell) ist $\mathfrak{X} = \{0,1\}^n$ und $\Theta := (0,1)$. Es scheint so zu sein, dass $x = (x_1, \dots, x_n) \in \mathfrak{X}$ den selben „Informationsgehalt" über ϑ enthält, wie $\sum_{k=1}^n x_k$, obwohl weniger Informationen über x bekannt sind.

Wie können wir diesen Verhalt mathematisch präzisieren?

Beim Gameshow-Modell „sollte" $\tilde{T}_n(x) := x_{(n)}$ suffizient sein, obwohl $\tilde{T}_n$ (oder $a\tilde{T}_n$ für irgendein $a \in \mathbb{R}$) kein bester Schätzer ist. ○

DEFINITION 1.4.1 (SUFFIZIENZ)
Sei $\big(\mathfrak{X}, \mathcal{F}, (\mathbb{P}_\vartheta)_{\vartheta\in\Theta}\big)$ ein statistisches Modell.

- Die σ-Algebra $\mathcal{B} \subset \mathcal{F}$ ist suffizient, wenn für alle $A \in \mathcal{F}$ eine messbare Funktion $f_A\colon (\mathfrak{X}, \mathcal{B}) \to \big(\mathbb{R}, \mathcal{B}(\mathbb{R})\big)$ existiert, sodass für alle $\vartheta \in \Theta$ suffizient

$$f_A(x) = \mathbb{P}_\vartheta[A \mid \mathcal{B}](x) \qquad \text{für } \mathbb{P}_\vartheta\text{-fast alle } x \in \mathfrak{X}\,.$$

- Die Statistik $T\colon (\mathfrak{X}, \mathcal{F}) \to (\Sigma, \mathscr{S})$ ist suffizient, wenn $\sigma(T) := \{T^{-1}(A) : A \in \mathscr{S}\} \subset \mathcal{F}$ suffizient ist, das heißt, wenn für alle $A \in \mathcal{F}$ $\mathbb{P}_\vartheta[A \mid \sigma(T)]$ nicht von ϑ abhängt.

Bemerkung 1.4.2 (Interpretation der Suffizienz) Beim Übergang von X zu $T(X)$ gehen typischerweise Informationen verloren. Die Statistik T ist *suffizient*, wenn T alle relevanten Informationen bezüglich des unbekannten Parameters enthält. ○

Bemerkung 1.4.3 Ist $\mathcal{B} \subset \mathcal{F}$ suffizient, so ist $T = \mathrm{id}$ suffizient. Ist T suffizient, so ist $\sigma(T)$ suffizient, es gibt also eine Art 1-1-Beziehung. ○

Wir betrachten zunächst die beiden Extremfälle.

Beispiel 1.4.4 ($T = \mathrm{id}$)
Sei $T\colon (\mathfrak{X}, \mathcal{F}) \to (\mathfrak{X}, \mathcal{F})$ die Identität („wir werfen keine Informationen weg"). Dann ist $\sigma(T) = \mathcal{F}$ suffizient, denn für $A \in \mathcal{B} := \mathcal{F}$ und die $\mathcal{F}$-messbare Abbildung $f_A := \mathbb{1}_A$ gilt für $\mathbb{P}_\vartheta$-fast alle $x \in \mathfrak{X}$

$$\mathbb{1}_A(x) = \mathbb{P}_\vartheta[A \mid \mathcal{B}](x) \qquad \forall \vartheta \in \Theta.$$ ◇

Beispiel 1.4.5 (T konstant)
Sei $T\colon (\mathfrak{X}, \mathcal{F}) \to (\Sigma, \mathscr{S})$ konstant, das heißt, T nimmt nur einen Wert an. Dann ist $\sigma(T) = \{\varnothing, \mathfrak{X}\}$ und für $\mathbb{P}_\vartheta$-fast alle $x \in \mathfrak{X}$ gilt

$$\mathbb{P}_\vartheta[A \mid \{\varnothing, \mathfrak{X}\}](x) = \mathbb{P}_\vartheta[A].$$

Wenn ein f_A wie in Definition 1.4.1 existiert, dann gilt $\mathbb{P}_\vartheta(A) = f_A(x)$ $\mathbb{P}_\vartheta$-fast sicher für alle $\vartheta \in \Theta$. Somit ist $\vartheta \mapsto \mathbb{P}_\vartheta[A]$ für alle A konstant, also sind alle $\mathbb{P}_\vartheta$, $\vartheta \in \Theta$ gleich. Also ist T, abgesehen von diesem uninteressanten Spezialfall, nicht suffizient. ◇

Lemma 1.4.6 (Charakterisierung der Suffizienz)
Es ist $\mathcal{B} \subset \mathcal{F}$ genau dann suffizient, wenn für alle messbaren Funktionen $g\colon (\mathfrak{X}, \mathcal{F}) \to \big(\mathbb{R}, \mathcal{B}(\mathbb{R})\big)$ mit $g \in L^1(\mathbb{P}_\vartheta)$ für alle $\vartheta \in \Theta$ eine messbare Abbildung $f_g\colon (\mathfrak{X}, \mathcal{B}) \to \big(\mathbb{R}, \mathcal{B}(\mathbb{R})\big)$ existiert, sodass für alle $\vartheta \in \Theta$ gilt

$$f_g(x) = \mathbb{E}_\vartheta[g(x) \mid \mathcal{B}] \qquad \text{für } \mathbb{P}_\vartheta\text{-fast alle } x \in \mathfrak{X}\,. \tag{22}$$

Beweis. " $\Longleftarrow$ ": Setze $g := \mathbb{1}_A$ für $A \in \mathcal{F}$.

" $\Longrightarrow$ ": (*Maßtheoretische Induktion*) Die Aussage (22) gilt für Funktionen $g := 1_A$ mit $A \in \mathcal{F}$ mit $f_g(x) := f_A(x)$ (letztere ist aus Definition 1.4.1, erster aus (22)).

Ist $g = \sum_{k=1}^n \alpha_k \mathbb{1}_{A_k}$ mit $\alpha_k \geqslant 0$ und $A_k \in \mathcal{F}$ für alle $k \in \{1, \dots, n\}$, dann folgt (22) mit $f_g(x) := \sum_{k=1}^n \alpha_k f_{A_k}(x)$.

Für $g \in L^1$ mit $g \geqslant 0$ benutzt man des Satz über monotone Konvergenz und für allgemeine $g \in L^1$ die Zerlegung in Positiv- und Negativteil. □

Ist T eine diskrete Zufallsvariable, dessen Suffizienz wir zeigen wollen, genügt es, zu zeigen, dass $\mathbb{P}_\vartheta[X \in A \mid T(x) = m]$ unabhängig von ϑ ist.

Beispiel 1.4.7 (Bernoulli-Produktmodell)
Wir zeigen, dass in diesem Modell $S(x_1, \dots, x_n) := \sum_{k=1}^n x_k$ suffizient ist.

Für $B \subset \{0,1\}^n = \mathfrak{X}$ und $x \in \mathfrak{X}$ gilt

$$\mathbb{P}_\vartheta[B \mid \sigma(S)](x) = \mathbb{P}_\vartheta[B \mid S = m],$$

wobei $m := \sum_{k=1}^n x_k$.

Für $m \in \{1, \dots, n\}$ gilt

$$\begin{aligned}
\mathbb{P}_\vartheta[B \mid S = m] &= \frac{\mathbb{P}_\vartheta[B \cap \{S = m\}]}{\mathbb{P}_\vartheta[S = m]} = \frac{1}{\mathbb{P}_\vartheta[S = m]} \sum_{b \in B} \mathbb{P}_\vartheta[\underbrace{\{b\} \cap \{S = m\}}_{=0 \text{ wenn } \sum_{k=1}^n b_k \neq m}] \\
&= \frac{1}{\mathbb{P}_\vartheta[S = m]} \sum_{\substack{b \in B \\ S(b) = m}} \mathbb{P}_\vartheta\big[\{b\}\big] = \frac{1}{\mathbb{P}_\vartheta[S = m]} \sum_{\substack{b \in B \\ S(b) = m}} \vartheta^m (1 - \vartheta)^{n-m} \\
&= \frac{\sum\limits_{\substack{b \in B \\ S(b) = m}} \vartheta^m (1 - \vartheta)^{n-m}}{\binom{n}{m} \vartheta^m (1 - \vartheta)^{n-m}} = \frac{1}{\binom{n}{m}} \#\{b \in B : S(b) = m\},
\end{aligned}$$

also ist $\mathbb{P}_\vartheta[B \mid S = m]$ unabhängig von ϑ.

Jedoch ist für $n \geqslant 2$ der Schätzer $\tilde{S}(x_1, \dots, x_n) := \sum_{k=1}^{n-1} x_k$ nicht suffizient: für $B \subset \mathcal{P}(\mathfrak{X})$

und $m \in \{0, \dots, n-1\}$ ist nach der obigen Rechnung $\mathbb{P}_\vartheta[B \mid \tilde{S} = m]$ gleich

$$\begin{aligned}
&\frac{1}{\mathbb{P}_\vartheta[\tilde{S} = m]} \Bigg(\sum_{\substack{b \in B \\ \tilde{S}(b)=m}} \vartheta^{m+1}(1-\vartheta)^{n-1-m} + \sum_{\substack{b \in B \\ \tilde{S}(b)=m}} \vartheta^{m}(1-\vartheta)^{n-m} \Bigg) \\
&= \frac{1}{\binom{n-1}{m}\vartheta^m(1-\vartheta)^{n-1-m}} \Bigg(\sum_{\substack{b \in B \\ \tilde{S}(b)=m}} \vartheta^{m+1}(1-\vartheta)^{n-1-m} + \sum_{\substack{b \in B \\ \tilde{S}(b)=m}} \vartheta^{m}(1-\vartheta)^{n-m} \Bigg) \\
&= \frac{1}{\binom{n-1}{m}} \Big(\vartheta \cdot \#\{b \in B : \tilde{S}(b) = m,\ b_n = 1\} + (1-\vartheta) \cdot \#\{b \in B : \tilde{S}(b) = m,\ b_n = 0\} \Big).
\end{aligned}$$

Die auftretenden Kardinalitäten sind verschieden, z.B. für $B = \{b \in \mathfrak{X} : b_n = 1\}$. ◇

Bemerkung 1.4.8 (Anschauung für Suffizienz (cf. Renesse-Skript, Satz 2.18))
Durch Bedingen / Festlegen auf das Ergebnis eine Hilfsbeobachtung $S = S(X)$ verbleibt zwar noch ein gewisser „Restzufall“ in der Beobachtung X, allerdings hängt dessen Verteilung nicht mehr von ϑ ab.

Ist S suffizient, so ist $\mathbb{P}_\vartheta$ durch $\mathbb{P}_\vartheta \circ S$ festgelegt. ○

Beispiel 1.4.9 (Geometrische Verteilung)
Wir werfen eine Münze und zählen die Anzahl der Misserfolge bis zum ersten Erfolg. Dies wiederholen wir n mal.

Sei $\big(\mathbb{N}_0^n, \mathbb{P}(\mathbb{N}_0^n), (\mathbb{P}_\vartheta)_{\vartheta \in \Theta := (0,1)}\big)$ das zugehörige statistische Modell mit $\mathbb{P}_\vartheta(\{k\}) = (1-\vartheta)^k \vartheta$. Die Statistik $S(x) := \sum_{k=1}^n x_k$ ist negativ-binomialverteilt:

$$\mathbb{P}_\vartheta[S(x) = s] = \binom{s+n-1}{n-1}(1-\vartheta)^s \vartheta^n \qquad \forall s \in \mathbb{N}_0\,.$$

Die Statistik S ist suffizient, denn für alle $k \in \mathbb{N}_0^n$ und $s \in \mathbb{N}_0$ hängt

$$\begin{aligned}
\mathbb{P}_\vartheta[X = k \mid S(X) = s] &= \frac{\mathbb{P}_\vartheta[X = k, S(X) = s]}{\mathbb{P}_\vartheta[S(X) = s]} = \mathbb{1}_{\{S(k)=s\}} \frac{\mathbb{P}_\vartheta[X = k]}{\mathbb{P}_\vartheta[S(X) = s]} \\
&= \mathbb{1}_{\{S(k)=s\}} \frac{\prod_{j=1}^n (1-\vartheta)^{k_j}\vartheta}{\binom{s+n-1}{n-1}(1-\vartheta)^s\vartheta^n} = \mathbb{1}_{\{S(k)=s\}} \frac{(1-\vartheta)^{\sum_{j=1}^n k_j}\vartheta^n}{\binom{s+n-1}{n-1}(1-\vartheta)^s\vartheta^n} \\
&= \mathbb{1}_{\{\sum_{j=1}^n k_j = s\}} \frac{(1-\vartheta)^{\sum_{j=1}^n k_j}\vartheta^n}{\binom{s+n-1}{n-1}(1-\vartheta)^s\vartheta^n} = \mathbb{1}_{\{\sum_{j=1}^n k_j = s\}} \frac{1}{\binom{s+n-1}{n-1}}
\end{aligned}$$

nicht von ϑ ab. ◇

Beispiel 1.4.10 (Suffizienz im Gauss-Produktmodell (Hausaufgabe 4.2))
Betrachte für $n > 1$ das n-fache Gausssche Produktmodell $\big(\mathfrak{X}^n, \mathcal{B}(\mathfrak{X}^n), (\mathbb{P}_\vartheta)_{\vartheta \in \Theta}\big)$, wobei $\mathfrak{X} := \mathbb{R}^d$ und $\mathbb{P}_\vartheta := \mathcal{N}_{\mu,\Sigma}$ die d-dimensionale Normalverteilung mit Erwartungswertvektor $\mu \in \mathbb{R}^d$ und Kovarianzmatrix $\Sigma \in \mathbb{R}^{d \times d}_{\geqslant 0}$, $d > 1$ und $\Theta := \mathbb{R}^d \times \mathbb{R}^{d \times d}_{\geqslant 0}$ sind. Definiere $M_n := M_n(X) := \frac{1}{n}\sum_{k=1}^n X_k$. Dann ist

$$U(X) := \left(M_n, \frac{1}{n-1} \sum_{k=1}^n (X_k - M_n)(X_k - M_n)^\mathsf{T} \right)$$

eine suffiziente Statistik für dieses Modell. ◇

Satz 1.4.1: Rao-Blackwell

Seien $\left(\mathfrak{X}, \mathcal{F}, (\mathbb{P}_\vartheta)_{\vartheta\in\Theta}\right)$ ein statistisches Modell, $\tau\colon \Theta \to \mathbb{R}$ eine Kenngröße, $S\colon (\mathfrak{X}, \mathcal{F}) \to (\Sigma, \mathscr{S})$ suffizient und $T\colon (\mathfrak{X}, \mathcal{F}) \to \left(\mathbb{R}, \mathcal{B}(\mathbb{R})\right)$ erwartungstreu für τ. Dann existiert eine messbare Abbildung $\tilde{T}\colon (\mathfrak{X}, \sigma(S)) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, sodass für alle $\vartheta \in \Theta$

$$\tilde{T}(x) = \mathbb{E}_\vartheta[T \mid \sigma(S)](x) \qquad \mathbb{P}_\vartheta - \text{fast sicher} \tag{23}$$

gilt. Dann ist $\tilde{T}$ erwartungstreu für τ und für alle $\vartheta \in \Theta$ gilt $\mathbb{V}_\vartheta[\tilde{T}] \leqslant \mathbb{V}_\vartheta[T]$.

Beweis. Existenz. Zu $g = T$ existiert nach Lemma 1.4.6 ein $f_T = \tilde{T}$ mit $\tilde{T}(x) = \mathbb{E}_\vartheta[T \mid \sigma(S)](x)$ $\mathbb{P}_\vartheta$-fast sicher für alle $\vartheta \in \Theta$.

Erwartungstreue. Für $\vartheta \in \Theta$ gilt nach der Turmeigenschaft (T)

$$\mathbb{E}_\vartheta[\tilde{T}] = \mathbb{E}_\vartheta\left[\mathbb{E}_\vartheta[T \mid \sigma(S)]\right] \overset{(T)}{=} \mathbb{E}_\vartheta[T] = \tau(\vartheta).$$

Varianz. Sei $\vartheta \in \Theta$. Ohne Beschränkung der Allgemeinheit sei $\mathbb{V}_\vartheta[T] < \infty$. Dann folgt mit der Jensen-Ungleichung für bedingte Erwartungswerte (J) (da $x \mapsto x^2$ konvex ist) und der Erwartungstreue von $\tilde{T}$

$$\begin{aligned}\mathbb{V}_\vartheta[\tilde{T}] = \mathbb{E}_\vartheta[\left(\tilde{T} - \tau(\vartheta)\right)^2] &\overset{(23)}{=} \mathbb{E}_\vartheta\left[(\mathbb{E}_\vartheta[T - \tau(\vartheta) \mid \sigma(S)])^2\right] \\ &\overset{(J)}{\leqslant} \mathbb{E}_\vartheta\left[\mathbb{E}_\vartheta\left[\left(T - \tau(\vartheta)\right)^2 \mid \sigma(S)\right]\right] \overset{(T)}{\leqslant} \mathbb{E}_\vartheta\left[\left(T - \tau(\vartheta)\right)^2\right] = \mathbb{V}_\vartheta[T]\end{aligned} \tag{24}$$

aufgrund der Erwartungstreue von T. □

Bemerkung 1.4.11 Mit diesem Satz lässt sich einen erwartungstreuen Schätzer T mittels einer suffizienten Statistik verbessern, in dem Sinne, dass die Erwartungstreue erhalten bleibt und die Varianz nicht größer wird. Ist S die Identität (suffizient nach Beispiel 1.4.4), so ist $\tilde{T} = T$ $\mathbb{P}_\vartheta$-fast überall für alle $\vartheta \in \Theta$, also ist keine Verbesserung erreicht worden. ○

Beispiel 1.4.12 (Poisson-Modell (Hausaufgabe 4.1))
Betrachte das statistische Produktmodell $\left(\mathbb{N}_0^n, \mathcal{P}(\mathbb{N}_0^n), (\mathrm{Poi}(\vartheta)^{\otimes n})_{\vartheta>0}\right)$. Dann ist $T(X) = X_1 \cdot X_2$ ein erwartungstreuer Schätzer für ϑ^2. Die Statistik $U(X) := \sum_{k=1}^n X_k$ ist in diesem Modell suffizient und vollständig. Mit Satz 1.4.1 kann man U zu einem Schätzer S verbessern, welcher ein bester erwartungstreuer Schätzer für ϑ^2 ist. ◇

11.05.2022

Satz 1.4.2: Neyman-Fisher Faktorisierungslemma (diskret)

Seien $\left(\mathfrak{X}, \mathcal{F}, (\mathbb{P}_\vartheta)_{\vartheta\in\Theta}\right)$ ein Standardmodell, wobei das dominierende Maß μ_0 das Zählmaß ist und $\rho\colon \Theta \times \mathfrak{X} \to (0, \infty)$ die Likelihoodfunktion.
Die Statistik $S\colon\ \mathfrak{X} \to \Sigma$ ist suffizient genau dann wenn Funktionen $h\colon \Theta \times \Sigma \to \mathbb{R}$ und $k\colon\ \mathfrak{X} \to \mathbb{R}$ existieren, sodass

$$\rho(\vartheta, x) = h(\vartheta, S(x))k(x)$$

für alle $x \in \mathfrak{X}$ und alle $\vartheta \in \Theta$ gilt.

Beweis. Für $s \in \Sigma$ mit $\mathbb{P}[S(X) = s] > 0$ und $x \in \mathfrak{X}$ gilt

$$\begin{aligned}\mathbb{P}_\vartheta[X = x \mid S(X) = s] &= \frac{\mathbb{P}_\vartheta[X = x, S(X) = s]}{\mathbb{P}_\vartheta[S(X) = s]} \\ &= \mathbb{1}_{\{S(X)=s\}} \frac{\mathbb{P}_\vartheta[X = x]}{\mathbb{P}_\vartheta[S(X) = s]} = \mathbb{1}_{\{S(X)=s\}} \frac{\rho(\vartheta, x)}{\sum_{y \in S^{-1}(s)} \rho(\vartheta, y)}.\end{aligned} \tag{25}$$

" $\Longrightarrow$ ": Sei S suffizient. Dann existiert eine Funktion $f\colon \mathfrak{X} \times \Sigma \to \mathbb{R}$ mit

$$f(x, s) = \mathbb{P}_\vartheta[X = x \mid S(X) = s] \qquad \forall (s, x) \in \mathfrak{X} \times \Sigma.$$

Für $x \in \mathfrak{X}$ und $s = S(X)$ gilt dann

$$\rho(\vartheta, x) \overset{(25)}{=} \underbrace{f(x, S(x))}_{=:k(x)} \underbrace{\sum_{y \in S^{-1}(s)} \rho(\vartheta, y)}_{=:h(\vartheta, s)}.$$

" $\Longleftarrow$ ": Sei $\rho(\vartheta, x) = h(\vartheta, S(x))k(x)$. Dann gilt für $x \in \mathfrak{X}$ und $s \in \Sigma$

$$\begin{aligned}\mathbb{P}_\vartheta[X = x \mid S(X) = s] &\overset{(25)}{=} \mathbb{1}_{\{S(X)=s\}} \frac{h(\vartheta, S(x))k(x)}{\sum_{y \in S^{-1}(s)} h(\vartheta, S(y))k(y)} \\ &= \mathbb{1}_{\{S(X)=s\}} \frac{h(\vartheta, s)k(x)}{\sum_{y \in S^{-1}(s)} h(\vartheta, S(y))k(y)} \\ &= \mathbb{1}_{\{S(X)=s\}} \frac{k(x)}{\sum_{y \in S^{-1}(s)} k(y)}.\end{aligned}$$

□

Beispiel 1.4.13 (Geometrische Verteilung)
Die Dichte bezüglich des Zählmaßes ist

$$\rho(x, \vartheta) = \prod_{k=1}^{n} (1 - \vartheta)^{x_k} \vartheta = (1 - \vartheta)^{S(x)} \vartheta^n =: h(\vartheta, S(x)),$$

somit ist S suffizient nach Satz 1.4.2. ◇

Beispiel 1.4.14 (Bernoulli-Produktmodell) **12.05.2022**
Sei wie in Beispiel 1.4.7 $S(x_1, \ldots, x_n) := \sum_{k=1}^n x_k$. Dann ist S suffizient. Sei $\tau(\vartheta) := \vartheta \in \Theta := (0, 1)$ und $T(x_1, \ldots, x_n) := x_1$. Dann ist T erwartungstreu.

Sei

$$\tilde{T}(x_1, \ldots, x_n) := \mathbb{E}_\vartheta[T \mid \sigma(S)](x_1, \ldots, x_n) = \mathbb{E}_\vartheta[T(x) \mid S(x) = m],$$

wobei $m = \sum_{k=1}^n x_k$.

Dann gilt

$$\begin{aligned}\frac{\mathbb{E}_\vartheta[T \cdot \mathbb{1}_{\{S(x)=m\}}]}{\mathbb{P}_\vartheta[S = m]} &= \frac{1}{\mathbb{P}_\vartheta[S = m]} \mathbb{E}_\vartheta[X_1 \cdot \mathbb{1}_{\{S=m\}}] = \frac{1}{\mathbb{P}_\vartheta[S = m]} \mathbb{P}_\vartheta[X_1 = 1, S = m] \\ &= \frac{1}{\mathbb{P}_\vartheta[S = m]} \mathbb{P}_\vartheta[X_1 = 1] \mathbb{P}_\vartheta[S = m \mid X_? = 1] \\ &= \frac{1}{\mathbb{P}_\vartheta[S = m]} \vartheta \vartheta^{m-1} (1 - \vartheta)^{n-1-(m-1)} \binom{n-1}{m-1} \\ &= \frac{\vartheta^m \cancel{(1 - \vartheta)^{n-m}} \binom{n-1}{m}}{\binom{n}{m} \cancel{\vartheta^m (1 - \vartheta)^{n-m}}} = \frac{m}{n}\end{aligned}$$

und diese Formel gilt auch für $m = 0$. Also ist $\tilde{T}(x) = \frac{m}{n}$ für alle $m \in \{0, \ldots, n\}$. Nach Satz 1.3.5 und Lemma 1.3.17 ist das der beste Schätzer. $\diamond$

Satz 1.4.3: Neyman-Kriterium für Suffizienz

Sei $\big(\mathfrak{X}, \mathcal{F}, (\mathbb{P}_\vartheta)_{\vartheta\in\Theta}\big)$ ein Standardmodell mit dominierendem Maß μ_0 und Likelihoodfunktion ρ. Eine Unter-σ-Algebra $\mathcal{B} \subset \mathcal{F}$ ist genau dann suffizient, wenn eine messbare Funktion $h\colon (\mathfrak{X}, \mathcal{F}) \to [0, \infty)$ und für jedes $\vartheta \in \Theta$ eine messbare Funktion $f_\vartheta\colon (\mathfrak{X}, \mathcal{B}) \to [0, \infty)$ existiert, sodass

$$\rho(x, \vartheta) = f_\vartheta(x) h(x)$$

μ_0-fast sicher und für alle $\vartheta \in \Theta$ gilt.

Beweis. Siehe [3]. $\square$

Bemerkung 1.4.15 (Schätzer im exponentiellen Modell ist suffizient) In einem exponentiellen Modell gilt

$$\rho(x, \vartheta) = h(x) \exp\left(a(\vartheta)T(x) - b(\vartheta)\right).$$

Es ist h positiv und wir können $f_\vartheta(x) := \exp\left(a(\vartheta)T(x) - b(\vartheta)\right)$ setzen. Dann ist $\mathcal{B} := \sigma(T)$ nach Satz 1.4.3 suffizient, also auch T. $\circ$

Beispiel 1.4.16 (Gameshow)
Der Schätzer $\tilde{T}_n := \max(\{x_1, \ldots, x_n\})$ ist suffizient, denn die Likelihoodfunktion ist

$$\rho(x, \vartheta) = \vartheta^{-n}\, \mathbb{1}_{[0,\vartheta]}(\tilde{T}_n(x)) =: f_\vartheta(x)$$

ist $\mathcal{B} := \sigma(\tilde{T}_n)$ messbar für alle $\vartheta \in \Theta$. $\diamond$

Bemerkung 1.4.17 In der Situation von Satz 1.4.3 folgt: ist $\mathcal{B}$ suffizient und ist $\tilde{\mathcal{B}} \subset \mathcal{F}$ ein σ-Algebra mit $\mathcal{B} \subset \tilde{\mathcal{B}}$, so ist $\tilde{\mathcal{B}}$ ebenfalls suffizient (aus der $\mathcal{B}$-Messbarkeit von f_ϑ folgt auch die $\tilde{\mathcal{B}}$-Messbarkeit von f_ϑ). Allgemein ist das tatsächlich falsch (cf. [3]). $\circ$

Definition 1.4.18 (Vollständige Statistik)
Sei $\big(\mathfrak{X}, \mathcal{F}, (\mathbb{P}_\vartheta)_{\vartheta\in\Theta}\big)$ ein statistisches Modell. Die Statistik $S\colon (\mathfrak{X}, \mathcal{F}) \to (\Sigma, \mathscr{S})$ ist vollständig, wenn für alle messbaren Funktionen $h\colon (\Sigma, \mathscr{S}) \to \big(\mathbb{R}, \mathcal{B}(\mathbb{R})\big)$ mit $\mathbb{E}_\vartheta[|h(S)|] < \infty$ gilt vollständig

$$\mathbb{E}_\vartheta[h(S)] = 0 \quad \forall \vartheta \in \Theta \qquad \Longrightarrow \qquad h(S) = 0 \quad \mathbb{P}_\vartheta\text{-fast sicher } \forall \vartheta \in \Theta.$$

Lemma 1.4.19 (Charakterisierung von Vollständigkeit)
Die Statistik S ist genau dann vollständig, wenn aus $f, g \in L^1(\mathfrak{X}, \sigma(S), \mathbb{P}_\vartheta)$ und $\int_{\mathfrak{X}} f(x)\, \mathrm{d}\mathbb{P}_\vartheta(x) = \int_{\mathfrak{X}} g(x)\, \mathrm{d}\mathbb{P}_\vartheta(x)$ für alle $\vartheta \in \Theta$ folgt, dass $\mathbb{P}_\vartheta[f = g] = 1$ für alle $\vartheta \in \Theta$ gilt.

Bemerkung 1.4.20 (Von Wikipedia) Die (äquivalente) Kontraposition der obigen Charakterisierung von Vollständigkeit ist: gilt nicht $f = g$ $\mathbb{P}_{\vartheta_0}$-fast sicher für ein $\vartheta_0 \in \Theta$, so folgt $\int_{\mathfrak{X}} f(x) - g(x)\, \mathrm{d}\mathbb{P}_{\vartheta_0}(x) \neq 0$, also ist $\sigma(S)$ klein genug, damit $(\mathbb{P}_\vartheta)_{\vartheta\in\Theta}$ alle Funktionen aus $L^1\big(\mathfrak{X}, \sigma(S), (\mathbb{P}_\vartheta)_{\vartheta\in\Theta}\big)$ unterscheiden kann. $\circ$

Beweis. " $\Longrightarrow$ ": Seien S vollständig und $f, g \in L^1(\mathfrak{X}, \sigma(S), \mathbb{P}_\vartheta)$ mit $\int_{\mathfrak{X}} f \, \mathrm{d}\mathbb{P}_\vartheta = \int_{\mathfrak{X}} g \, \mathrm{d}\mathbb{P}_\vartheta$ für alle $\vartheta \in \Theta$. Aufgrund des Faktorisierungssatzes [4, Satz 8.11] existiert eine messbare Funktion $h\colon (\Sigma, \mathscr{S}) \to \big(\mathbb{R}, \mathcal{B}(\mathbb{R})\big)$ mit $h \circ S = f - g$. Dann gilt für alle $\vartheta \in \Theta$

$$\mathbb{E}_\vartheta[|h(S)|] = \mathbb{E}_\vartheta[|f - g|] \leqslant \mathbb{E}_\vartheta[|f|] + \mathbb{E}_\vartheta[|g|] \overset{f,g \in L^1}{<} \infty$$

sowie

$$0 = \int_{\mathfrak{X}} f(x) - g(x) \, \mathrm{d}\mathbb{P}_\vartheta(x) = \int_{\mathfrak{X}} h(S(x)) \, \mathrm{d}\mathbb{P}_\vartheta(x) = \mathbb{E}_\vartheta[h(S)].$$

$$\begin{tikzcd} (\mathfrak{X}, \sigma(S)) \arrow[r, "f-g"] \arrow[dr, "S"'] & (\mathbb{R}, \mathcal{B}(\mathbb{R})) \\ & (\Sigma, \mathscr{S}) \arrow[u, "\exists h"'] \end{tikzcd}$$

Abb. 3: Kommutatives Diagram für den Faktorisierungssatz.

Aufgrund der Vollständigkeit von S folgt $h(S) = f - g = 0$ $\mathbb{P}_\vartheta$-fast sicher für alle $\vartheta \in \Theta$.

" $\Longleftarrow$ ": Sei $h\colon \Sigma \to \mathbb{R}$ messbar, sodass $\mathbb{E}_\vartheta[|h(S)|] < \infty$ und $\mathbb{E}_\vartheta[h(S)] = 0$ für alle $\vartheta \in \Theta$ gilt. Dann können wir h als $h = h^+ - h^-$ zerlegen, wobei h^+ und h^- nichtnegativ und integrierbar sind. Aus der Linearität von $\mathbb{E}_\vartheta[h(S)] = 0$ folgt sowie

$$\mathbb{E}_\vartheta[h^+(S)] = \mathbb{E}_\vartheta[h^-(S)] \qquad \forall \vartheta \in \Theta$$

und somit nach Voraussetzung $\mathbb{P}_\vartheta[h^+(S) = h^-(S)] = 1$ und somit $\mathbb{P}_\vartheta[h(S) = 0] = 1$ für alle $\vartheta \in \Theta$. □

Lemma 1.4.21 (Konstante Schätzer sind vollständig)
Nimmt S nur einen Wert an, so ist S vollständig.

Beweis. Sei $S(x) := s$ für alle $x \in \mathfrak{X}$. Für eine messbare Funktion $h\colon \Sigma \to \mathbb{R}$ mit $\mathbb{E}_\vartheta[|h(S)|] < \infty$ mit $\mathbb{E}_\vartheta[h(S)] = 0$ für alle $\vartheta \in \Theta$ gilt

$$0 = \mathbb{E}_\vartheta[h(S)] = \mathbb{E}_\vartheta[h(s)] = h(s)$$

für alle $\vartheta \in \Theta$. Somit ist $h(s) = 0$, also folgt $h(S) = 0$. □

Beispiel 1.4.22 (Bernoulli-Produktmodell)
Im Bernoulli-Produktmodell ist $\mathfrak{X} = \{0,1\}^n$. Die Statistik $S(x) := \sum_{k=1}^n x_k$ ist vollständig, denn für $h\colon \{0, \dots, n\} \to \mathbb{R}$ ist

$$\begin{aligned} \mathbb{E}_\vartheta[h(S)] = \sum_{m=0}^n h(m) \mathbb{P}_\vartheta[S = m] &= \sum_{m=0}^n h(m) \binom{n}{m} \vartheta^m (1 - \vartheta)^{n-m} \\ &= \underbrace{(1-\vartheta)^n}_{>0} \sum_{m=0}^n h(m) \binom{n}{m} \left(\frac{\vartheta}{1 - \vartheta} \right)^m \end{aligned}$$

ein Polynom in $y := \frac{\vartheta}{1-\vartheta}$, welches genau dann für alle ϑ (das heißt, für alle $y \in (0, \infty)$) Null ist, wenn $h(m) = 0$ für alle $m \in \{0, \dots, n\}$ ist.

Jedoch ist $S(x) := \big(x_1, \frac{1}{n} \sum_{k=1}^n x_k\big)$ *nicht* vollständig für $n \geqslant 2$. Das kann man so sehen: für $h\colon [0,1]^2 \to \mathbb{R}$, $(s_1, s_2) \mapsto s_1 - s_2$ gilt

$$\mathbb{E}_\vartheta[h(S)] = \mathbb{E}_\vartheta \left[X_1 - \frac{1}{n} \sum_{k=1}^n X_k \right] = \vartheta - \frac{1}{n} \cdot n\vartheta = 0$$

für alle $\vartheta \in \Theta$, aber $\mathbb{P}_\vartheta[h(S) = 0] < 1$ für alle $\vartheta \in (0,1)$, da $n \geqslant 2$ ist (es ist $h(S(x)) = 0$ genau dann wenn $x_1 = \frac{1}{n-1} \sum_{k=2}^n x_k$ ist). ◇

Satz 1.4.4: Lehmann-Scheffé (1950, 1955)

Seien $\big(\mathfrak{X}, \mathcal{F}, (\mathbb{P}_\vartheta)_{\vartheta\in\Theta}\big)$ ein statistisches Modell und $T\colon \mathfrak{X} \to \mathbb{R}$ ein erwartungstreuer Schätzer für die Kenngröße $\tau\colon \Theta \to \mathbb{R}$ mit $\mathbb{E}_\vartheta[T^2] < \infty$. Ist $S\colon (\mathfrak{X}, \mathcal{F}) \to (\Sigma, \mathscr{S})$ suffizient und vollständig, dann ist

$$\tilde{T} := \mathbb{E}_\vartheta[T \mid \sigma(S)]$$

bester Schätzer für τ (nach Satz 1.4.1 kennen wir eine Version von $\tilde{T}$, welche unabhängig von ϑ ist).

Beweis. Existenz. Aus Satz 1.4.1 folgt: $\tilde{T}$ existiert und hängt nicht von ϑ ab und $\tilde{T}$ ist erwartungstreu mit $\mathbb{V}_\vartheta[\tilde{T}] \leqslant \mathbb{V}_\vartheta[T]$ für alle $\vartheta \in \Theta$. **17.05.2022**

UMVU. Zunächst folgt aus $\mathbb{E}_\vartheta[T^2] < \infty$ wie in (24) $\mathbb{E}_\vartheta[\tilde{T}^2] < \infty$ und somit ist $\mathbb{V}_\vartheta[\tilde{T}] < \infty$. Sei nun U ein erwartungstreuer Schätzer für τ. Sei analog zu $\tilde{T}$

$$\tilde{U} := \mathbb{E}_\vartheta[U \mid \sigma(S)].$$

Nach dem Satz von Rao-Blackwell gilt $\mathbb{V}_\vartheta[\tilde{U}] \leqslant \mathbb{V}_\vartheta[U]$ für alle $\vartheta \in \Theta$.

Nach dem Faktorisierungssatz existieren messbare Funktionen $t, u\colon (\Sigma, \mathscr{S}) \to \big(\mathbb{R}, \mathcal{B}(\mathbb{R})\big)$ mit $\tilde{T} = t \circ S$ und $U = u \circ S$. Für $\vartheta \in \Theta$ folgt

Alternative: Es gilt $\tilde{T}, \tilde{U} \in L^1(\mathfrak{X}, \sigma(S), \mathbb{P}_\vartheta)$ und $\mathbb{E}_\vartheta[\tilde{T}] = \mathbb{E}_\vartheta[\tilde{U}]$ für alle $\vartheta \in \Theta$ wegen der Turmeigenschaft und der Erwartungstreue von U und T, also folgt aus Lemma 1.4.19 $\tilde{T} = \tilde{U}$ $\mathbb{P}_\vartheta$-fast sicher für alle $\vartheta \in \Theta$.

$$0 = \tau(\vartheta) - \tau(\vartheta) = \mathbb{E}_\vartheta[\tilde{U} - \tilde{T}] = \mathbb{E}_\vartheta[(u - t)(S)].$$

Mit der Vollständigkeit folgt nach Lemma 1.4.19 $(u - t)(S) = 0$ $\mathbb{P}_\vartheta$-fast sicher für alle $\vartheta \in \Theta$ und somit $\tilde{T} = \tilde{U}$ $\mathbb{P}_\vartheta$-fast sicher.

Also folgt für alle $\vartheta \in \Theta$

$$\mathbb{V}_\vartheta[\tilde{T}] = \mathbb{V}_\vartheta[\tilde{U}] \overset{1.4.1}{\leqslant} \mathbb{V}_\vartheta[U]$$

und somit ist $\tilde{T}$ bester Schätzer. □

Lemma 1.4.23
Ist ein Modell exponentiell bezüglich einer Statistik T, dann ist T vollständig.

Beweis. Sei die Likelihoodfunktion

$$\rho(x, \vartheta) = \exp\left(\vartheta T(x) - b(\vartheta)\right) \tilde{h}(x),$$

wobei $\tilde{h}(x) > 0$, für alle $x \in \mathfrak{X}$ und $\vartheta \in \Theta$. Wir können wieder ohne Beschränkung der Allgemeinheit $a(\vartheta) = \vartheta$ wählen. Ferner sei $h\colon \mathbb{R} \to \mathbb{R}$ messbar mit $\mathbb{E}_\vartheta[|h(T)|] < \infty$ und $\mathbb{E}_\vartheta[h(T)] = 0$ für alle $\vartheta \in \Theta$.

Dann folgt mit $\mu_1 := \tilde{h}\mu_0$ und $\nu_1 := \mu_1 \circ T^{-1}$ für alle $\vartheta \in \Theta$

$$\begin{aligned}
0 = \mathbb{E}_\vartheta[h(T)] &= \int_{\mathfrak{X}} h(T(x)) \,\mathrm{d}\mathbb{P}_\vartheta(x) \\
&= \int_{\mathfrak{X}} h(T(x)) \exp\left(\vartheta T(x) - b(\vartheta)\right) \tilde{h}(x) \,\mathrm{d}\mu_0(x) \\
&= \int_{\mathfrak{X}} h(T(x)) \exp\left(\vartheta T(x) - b(\vartheta)\right) \mathrm{d}\mu_1(x) = e^{-b(\vartheta)} \int_{\mathbb{R}} h(y) e^{\vartheta y} \,\mathrm{d}\nu_1(y)
\end{aligned}$$

und somit ($h = h^+ - h^-$)

$$\int_{\mathbb{R}} h^+(y)e^{\vartheta y}\,\mathrm{d}\nu_1(y) = \int_{\mathbb{R}} h^-(y)e^{\vartheta y}\,\mathrm{d}\nu_1(y). \tag{26}$$

(Jetzt orientieren wir uns an [3].) Ohne Beschränkung der Allgemeinheit sei $0 \in \Theta$. Einsetzen von $\vartheta = 0$ gibt

$$w := \int_{\mathbb{R}} h^+(y)\,\mathrm{d}\nu_1(y) = \int_{\mathbb{R}} h^-(y)\,\mathrm{d}\nu_1(y) \geqslant 0.$$

Wir unterscheiden in zwei Fälle.

1. Ist $w = 0$, folgt $h^+(y) = h^-(y) = 0$ für ν_1-fast alle $y \in \mathbb{R}$ und somit $h(y) = 0$ für ν_1-fast alle $y \in \mathbb{R}$. Daraus folgt $h\big(T(x)\big) = 0$ für μ_1-fast alle $x \in \mathfrak{X}$, also $h\big(T(x)\big) = 0$ für $\mathbb{P}_\vartheta$-fast alle $x \in \mathfrak{X}$ für alle $\vartheta \in \Theta$.

2. Ist $w > 0$, definiere $\mathrm{d}\kappa_1(y) = \frac{h^+(y)}{w}\,\mathrm{d}\nu_1(y)$, das heißt $\kappa_1(A) = \int_A \frac{h^+(y)}{w}\,\mathrm{d}\nu_1(y)$ für $A \in \mathcal{B}(\mathbb{R})$. Analog definiere $\mathrm{d}\kappa_2(y) = \frac{h^-(y)}{w}\,\mathrm{d}\nu_1(y)$. Dann sind κ_1 und κ_2 Wahrscheinlichkeitsmaße auf $\big(\mathbb{R}, \mathcal{B}(\mathbb{R})\big)$ und wegen (26) gilt

$$\int_{\mathbb{R}} e^{\vartheta y}\,\mathrm{d}\kappa_1(y) = \int_{\mathbb{R}} e^{\vartheta y}\,\mathrm{d}\kappa_2(y)$$

für $\vartheta \in \Theta$. Auf der linken Seite steht die *momenterzeugende Funktion* von κ_1. Aus dem Eindeutigkeitssatz für momenterzeugende Funktionen folgt (da die Integrale für ϑ in einer Umgebung von 0 konvergieren) $\kappa_1 = \kappa_2$. Es folgt $h^+(y) = h^-(y)$ für ν_1-fast alle $y \in \mathbb{R}$ und somit wie in 1 $h\big(T(x)\big) = 0$ für $\mathbb{P}_\vartheta$-fast alle $x \in \mathfrak{X}$ für alle $\vartheta \in \Theta$. □

1.5 Konsistenz

Literatur: [2].

Definition 1.5.1 (Konsistente Schätzfolge)
Sei $\big((\mathfrak{X}_n, \mathcal{F}_n, (\mathbb{P}_{\vartheta,n})_{\vartheta\in\Theta})\big)_{n\in\mathbb{N}}$ eine Folge statistischer Modelle mit gleichem Θ. Seien $\tau\colon \Theta \to \mathbb{R}^d$ eine Kenngröße und $T_n\colon \mathfrak{X}_n \to \mathbb{R}^d$ ein Schätzer für τ für jedes $n \in \mathbb{N}$. Die Folge $(T_n)_{n\in\mathbb{N}}$ ist konsistent, wenn für alle $\varepsilon > 0$ und alle $\vartheta \in \Theta$ gilt, dass

konsistent

$$\mathbb{P}_{\vartheta,n}[|T_n - \tau(\vartheta)| \geqslant \varepsilon] \xrightarrow{n\to\infty} 0.$$

Bemerkung 1.5.2
Von besonderem Interesse ist der Produktfall: $\mathfrak{X}_n = E^n$ $\mathcal{F}_n = \mathfrak{E}^{\otimes n}$ und $\mathbb{P}_{\vartheta,n} = Q_\vartheta^{\otimes n}$, wobei $\big(E, \mathfrak{E}, (Q_\vartheta)_{\vartheta\in\Theta}\big)$ ein statistisches Modell ist. Der Einfachheit halber betrachten wir das unendliche Produkt $\big(\mathfrak{X}, \mathcal{F}, (\mathbb{P}_\vartheta)_{\vartheta\in\Theta}\big) = \big(E^{\mathbb{N}}, \mathfrak{E}^{\otimes\mathbb{N}}, (Q_\vartheta^{\otimes\mathbb{N}})_{\vartheta\in\Theta}\big)$ und vereinbaren, dass $T_n\colon \mathfrak{X} \to \mathbb{R}^d$ nur von den ersten n Komponenten abhängt. (In diesem Fall ist die obigen Konvergenz Konvergenz in Wahrscheinlichkeit. Konvergenz im p-ten Mittel und fast sichere Konvergenz implizieren Konvergenz in Wahrscheinlichkeit.) ○

Satz 1.5.1: Konsistenz von Stichprobenmittel und Stichprobenvarianz

Seien $(E, \mathfrak{E}) = \big(\mathbb{R}, \mathcal{B}(\mathbb{R})\big)$und $\int_{\mathbb{R}} x^2 \,\mathrm{d}Q_\vartheta(x) < \infty$ für alle $\vartheta \in \Theta$. Im Produktmodell seien $M_n := \frac{1}{n}\sum_{k=1}^n X_k$ das Stichprobenmittel nach n Beobachtungen sowie $V_n^* := \frac{1}{n-1}\sum_{k=1}^n (X_k - M_n)^2$ die korrigierte Stichprobenvarianz.
Dann sind M_n bzw. V_n^* erwartungstreue Schätzer für $m(\vartheta) := \int_{\mathbb{R}} x \,\mathrm{d}Q_\vartheta(x)$ bzw. $v(\vartheta) := \int_{\mathbb{R}}(x - m(\vartheta))^2 \,\mathrm{d}Q_\vartheta(x)$ und die Folgen $(M_n)_{n\in\mathbb{N}}$ bzw. $(V_n^*)_{n\in\mathbb{N}}$ konsistente Schätzfolgen für m bzw. v.

Beweis. (1) Die **Erwartungstreue** wurde in Lemma 1.2.16 gezeigt.

(2) **Konsistenz von M_n.** Für $\varepsilon > 0$ gilt

$$\mathbb{P}_{\vartheta,n}\big[|M_n - m(\vartheta)| > \varepsilon\big] = \mathbb{P}_{\vartheta,n}\left[\left|\frac{1}{n}\sum_{k=1}^n X_k - \mathbb{E}_{\vartheta,n}\right| > \varepsilon\right] \xrightarrow{n\to\infty} 0$$

nach dem Gesetz der großen Zahlen.

(3) **Konsistenz von V_n^*.** Vorbetrachtung: Im Hilbertraum $(\mathbb{R}^n, |\cdot|_2)$ ist die orthogonale Projektion von $x \in \mathbb{R}^n$ auf den eindimensionalen Unterraum $U := \operatorname{span}\{(1,\dots,1)\}$ gerade $M_n := (m_n, \dots, m_n)$, wobei $m_n := \frac{1}{n}\sum_{k=1}^n x_k$ (denn $\arg\min_m \sum_{k=1}^n (x_k - m)^2 = \frac{1}{n}\sum_{k=1}^n x_k$). Es gilt[5]

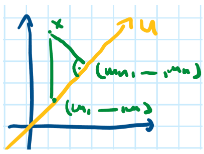


Abb. 4: Die orthogonale Projektion von x auf den Unterraum U.

[5] Seien $X := (x_1, \dots, x_n)$ und $M := (m, \dots, m) \in U$. Per Definition der orthogonalen Projektion ist $X - M_n$ orthogonal zu dem von M_n aufgespannten Unterraum U, insbesondere zu $M - M_n \in U$ und daher gilt $|X - M|_2^2 = |X - M_n + M_n - M|_2^2 = |X - M_n|_2^2 + |M - M_n|_2^2$. (Es gilt $|x + y|_2^2 = |x|_2^2 + |y|_2^2$ genau dann, wenn $x \perp y$.)

$$\sum_{k=1}^{n}(x_k - m_n)^2 = \sum_{k=1}^{n}(x_k - m)^2 - n(m - m_n)^2 \tag{27}$$

und somit

$$\begin{aligned} V_n^* &= \frac{n}{n-1}\frac{1}{n}\sum_{k=1}^{n}(x_k - m_n)^2 \\ &\overset{(27)}{=} \underbrace{\frac{n}{n-1}}_{\xrightarrow{n\to\infty}1} \underbrace{\frac{1}{n}\sum_{k=1}^{n}\big(x_k - m(\vartheta)\big)^2}_{\xrightarrow[\text{GGZ}]{n\to\infty}v(\vartheta)} - \underbrace{\frac{n}{n-1}}_{\xrightarrow{n\to\infty}1} \underbrace{\big(m_n - m(\vartheta)\big)^2}_{\xrightarrow{n\to\infty}0} \xrightarrow{n\to\infty} v(\vartheta). \end{aligned}$$

□

Beispiel 1.5.3 (Verschobene Exponentialverteilung) **18.05.2022**

Sei $\big(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), (\mathbb{P}_\vartheta^{\otimes n})_{\vartheta\in\mathbb{R}}\big)$ ein statistisches Modell, wobei $\mathbb{P}_\vartheta$ die Dichte $\rho(\vartheta,\cdot) = \frac{1}{\exp(\cdot)}e^{\vartheta}\,\mathbb{1}_{[\vartheta,\infty)}$ hat. Aus Beispiel 1.3.13 wissen wir, dass $T_{\mathrm{ML}}^{(n)}(x) =: x_{(1)}$ der Maximum-Likelihood-Schätzer für ϑ ist.

Die Folge $\big(T_{\mathrm{ML}}^{(n)}\big)_{n\in\mathbb{N}}$ ist konsistent, da für alle $\varepsilon > 0$ und alle $\vartheta \in \mathbb{R}$ gilt:

$$\mathbb{P}_\vartheta^{\otimes n}\Big[|T_{\mathrm{ML}}^{(n)} - \vartheta| > \varepsilon\Big] = \mathbb{P}_\vartheta^{\otimes n}\Big[\min_{1\leqslant k\leqslant n} X_k > \vartheta + \varepsilon\Big] = \mathbb{P}_\vartheta[X_1 > \vartheta + \varepsilon]^n = e^{-\varepsilon n} \xrightarrow{n\to\infty} 0,$$

da

$$\mathbb{P}_\vartheta[X_1 > \vartheta + \varepsilon] = \int_{\vartheta+\varepsilon}^{\infty} e^{\vartheta - x}\,\mathrm{d}x = \int_{\varepsilon}^{\infty} e^{-x}\,\mathrm{d}x = e^{-\varepsilon}.$$

◇

Satz 1.5.2: Hinreichendes Konsistenz-Kriterium

Gelten $\mathbb{E}_{\vartheta,n}[T_n] \xrightarrow{n\to\infty} \tau(\vartheta)$ und $\mathbb{V}_{\vartheta,n}[T_n] \xrightarrow{n\to\infty} 0$, so ist Folge $(T_n)_{n\in\mathbb{N}}$ konsistent.

Somit m

Beweis. Für feste $\varepsilon > 0$ und $\vartheta \in \Theta$ existiert ein $N_{\varepsilon,\vartheta} \in \mathbb{N}$, sodass $|\,\mathbb{E}_{\vartheta,n}[T_n] - \tau(\vartheta)| \leqslant \frac{\varepsilon}{2}$ für alle $n > N_{\varepsilon,\vartheta}$ gilt (wegen $\mathbb{E}_{\vartheta,n}[T_n] \xrightarrow{n\to\infty} \tau(\vartheta)$). Für $n > N_{\varepsilon,\vartheta}$ gilt mit der Markov-Ungleichung (M)

$$\begin{aligned} \mathbb{P}_{\vartheta,n}[|T_n - \tau(\vartheta)| > \varepsilon] &\overset{\triangle\neq}{\leqslant} \mathbb{P}_{\vartheta,n}[|T_n - \mathbb{E}_{\vartheta,n}[T_n]| + |\,\mathbb{E}_{\vartheta,n}[T_n]| - \tau(\vartheta)| > \varepsilon] \\ &\overset{(\star)}{\leqslant} \mathbb{P}_{\vartheta,n}\Big(|T_n - \mathbb{E}_{\vartheta,n}[T_n]| > \frac{\varepsilon}{2}\Big) + \mathbb{P}_{\vartheta,n}\Big(|\,\mathbb{E}_{\vartheta,n}[T_n] - \tau(\vartheta)| > \frac{\varepsilon}{2}\Big) \\ &\overset{(\mathrm{M})}{\leqslant} \frac{1}{\frac{\varepsilon^2}{4}}\underbrace{\mathbb{V}_{\vartheta,n}[T_n]}_{\xrightarrow{n\to\infty}0} + \underbrace{\mathbb{1}_{\{|\,\mathbb{E}_{\vartheta,n}[T_n]-\tau(\vartheta)|>\frac{\varepsilon}{2}\}}(n)}_{=0} \xrightarrow{n\to\infty} 0. \end{aligned}$$

In $(\star)$ nutzen wir, dass für nichtnegative Zufallsvariablen X und Y und $\delta > 0$ gilt:

$$\mathbb{P}[X + Y > \delta] \leqslant \mathbb{P}\Big[X > \frac{\delta}{2} \text{ oder } Y > \frac{\delta}{2}\Big] \leqslant \mathbb{P}\Big[X > \frac{\delta}{2}\Big] + \mathbb{P}\Big[Y > \frac{\delta}{2}\Big].$$

□

Bemerkung 1.5.4 (Verschobene Exponentialverteilung, Gameshow)
Die Aussage von Beispiel 1.5.3 folgt auch mit Satz 1.5.2, da wir aus Hausaufgabe 3.2 wissen, dass $\mathbb{E}_\vartheta[T^n_{\mathrm{ML}}] = \vartheta + \frac{1}{n}$ und $\mathbb{V}_\vartheta[T^{(n)}_{\mathrm{ML}}] = \frac{1}{n^2}$ gelten.

Die Konsistenzaussage aus Beispiel 1.1.15 folgt auch mit Satz 1.5.2, da wir aus Beispiel 1.1.15 $\mathbb{E}_\vartheta[T^n_{\mathrm{ML}}] = \frac{n}{n+1}\vartheta + \frac{1}{n}$ und $\mathbb{V}_\vartheta[T^{(n)}_{\mathrm{ML}}] = \frac{n}{(n+1)^2(n+2)}\vartheta^2$ wissen. ○

Beispiel 1.5.5 (Konsistente Schätzfolge für das Gameshow-Modell)
Betrachte $\big(\mathbb{R}^n_+, \mathcal{B}(\mathbb{R}^n_+), (\mathbb{P}^{\otimes n}_\vartheta := \mathcal{U}_{[0,\vartheta]^n})_{\vartheta\in(0,\infty)}\big)$ und setze $T_n(x) := (\prod_{k=1}^n x_k)^{\frac{1}{n}}$. Dann ist $(T_n)_{n\in\mathbb{N}}$ eine konsistente Schätzfolge für $\frac{\vartheta}{e}$. ◇

Beweis. Für alle $\vartheta \in (0,\infty)$ gilt

$$\begin{aligned}\mathbb{E}_\vartheta[T_n] &\overset{\text{u.i.v.}}{=} \mathbb{E}\left[X_1^{\frac{1}{n}}\right]^n = \left(\int_0^\vartheta \frac{1}{\vartheta} x^{\frac{1}{n}}\,\mathrm{d}x\right)^n \\ &= \left(\frac{1}{\vartheta}\frac{1}{1+\frac{1}{n}} x^{1+\frac{1}{n}}\bigg|_{x=0}^{\vartheta}\right)^n = \frac{1}{(1+\frac{1}{n})^n}\vartheta \xrightarrow{n\to\infty} \frac{\vartheta}{e}\end{aligned}$$

sowie analog

$$\mathbb{E}_\vartheta\left[T_n^2\right] = \left(\int_0^\vartheta \frac{1}{\vartheta} x^{\frac{2}{n}}\,\mathrm{d}x\right)^n = \left(\frac{1}{\vartheta}\frac{1}{1+\frac{2}{n}} x^{1+\frac{2}{n}}\bigg|_{x=0}^{\vartheta}\right)^n = \frac{1}{(1+\frac{2}{n})^n}\vartheta^2 \xrightarrow{n\to\infty} \frac{\vartheta^2}{e^2}$$

und somit

$$\mathbb{V}_\vartheta[T_n] = \mathbb{E}_\vartheta[T_n^2] - \mathbb{E}_\vartheta[T_n]^2 \xrightarrow{n\to\infty} \frac{\vartheta^2}{e^2} - \frac{\vartheta^2}{e^2} = 0.$$

Die Aussage folgt mit Satz 1.5.2. □

Beispiel 1.5.6 (Konsistente Schätzfolgen für besondere Kovarianzmatrizen)
Betrachte das statistische Modell $\big(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), (\mathcal{N}_{\vartheta\cdot\mathbf{1},C})_{\vartheta\in\mathbb{R}}\big)$, wobei $C = (C_{i,j})_{i,j=1}^n$ die Kovarianzmatrix mit $C_{i,j} = \rho j - i$ für $i \leqslant j$ ist, wobei $\rho_0 = \sigma^2$ und $\rho_k \in \mathbb{R}$ sind. Weiterhin sei $M_n := \frac{1}{n}\sum_{k=1}^n X_k$ ein Schätzer für ϑ. Es ist Hausaufgabe 5.1 zu zeigen, dass M_n

1. *keine* konsistente Schätzfolge für ϑ ist, wenn $\rho_{j-i} = \rho \in (0,\sigma^2]$ für alle $i<j$ gilt.
2. eine konsistente Schätzfolge für ϑ ist, wenn $\rho_{j-i} = c\gamma^{j-i}$ mit $\gamma \in (-1,1)$ für $i<j$ und $c \geqslant 0$. ◇

Beispiel 1.5.7 (Schätzfolge für Funktionen der ersten r Momente)
Seien $\big(\mathbb{R}, \mathcal{B}(\mathbb{R}), (\mathbb{P}_\vartheta)_{\vartheta\in\Theta}\big)$ ein statistisches Modell und $r \in \mathbb{N}$. Zu jedem $\vartheta \in \Theta$ und jedem $k \in \{1,\dots,r\}$ existiere das k-te Moment $m_k(\vartheta) := \mathbb{E}_\vartheta[X^k]$ von $\mathbb{P}_\vartheta$. Sei ferner $g\colon \mathbb{R}^r \to \mathbb{R}$ stetig und $\tau(\vartheta) := g\big(m_1(\vartheta),\dots,m_r(\vartheta)\big)$. Im zugehörigen Produktmodell $\big(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), (\mathbb{P}^{\otimes n})_{\vartheta\in\Theta}\big)$ ist

$$T_n(x) := g\left(\frac{1}{n}\sum_{k=1}^n x_k, \frac{1}{n}\sum_{k=1}^n x_k^2, \dots, \frac{1}{n}\sum_{k=1}^n x_k^r\right)$$

ein Schätzer für τ und $(T_n)_{n\in\mathbb{N}}$ eine konsistente Schätzfolge für τ (Hausaufgabe 5.2). ◇

Konsistenz von ML-Schätzern

19.05.2022

Definition 1.5.8 (Relative Entropie)
Seien P und Q Wahrscheinlichkeitsmaße auf $(E, \mathfrak{E})$ und μ_0 ein σ-endliches Maß auf $(E, \mathfrak{E})$ mit $P \ll \mu_0$ und $Q \ll \mu_0$. (So ein Maß existiert immer, z.B. $\mu_0 = P + Q$.) Seien $\rho := \frac{\mathrm{d}P}{\mathrm{d}\mu_0}$ und $\sigma := \frac{\mathrm{d}Q}{\mathrm{d}\mu_0}$. Dann ist

$$H(P;Q) := \int_E \rho(x) \ln\left(\frac{\rho(x)}{\sigma(x)}\right) \mathrm{d}\mu_0(x)$$

die relative Entropie (oder: Kullback-Leibler-Divergenz/Abstand/Information) von P bezüglich Q. relative Entropie

Hierbei seien $0 \cdot \ln\left(\frac{0}{0}\right) := 0$, $a \cdot \ln\left(\frac{a}{0}\right) := \infty$ und $0 \cdot \ln\left(\frac{0}{a}\right) := 0$ für alle $a > 0$.

Bemerkung 1.5.9 Es ist zumindest unklar, ob das Integral definiert ist und ob $H(P;Q)$ von μ_0 abhängt. ○

Lemma 1.5.10
Es gilt $P \ll Q$ genau dann wenn $P(\{\sigma = 0\}) = 0$ ist.

Beweis. Es gilt immer $Q(\{\sigma = 0\}) = \int_{\{\sigma=0\}} \sigma(x)\,\mathrm{d}\mu_0(x) = 0$.

"$\implies$": Sei $P \ll Q$, dann gilt wegen $Q(\{\sigma = 0\}) = 0$ auch $P(\{\sigma = 0\}) = 0$.

"$\impliedby$": Seien $P(\{\sigma = 0\}) = 0$ und $A \in \mathfrak{E}$ mit $Q(A) = 0$. Angenommen, $P(A) > 0$. Dann gilt

$$0 < P(A) = P(A \backslash \{\sigma = 0\}) + P(\{\sigma = 0\}) = P(A \backslash \{\sigma = 0\}).$$

Aus $P(A \backslash \{\sigma = 0\}) > 0$ und $P \ll \mu_0$ folgt $\mu_0(A \backslash \{\sigma = 0\}) > 0$. Es folgt

$$0 = Q(A) = \int_A \sigma(x)\,\mathrm{d}\mu_0(x) \geqslant \int_{A \backslash \{\sigma=0\}} \sigma(x)\,\mathrm{d}\mu_0(x) > 0,$$

was einen Widerspruch darstellt. □

Lemma 1.5.11
Es gilt

$$H(P;Q) = \begin{cases} \int_E \frac{\mathrm{d}P}{\mathrm{d}Q}(x) \ln\left(\frac{\mathrm{d}P}{\mathrm{d}Q}(x)\right) \mathrm{d}Q(x), & \text{wenn } P \ll Q, \\ \infty, & \text{sonst.} \end{cases}$$

Insbesondere ist $H(P;Q)$ unabhängig von μ_0 und das Integral ist wohldefiniert und größer als $-\infty$.

Beweis. Sei $P \not\ll Q$. Nach Lemma 1.5.10 folgt $P(\{\sigma = 0\}) > 0$. Weiter gilt $P(\{\rho > 0\}) = 1$, also folgt $P(\{\rho > 0\} \cap \{\sigma = 0\}) > 0$ und somit $\mu_0(\{\rho > 0\} \cap \{\sigma = 0\}) > 0$. Betrachte

$$H(P;Q) = \underbrace{\int_{\{\rho=0\}} 0\,\mathrm{d}\mu_0(x)}_{=0} + \underbrace{\int_{\{\sigma=0\}\cap\{\rho>0\}} \infty\,\mathrm{d}\mu_0(x)}_{=\infty,\text{ da } \mu_0(\{\rho>0\}\cap\{\sigma=0\})>0} + \int_{\{\sigma>0\}\cap\{\rho>0\}} \rho(x) \ln\left(\frac{\rho(x)}{\sigma(x)}\right) \mathrm{d}\mu_0(x).$$

Der letzte Summand ist gleich

$$\int_{\{\sigma>0\}\cap\{\rho>0\}} \sigma(x)\frac{\rho(x)}{\sigma(x)} \ln\left(\frac{\rho(x)}{\sigma(x)}\right) \mathrm{d}\mu_0(x) = \int_{\{\sigma>0\}\cap\{\rho>0\}} \underbrace{f(x)\ln\left(f(x)\right)}_{\geqslant C} \mathrm{d}Q(x)$$

für ein $C \in \mathbb{R}$ und somit nach unten beschränkt. Daher ist $H(P;Q)$ wohldefiniert und es folgt $H(P;Q) = \infty$.

Seien $P \ll Q$ und $f := \frac{\mathrm{d}P}{\mathrm{d}Q}$. Dann gilt $\frac{\mathrm{d}P}{\mathrm{d}\mu_0} = \frac{\mathrm{d}P}{\mathrm{d}Q}\frac{\mathrm{d}Q}{\mathrm{d}\mu_0}$ μ_0-fast sicher. □

Lemma 1.5.12 (Eigenschaften der relativen Entropie)
Es gilt $H(P;Q) \geqslant 0$ und $H(P;Q) = 0$ genau dann, wenn $P = Q$ gilt.

Beweis. (1) Die Aussage $H(P;Q) \geqslant 0$ ist klar, wenn $P \not\ll Q$ gilt. Sei also $P \ll Q$. Da die Funktion $\psi(y) := y\ln(y)\,\mathbb{1}_{(0,\infty)}(y)$ strikt konvex ist, folgt mit der JENSEN-Ungleichung

$$H(P;Q) = \int_E \frac{\mathrm{d}P}{\mathrm{d}Q} \ln\left(\frac{\mathrm{d}P}{\mathrm{d}Q}\right) \mathrm{d}Q = \int_E \psi\left(\frac{\mathrm{d}P}{\mathrm{d}Q}\right) \mathrm{d}Q \geqslant \psi\left(\int_E \frac{\mathrm{d}P}{\mathrm{d}Q}\,\mathrm{d}Q\right) = \psi(1) = 0.$$

(2) Ist $P = Q$, so ist $\frac{\mathrm{d}P}{\mathrm{d}Q} \equiv 1$ fast sicher und daher $H(P;Q) = 0$. Sei $H(P;Q) = 0$, dann gilt $P \ll Q$ und somit Q-fast sicher $\frac{\mathrm{d}P}{\mathrm{d}Q} \equiv C$, da ψ strikt konvex ist (aus der vorherigen Ungleichung). Wegen $1 = P(E) = \int_E \frac{\mathrm{d}P}{\mathrm{d}Q}(x)\,\mathrm{d}Q(x)$ folgt $\frac{\mathrm{d}P}{\mathrm{d}Q} \equiv 1$ Q-fast sicher und daher $P = Q$. □

Bemerkung 1.5.13 (Anschauung für die Entropie) Im Spezialfall, dass $|E| < \infty$ gilt und μ_0 das Zählmaß ist, ist die „Entropie von P"

$$H(P) := -\sum_{p\in E} p\ln(p),$$

welche z.B. die mittlere Anzahl von Bits pro Buchstabe bei optimaler Codierung gibt. Dann ist $H(P;Q)$ die Zusatzkosten, die man benötigt, wenn man als (HUFFMANN-)Codierung die Häufigkeit in z.B. deutscher Sprache verwendet hat, der Text jedoch in z.B. englisch ist. ∘

SATZ 1.5.3: KONSISTENZ VON ML-SCHÄTZERN

Sei $\left(E, \mathfrak{E}, (Q_\vartheta)_{\vartheta\in\Theta}\right)$ ein Standardmodell mit Likelihoodfunktion ρ. Wir betrachten das (abzählbar) unendliche Produktmodell. Es gelte:

(1) Θ ist ein offenes Intervall und für $\vartheta \neq \vartheta'$ ist $Q_\vartheta \neq Q_{\vartheta'}$.

(2) Für jedes $n \in \mathbb{N}$ und jeden Vektor $(x_1, \dots, x_n) \in E^n$ ist $\prod_{k=1}^n \rho(x_k,\cdot)$ unimodal, das heißt für einen ML-Schätzer T_n für ϑ ist $\prod_{k=1}^n \rho(x_k,\cdot)$ wachsend für $\vartheta < T_n(x)$ und fallend für $\vartheta > T_n(x)$. unimodal

Dann ist die Folge $(T_n)_{n\in\mathbb{N}}$ für ϑ konsistent.

Bemerkung 1.5.14 (Hinreichende Bedingung für (2)) Die Bedingung (2) ist erfüllt, wenn $\ln\left(\rho(y,\cdot)\right)$ für alle $y \in E$ konkav ist mit zuerst positiver und am Ende negativer strikter Veränderung (bzw. Ableitung), denn dasselbe gilt für $\ln\left(\prod_{k=1}^n \rho(x_k,\cdot)\right)$, da endliche Summen konkaver Funktionen konkav sind. ∘

Bemerkung 1.5.15 (Vererbung der Unimodalität) Das Produkt unimodaler Funktionen ist im Allgemeinen *nicht* unimodal. ○

Beweis. (von Satz 1.5.3) Seien $\vartheta \in \Theta$ und $\varepsilon > 0$, sodass $\vartheta \pm \varepsilon \in \Theta$. Nach Lemma 1.5.12 und Voraussetzung 1 existiert ein $\delta \in \big(0, H(Q_\vartheta; Q_{\vartheta\pm\varepsilon})\big)$. Wir zeigen (später)

$$Q_\vartheta^{\otimes n}\left[\frac{1}{n}\ln\left(\frac{\rho_\vartheta^{\otimes n}}{\rho_{\vartheta\pm\varepsilon}^{\otimes n}}\right) > \delta\right] \xrightarrow{n\to\infty} 1, \tag{28}$$

wobei $\rho_\vartheta^{\otimes n} := \prod_{k=1}^n \rho(x_k, \vartheta)$.

Wenn (28) gilt, so folgt für $n \in \mathbb{N}$

$$\bigcap_{\sigma\in\{-1,1\}} \left\{\frac{1}{n}\ln\left(\frac{\rho_\vartheta^{\otimes n}}{\rho_{\vartheta+\sigma\varepsilon}^{\otimes n}}\right) > \delta\right\} \subset \left\{\rho_{\vartheta-\varepsilon}^{\otimes n} < \rho_\vartheta^{\otimes n} > \rho_{\vartheta+\varepsilon}^{\otimes n}\right\} \subset \{\vartheta - \varepsilon < T_n < \vartheta + \varepsilon\},$$

Die erste Inklusion folgt so: Wenn $\frac{1}{n}\ln\left(\frac{\rho_\vartheta^{\otimes n}}{\rho_{\vartheta+\sigma\varepsilon}^{\otimes n}}\right) > \delta$ gilt, dann auch

$$\frac{\rho_\vartheta^{\otimes n}}{\rho_{\vartheta+\sigma\varepsilon}^{\otimes n}} > e^{n\delta} > 1.$$

Die zweite Inklusion folgt aus der Unimodalität.

Wegen (28) folgt

$$\mathbb{P}_\vartheta[\vartheta - \varepsilon < T_n < \vartheta + \varepsilon] \xrightarrow{n\to\infty} 1.$$

Wir zeigen nun (28) mit $\pm = +$. **24.05.2022**

1 Fall 1. $Q_\vartheta \ll Q_{\vartheta+\varepsilon}$. Sei $f := \frac{\mathrm{d}Q_\vartheta}{\mathrm{d}Q_{\vartheta+\varepsilon}}$ eine Version der Radon-Nikodým-Dichte. Aufgrund der Absolutstetigkeit aller Maße bezüglich μ_0 folgt $f = \frac{\rho_\vartheta}{\rho_{\vartheta+\varepsilon}}$ (vgl. Beweis von 1.54) fast sicher.

Dann gilt

$$\frac{1}{n}\ln\left(\frac{\rho_\vartheta^{\otimes n}}{\rho_{\vartheta\pm\varepsilon}^{\otimes n}}\right) = \frac{1}{n}\sum_{k=1}^n \ln\left(\frac{\rho_\vartheta}{\rho_{\vartheta+\varepsilon}}(X_k)\right) = \frac{1}{n}\sum_{k=1}^n \ln\big(f(X_k)\big) =: (\star\star).$$

Es sind $\big(\ln\big(f(X_i)\big)\big)_{k=1}^n$ u.i.v. bezüglich $\mathbb{P}_\vartheta = Q_\vartheta^{\otimes n}$. Wir wollen ein Gesetz der großen Zahlen anwenden.

Es gilt

$$\mathbb{E}_\vartheta\big[\ln\big(f(X_1)\big)\big] = \int_{\mathfrak{X}} \ln(f)\,\mathrm{d}Q_\vartheta = \int_{\mathfrak{X}} f\ln(f)\,\mathrm{d}Q_{\vartheta+\varepsilon} \overset{1.54}{=} H\big(Q_\vartheta; Q_{\vartheta+\varepsilon}\big) > \delta > 0.$$

Wenn zusätzlich $H\big(Q_\vartheta; Q_{\vartheta+\varepsilon}\big) < \infty$ gilt, so folgt aus beiden großen Gesetzen der großen Zahlen

$$\lim_{n\to\infty}\frac{1}{n}\ln\left(\frac{\rho_\vartheta^{\otimes n}}{\rho_{\vartheta\pm\varepsilon}^{\otimes n}}\right) = H\big(Q_\vartheta; Q_{\vartheta+\varepsilon}\big) > \delta$$

und somit (28).

Ist $H(Q_\vartheta; Q_{\vartheta+\varepsilon}) = \infty$, dann folgt mit einem „Abschneideargument“ die Behauptung (28), denn für $c > 0$ gilt

$$\mathbb{E}_\vartheta \left[\ln\left(f(X_1)\right) \wedge c\right] \underset{c\to\infty}{\uparrow} H(Q_\vartheta; Q_{\vartheta+\varepsilon}) = \infty$$

(mit dem Satz über monotone Konvergenz und da $x \mapsto x\ln(x)$ nach unten beschränkt ist). Wähle nun $c > 0$ so groß, dass die linke Seite größer als δ ist. Dann folgt, da ln monoton wachsend ist,

$$\frac{1}{n}\sum_{k=1}^n \ln\left(f(X_k)\right) \geqslant \frac{1}{n}\sum_{k=1}^n \left(\ln\left(f(X_k)\right) \wedge c\right) \xrightarrow{n\to\infty} \mathbb{E}_\vartheta\left[\ln\left(f(X_1)\right) \wedge c\right] > \delta$$

mit dem Gesetz der großen Zahlen, und das zeigt (28).

2 $Q_\vartheta \not\ll Q_{\vartheta+\varepsilon}$. Dann gilt $Q_\vartheta[\rho_{\vartheta+\varepsilon} = 0] = \alpha > 0$ und $Q_\vartheta^{\otimes n}[\rho_\vartheta^{\otimes n} > 0] = 1$ (cf. (1.53) mit anderen Bezeichnungen). Dann folgt

$$\begin{aligned}
Q_\vartheta^{\otimes n}\left[\frac{1}{n}\ln\left(\frac{\rho_\vartheta^{\otimes n}}{\rho_{\vartheta+\varepsilon}^{\otimes n}}\right) = \infty\right] &= Q_\vartheta^{\otimes n}[\rho_\vartheta^{\otimes n} > 0 \text{ und } \rho_{\vartheta+\varepsilon}^{\otimes n} = 0] \\
&= Q_\vartheta^{\otimes n}[\rho_{\vartheta+\varepsilon}^{\otimes n} = 0] \\
&= 1 - Q_\vartheta^{\otimes n}[\rho_{\vartheta+\varepsilon}^{\otimes n} > 0] \\
&= 1 - Q_\vartheta^{\otimes n}\left[\bigcap_{i=1}^n \{\rho_{\vartheta+\varepsilon}(X_i) > 0\}\right] \\
&= 1 - \prod_{i=1}^n Q_\vartheta^{\otimes n}[\rho_{\vartheta+\varepsilon}(X_i) > 0] \\
&= 1 - (1 - Q_\vartheta[\rho_{\vartheta+\varepsilon} = 0])^n \xrightarrow[\alpha>0]{n\to\infty} 1.
\end{aligned}$$

□

Beispiel 1.5.16 (Konsistente Folge von ML-Schätzern für das Poisson-Modell) Seien $\rho(x,\vartheta) := e^{-\vartheta}\frac{\vartheta^x}{x!}$, $\vartheta \in \Theta := (0,\infty)$, $E := \mathbb{N}_0$ und μ_0 das Zählmaß. Die Bedingung 1 ist offensichtlich erfüllt. Wir überprüfen die log-Konkavität der Dichte, um die Bedingung 2 zu verifizieren. Es gilt

$$\ln\left(\rho(x,\vartheta)\right) = -\vartheta + x\log(\vartheta) - \log(x!)$$

und somit ist

$$\frac{\partial}{\partial\vartheta}\ln\left(\rho(x,\vartheta)\right) = \frac{x}{\vartheta} - 1$$

fallend bezüglich ϑ.

Ist $(T_n)_{n\in\mathbb{N}}$ eine ML-Folge, dann folgt aus Satz 1.5.3 die Konsistenz.

Wir berechnen nun den ML-Schätzer. Die Produktdichte ist $\rho^{\otimes n}(x,\vartheta) = \prod_{k=1}^n e^{-\vartheta}\frac{\vartheta^{x_k}}{x_k!}$ und somit gilt

$$\ln\left(\rho^{\otimes n}(x,\vartheta)\right) = -n\vartheta + \sum_{k=1}^n x_k\log(\vartheta) - \log(x_k!).$$

Nullsetzen der Ableitung bezüglich ϑ ergibt

$$\frac{\partial}{\partial\vartheta}\ln\left(\rho^{\otimes n}(x,\vartheta)\right) = -n + \frac{1}{\vartheta}\sum_{k=1}^n x_k \overset{!}{=} 0.$$

Also wird $\rho^{\otimes n}(x,\vartheta)$ maximal für $\vartheta_0 := \frac{1}{n}\sum_{k=1}^n x_k$. ◇

Beispiel 1.5.17 (Konsistente Folge von ML-Schätzern fürs Gameshow-Modell)
Wir wissen, dass $\tilde{T}_n(x) := x_{(n)}$ ein Maximum-Likelihood-Schätzer ist und, dass die Folge $\big(\tilde{T}_n\big)_{n\in\mathbb{N}}$ nach Beispiel 1.1.15 konsistent ist. Die Konsistenz folgt auch mit Satz 1.5.3, denn

$$\prod_{k=1}^{n} \rho(x_k, \vartheta) = \vartheta^{-n}\, \mathbb{1}_{[0,\vartheta]}(x_{(n)})$$

ist unimodal für jedes $n \in \mathbb{N}$, wie man in Abb. 5 sehen kann.

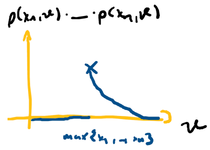


Abb. 5: Die Likelihoodfunktion im Gameshow-Modell ist unimodal.

Aber die Dichte ist nicht log-konkav:

$$\ln\left(\frac{1}{\vartheta}\, \mathbb{1}_{[0,\vartheta]}(x_1)\right) = -\ln(\vartheta) + \ln\big(\mathbb{1}_{[0,\vartheta]}(x_1)\big)$$

◇

1.6 Bayes-Schätzer

Definition 1.6.1 (a-priori Verteilung, a-posteriori Verteilung)
Seien $\big(\mathfrak{X}, \mathcal{F}, (\mathbb{P}_\vartheta)_{\vartheta\in\Theta}\big)$ ein Standardmodell und $(\Theta, \mathcal{T})$ ein Messraum. Sei ferner die Likelihoodfunktion ρ gemeinsam $\big(\mathbb{R}, \mathcal{B}(\mathbb{R})\big)$-messbar auf $\mathcal{F}\otimes\mathcal{T}$. Ein Wahrscheinlichkeitsmaß μ auf $(\Theta, \mathcal{T})$ ist eine a-priori Verteilung (oder: Vorbewertung). a-priori Verteilung

Die a-posteriori Verteilung zur a-priori Verteilung μ ist für $x \in \mathfrak{X}$ ist das Wahrscheinlichkeitsmaß ν_x auf $(\Theta, \mathcal{T})$ definiert durch a-posteriori Verteilung

$$\nu_x(A) = \begin{cases} \frac{1}{Z_x}\int_A \rho(x,\vartheta)\,\mathrm{d}\mu(\vartheta), & \text{wenn } Z_x \in (0,\infty), \\ \mu(A), & \text{wenn } Z_x \in \{0,\infty\}, \end{cases}$$

für $A \in \mathcal{T}$, wobei $Z_x := \int_\Theta \rho(x,\vartheta)\,\mathrm{d}\mu(\vartheta)$ eine Normalisierungskonstante ist.

Bemerkung 1.6.2 Die a-posteriori Verteilung ν_x ist ein Wahrscheinlichkeitsmaß für alle $x \in \mathfrak{X}$ und hat die Dichte $\frac{1}{Z_x}\rho(x,\cdot)$ bezüglich μ, wenn $Z_x \in (0,\infty)$ ist und 1 sonst. ○

Beispiel 1.6.3 ($\mu = \delta_{\vartheta_0}$) Ist $\mu = \delta_{\vartheta_0}$ für $\vartheta_0 \in \Theta$ (drückt aus, dass wir denken, dass ϑ_0 genau das richtige ϑ ist), so folgt $\nu_x = \mu$ für alle $x \in \mathfrak{X}$. ◇

Bemerkung 1.6.4 In diesem Szenario ist $\mathbb{P}_\vartheta$ die bedingte Verteilung der Stichprobe X gegeben, dass der Parameter ϑ ist, μ die Randverteilung der Parameters und ν_x die bedingte Verteilung des Parameters gegeben der Stichprobe $X = x$. ○

Beispiel 1.6.5 ($\mathbb{P}_\vartheta = \mathrm{Bin}(n,\vartheta)$, $\mu = \mathcal{U}_\Theta$, $\nu_x^{(n)} = \mathrm{Beta}(x+1, n-x+1)$)
Seien $\Theta := (0,1)$, $\mathfrak{X} := \{0,\dots,n\}$ und $\mathbb{P}_\vartheta(\{x\}) = \rho(x,\vartheta) := \binom{n}{x}\vartheta^x(1-\vartheta)^{n-x}$. Sei μ eine a-priori Verteilung auf Θ. Für die a-posteriori Verteilung gilt nach Bemerkung 1.6.2

$$\frac{\mathrm{d}\nu_x^{(n)}}{\mathrm{d}\mu}(\vartheta) = \frac{\rho(x,\vartheta)}{Z_x},$$

wobei $Z_x := \int_0^1 \rho(x,\tilde{\vartheta})\,\mathrm{d}\mu(\tilde{\vartheta})$.

In Spezialfall, dass μ die Gleichverteilung auf Θ ist, gilt

$$Z_x = \int_0^1 \binom{n}{x}\vartheta^x(1-\vartheta)^{n-x}\,\mathrm{d}\vartheta = \binom{n}{x}B(x+1, n-x+1),$$

wobei B die Betafunktion ist. Es folgt

$$\frac{\mathrm{d}\nu_x^{(n)}}{\mathrm{d}\mu}(\vartheta) = \frac{\vartheta^x(1-\vartheta)^{n-x}}{B(x+1,n-x+1)},$$

und das ist die Dichte der Betaverteilung.

Es ist Hausaufgabe 6.2 zu zeigen, dass für alle $\varepsilon > 0$ gilt:

$$\sup_{x\in\mathfrak{X}} \nu_x^{(n)}\left(\left[\frac{x+1}{n+2}-\varepsilon, \frac{x+1}{n+2}+\varepsilon\right]\right) \xrightarrow{n\to\infty} 1$$

$$\mathbb{P}_\vartheta^{\otimes n}\left[x\in\mathfrak{X} : \nu_x^{(n)}\left([\vartheta-\varepsilon, \vartheta+\varepsilon]\right) \geqslant 1-\varepsilon\right] \xrightarrow{n\to\infty} 0 \qquad \forall\vartheta\in[0,1].$$

◇

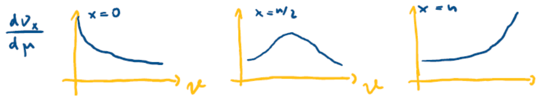


Abb. 6: Die Steigung der Dichte der a-posteriori-Verteilung wächst mit n.

Beispiel 1.6.6 ($\mathbb{P}_\vartheta = \mathcal{U}_{[0,\vartheta]}$, $\mu \sim \mathrm{Par}(k,a)$, $\nu_x^{(n)} \sim \mathrm{Par}\left(\max\{x_{(n)},k\}, n+a\right)$)
Gegeben sei das n-fache Produktmodell von $\left(\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+), (\mathcal{U}_{[0,\vartheta]})_{\vartheta>0}\right)$ und $\mathrm{Par}(k,a)$ für $k,a>0$ die a-priori Verteilung. Es ist Hausaufgabe 6.1(i) zu zeigen, dass die a-priori Verteilung ν_x $\mathrm{Par}\left(\max\{x_{(n)},k\}, n+a\right)$-verteilt ist. ◇

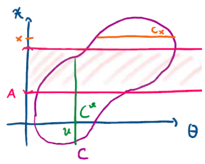
Beispiel 1.6.7 ($\mathbb{P}_\vartheta = \mathrm{Bin}(n,\vartheta)$, $\mu \sim \mathrm{Beta}(a,b)$, $\nu_k \sim \mathrm{Beta}(k+a, n-k+b)$)
Betrachte das statistische Modell mit $\mathfrak{X} := \{0,\dots,n\}$, $\mathcal{F} := \mathcal{P}(\mathfrak{X})$ und $\mathbb{P}_\vartheta := \mathrm{Bin}(n,\vartheta)$ für $\vartheta \in \Theta := (0,1)$. Die a-priori Verteilung sei die $\mathrm{Beta}(a,b)$-Verteilung mit $a,b>0$. Es ist Hausaufgabe 6.1(ii) zu zeigen, dass die a-priori Verteilung ν_k $\mathrm{Beta}(k+a, n-k+b)$-verteilt ist. ◇

31.05.2022

Definition 1.6.8 (Q)
Wir definieren auf $(\mathfrak{X}\times\Theta, \mathcal{F}\otimes\mathcal{T})$ das Wahrscheinlichkeitsmaß Q durch

$$Q(C) := \int_C \rho(x,\vartheta)\,\mathrm{d}\mu_0(x)\,\mathrm{d}\mu(\vartheta), \qquad C \in \mathcal{F}\otimes\mathcal{T},$$

welches die gemeinsame Verteilung von x und ϑ beschreibt.



Es ist Q ein Wahrscheinlichkeitsmaß: mit Fubini folgt

$$Q(\mathfrak{X}\times\Theta) = \int_\Theta \int_{\mathfrak{X}} \rho(x,\vartheta)\,\mathrm{d}\mu_0(x)\,\mathrm{d}\mu(\vartheta) = \int_\Theta \underbrace{\mathbb{P}_\vartheta(\mathfrak{X})}_{=1}\,\mathrm{d}\mu(\vartheta) = 1,$$

da $\mathbb{P}_\vartheta$ und μ Wahrscheinlichkeitsmaße sind. Nach Konstruktion hat Q die Dichte ρ bezüglich des Produktmaßes $\mu_0 \otimes \mu$.

Für $A \in \mathcal{F}$ gilt mit Fubini

$$Q(A\times\Theta) = \int_A \underbrace{\left(\int_\Theta \rho(x,\vartheta)\,\mathrm{d}\mu(\vartheta)\right)}_{=Z_x}\mathrm{d}\mu_0(x) \overset{(\mathrm{F})}{=} \int_\Theta \mathbb{P}_\vartheta[A]\,\mathrm{d}\mu(\vartheta) \tag{29}$$

und für $B \in \mathcal{T}$ mit Fubini

$$Q(\mathfrak{X}\times B) = \int_B \underbrace{\left(\int_{\mathfrak{X}} \rho(x,\vartheta)\,\mathrm{d}\mu_0(x)\right)}_{=\mathbb{P}_\vartheta[\mathfrak{X}]=1}\mathrm{d}\mu(\vartheta) = \mu(B). \tag{30}$$

Bezeichnen $\pi_{\mathfrak{X}}$ und π_Θ die Projektionen von $\mathfrak{X}\times\Theta$ auf $\mathfrak{X}$ bzw. Θ, so folgt wegen (29) $(Q\circ\pi_{\mathfrak{X}}^{-1})(A) = \int_A Z_x\,\mathrm{d}\mu_0(x)$ für $A\in\mathcal{F}$ und wegen (30) $Q\circ\pi_\Theta^{-1} = \mu$.

Insbesondere ist $Z_x > 0$ $Q\circ\pi_{\mathfrak{X}}^{-1}$-fast sicher, da $(Q\circ\pi_{\mathfrak{X}}^{-1})(\{x : Z_x = 0\}) = \int_{\{x:Z_x=0\}} Z_x\,\mathrm{d}\mu_0(x) = 0$. Weiter gilt für $C\in\mathcal{F}\otimes\mathcal{T}$ mit

$$C^{(\vartheta)} := \{x\in\mathfrak{X} : (x,\vartheta)\in C\} \qquad C_{(x)} := \{\vartheta\in\Theta : (x,\vartheta)\in C\}$$

(welche messbare Mengen sind)

$$Q(C) = \int_C \rho(x,\vartheta)\,\mathrm{d}\mu_0(x)\,\mathrm{d}\mu(\vartheta) = \int_\Theta \underbrace{\left(\int_{C^{(\vartheta)}} \rho(x,\vartheta)\,\mathrm{d}\mu_0(x)\right)}_{=\mathbb{P}_\vartheta(C^{(\vartheta)})} \mathrm{d}\mu(\vartheta)$$

aber auch

$$\begin{aligned}Q(C) &= \int_{\mathfrak{X}} \left(\int_{C_{(x)}} \rho(x,\vartheta)\,\mathrm{d}\mu(\vartheta)\right)\mathrm{d}\mu_0(x) = \int_{\mathfrak{X}\cap\{x:Z_x>0\}} \left(\int_{C_{(x)}} \rho(x,\vartheta)\,\mathrm{d}\mu(\vartheta)\right)\mathrm{d}\mu_0(x)\\ &= \int_{\mathfrak{X}\cap\{x:Z_x>0\}} \underbrace{\left(\int_{C_{(x)}} \frac{\rho(x,\vartheta)}{Z_x}\,\mathrm{d}\mu(\vartheta)\right)}_{=\nu_x(C_{(x)})} Z_x\,\mathrm{d}\mu_0(x).\end{aligned}$$

Also ist ν_x die bedingte Verteilung für ϑ gegeben einer Beobachtung x.

Wir wollen nun das ϑ schätzen.

Definition 1.6.9 (Bayes-Schätzer, $\mathbb{F}_T$)
Sei $\tau\colon \Theta \to \mathbb{R}$ messbar mit $\mathbb{E}_\mu[\tau^2] < \infty$. Die Statistik $T\colon \mathfrak{X} \to \mathbb{R}$ ist ein Bayes-Schätzer für τ zur a-priori-Verteilung μ, wenn

Bayes-Schätzer

$$\begin{aligned}\mathbb{F}_T &:= \int_{\mathfrak{X}\times\Theta} \big(T(x)-\tau(\vartheta)\big)^2\,\mathrm{d}Q(x,\vartheta)\\ &= \int_\Theta \left(\int_{\mathfrak{X}} \big(T(x)-\tau(\vartheta)\big)^2 \rho(x,\vartheta)\,\mathrm{d}\mu_0(x)\right)\mathrm{d}\mu(\vartheta) = \int_\Theta \mathbb{F}_\vartheta[T]\,\mathrm{d}\mu(\vartheta)\end{aligned}$$

minimal ist unter allen Schätzern T für τ.

Bemerkung 1.6.10 Sei $\|\cdot\|_2$ die L^2-Norm auf $(\mathfrak{X}\times\Theta, \mathcal{F}\otimes\mathcal{T}, Q)$. Dann ist

$$\mathbb{F}_T = \|T(\pi_X) - \tau(\pi_\Theta)\|_2^2.$$

○

Bemerkung 1.6.11 (Endliches Risiko des Bayes-Schätzers)
Existiert ein Bayes-Schätzer für τ, so ergibt der Vergleich mit dem Nullschätzer

$$\mathbb{F}_T \leqslant \int_{\mathfrak{X}\times\Theta} \tau^2(\vartheta)\,\mathrm{d}Q(x,\vartheta) = \int_\Theta \tau^2(\vartheta)\,\mathrm{d}\mu(\vartheta) = \mathbb{E}_\mu[\tau^2] < \infty.$$

○

Bemerkung 1.6.12 Es ist naheliegend zu vermuten, dass für den Bayes-Schätzer die Formel $T(x) = \int_\Theta \tau(\vartheta)\,\mathrm{d}\nu_x(\vartheta)$ gilt, da $\int_\Theta \tau(\vartheta)\,\mathrm{d}\nu_x(\vartheta)$ eine Minimalwert der Funktion $\mathbb{R} \ni C \mapsto \int_\Theta \big(C-\tau(\vartheta)\big)^2\,\mathrm{d}\nu_x(\vartheta)$ ist. ○

Satz 1.6.1: Existenz + Eindeutigkeit des Bayes-Schätzers

Der Bayes-Schätzer T für τ zur a-priori-Verteilung μ existiert, ist $Q\circ\pi_{\mathfrak{X}}^{-1}$-fast sicher eindeutig und es gilt $T(x) = \int_\Theta \tau(\vartheta)\,\mathrm{d}\nu_x(\vartheta)$ $Q\circ\pi_{\mathfrak{X}}^{-1}$-fast sicher.

Beweis. 1 Wir zeigen erst, dass $\int_\Theta \tau(\vartheta)\,\mathrm{d}\nu_x(\vartheta)$ für $Q\circ\pi_{\mathfrak{X}}^{-1}$-fast alle $x\in\mathfrak{X}$ definiert ist und messbar in x und reellwertig und endlich ist.

Es gilt $Q\circ\pi_{\mathfrak{X}}^{-1}$-fast sicher

$$\int_\Theta \tau^2(\vartheta)\,\mathrm{d}\nu_x(\vartheta) = \int_\Theta \tau^2(\vartheta)\frac{\rho(x,\vartheta)}{Z_x}\,\mathrm{d}\mu(\vartheta)$$

und somit ist dieser Ausdruck nach dem Satz von FUBINI messbar in x (aber eventuell ∞).

Integration über $\mathfrak{X}$ mit $Z_x\,\mathrm{d}\mu_0(x)$ ergibt

$$\int_{\mathfrak{X}\times\Theta} \tau^2(\vartheta)\,\mathrm{d}Q(x,\vartheta) = \int_\Theta \tau^2(\vartheta)\,\mathrm{d}\mu(\vartheta) < \infty.$$

Also gilt $\int_\Theta \tau^2(\vartheta)\,\mathrm{d}\mu(\vartheta) < \infty$ für $Q\circ\pi_{\mathfrak{X}}^{-1}$-fast alle $x\in\mathfrak{X}$.

Damit gilt

$$T^2(x) = \left(\int_\Theta \tau(\vartheta)\,\mathrm{d}\nu_x(\vartheta)\right)^2 \overset{(J)}{\leqslant} \int_\Theta \tau(\vartheta)^2\,\mathrm{d}\nu_x(\vartheta) < \infty$$

für $Q\circ\pi_{\mathfrak{X}}^{-1}$-fast alle $x\in\mathfrak{X}$.

Weiter ist T messbar (wie in voriger Zeile). Nach Integration über $\mathfrak{X}$ mit $Z_x\,\mathrm{d}\mu_0(x)$ folgt

$$\int_{\mathfrak{X}} T^2(x)Z_x\,\mathrm{d}\mu_0(x) \leqslant \int_{\mathfrak{X}}\int_\Theta \tau^2(\vartheta)\,\mathrm{d}\nu_x(\vartheta)Z_x\,\mathrm{d}\mu_0(x) = \int_\Theta \tau^2(\vartheta)\,\mathrm{d}\mu(\vartheta) < \infty$$

und

$$\int_{\mathfrak{X}} T^2(x)Z_x\,\mathrm{d}Q(x,\vartheta) = \int_\Theta \tau^2(\vartheta)\,\mathrm{d}\mu(\vartheta) < \infty.$$

Insbesondere gilt $\mathbb{F}_T < \infty$.

2 Sei nun S ein Schätzer für τ. Wir zeigen $\mathbb{F}_T \leqslant \mathbb{F}_S$. Ohne Beschränkung der Allgemeinheit sei $\mathbb{F}_S < \infty$.

Es gilt

$$\mathbb{F}_S - \mathbb{F}_T = \int_{\mathfrak{X}}\int_\Theta \left(S^2 - 2S\tau - T^2 + 2T\tau\right)\mathrm{d}Q(x,\vartheta).$$

Nebenrechnung: es gilt

$$\int_{\mathfrak{X}}\int_\Theta T\tau\,\mathrm{d}Q(x,\vartheta) = \int_{\mathfrak{X}} T(x)\underbrace{\left(\int_\Theta \tau(\vartheta)\underbrace{\frac{\rho(x,\vartheta)}{Z_x}\,\mathrm{d}\mu(\vartheta)}_{=\mathrm{d}\nu_x(\vartheta)}\right)}_{=T(x)} Z_x\,\mathrm{d}\mu_0(x)$$

$$= \int_{\mathfrak{X}\times\Theta} T^2(x)\,\mathrm{d}Q(x,\vartheta)$$

und analog

$$\int_{\mathfrak{X}}\int_\Theta S\tau\,\mathrm{d}Q(x,\vartheta) = \int_{\mathfrak{X}\times\Theta} S(x)T(x)\,\mathrm{d}Q(x,\vartheta).$$

Einsetzen ergibt

$$\begin{aligned}
\mathbb{F}_S - \mathbb{F}_T &= \int_{\mathfrak{X}} \int_{\Theta} S^2 - 2ST - T^2 + 2T^2 \,\mathrm{d}Q(x,\vartheta) \\
&= \int_{\mathfrak{X}} \int_{\Theta} S^2 - 2ST + T^2 \,\mathrm{d}Q(x,\vartheta) \\
&= \int_{\mathfrak{X}} \int_{\Theta} (S-T)^2 \,\mathrm{d}Q(x,\vartheta) = \int_{\mathfrak{X}} \int_{\Theta} (S-T)^2 Z_x \,\mathrm{d}\mu_0(x) \geqslant 0
\end{aligned}$$

und somit $\mathbb{F}_T \leqslant \mathbb{F}_S$, somit ist T ein BAYES-Schätzer.

Wenn auch S ein BAYES-Schätzer ist, dann gilt $\mathbb{F}_S = \mathbb{F}_T$. Gleichheit gilt genau dann, wenn $Q \circ \pi_X^{-1}$-fast sicher $S = T$ gilt. □

Beispiel 1.6.13 (BAYES-Schätzer des E-Werts einer Normalverteilung)
Dieses Beispiel folgt [2, 7.39].

Sei $\left(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), (\mathcal{N}_{\vartheta,v}^{\otimes n})_{\vartheta\in\mathbb{R}}\right)$ das n-fache GAUSSsche Produktmodell mit bekannter Varianz $v > 0$. Dann ist (bezüglich des n-dimensionalen LEBESGUEmaßes)

$$\rho(x,\vartheta) = (2\pi v)^{-\frac{n}{2}} \exp\left(-\frac{1}{2v}\sum_{k=1}^{n}(x_k-\vartheta)^2\right).$$

Sei die a-priori-Verteilung $\mu = \mathcal{N}_{m,u}$ mit $m \in \mathbb{R}$ und $u > 0$. Je kleiner wir u wählen, desto sicherer sind wir uns, dass der Erwartungswert nahe bei m liegt.

Wir bestimmen den BAYES-Schätzer für $\tau(\vartheta) := \vartheta$.

Wir schreiben im Folgenden c_x, c_x' und c_x'' für von ϑ unabhängig (aber von x abhängige) Konstanten. Weil μ absolutstetig bezüglich des Lebesgue-Maßes λ ist, gilt

$$\begin{aligned}
\frac{\mathrm{d}\nu_x}{\mathrm{d}\lambda}(\vartheta) &= \frac{\mathrm{d}\nu_x}{\mathrm{d}\mu}(\vartheta)\frac{\mathrm{d}\mu}{\mathrm{d}\lambda}(\vartheta) = \frac{\rho(x,\vartheta)}{Z_x}\frac{1}{\sqrt{2\pi u}}\exp\left(-\frac{(\vartheta-m)^2}{2u}\right) \\
&= c_x \exp\left(-\frac{1}{2v}\sum_{k=1}^{n}(x_k-\vartheta)^2 - \frac{(\vartheta-m)^2}{2u}\right).
\end{aligned} \tag{31}$$

Durch Ausmultiplizieren erhalten wir

$$\begin{aligned}
-\frac{1}{2v}\sum_{k=1}^{n}(x_k-\vartheta)^2 - \frac{(\vartheta-m)^2}{2u} &= -\frac{1}{2v}\sum_{k=1}^{n}(x_k^2 - 2x_k\vartheta + \vartheta^2) - \frac{(\vartheta^2 - 2\vartheta m + m^2)}{2u} \\
&= -\frac{1}{2}\left(\frac{1}{u} + \frac{n}{v}\right)\vartheta^2 + \left(\frac{m}{u} + \frac{1}{v}\sum_{k=1}^{n} x_k\right)\vartheta + c_x'.
\end{aligned}$$

Quadratische Ergänzung liefert dann

$$-\frac{1}{2v}\sum_{k=1}^{n}(x_k-\vartheta)^2 - \frac{(\vartheta-m)^2}{2u} = -\frac{\frac{1}{u}+\frac{n}{v}}{2}\left(\vartheta - \left(\frac{\frac{1}{u}m + \frac{1}{v}\sum_{k=1}^{n} x_k}{\frac{1}{u}+\frac{n}{v}}\right)\right)^2 + c_x''$$

und durch Einsetzen in Gleichung (31) folgt

$$\frac{\mathrm{d}\nu_x}{\mathrm{d}\lambda}(\vartheta) = \tilde{c}_x \exp\left[-\frac{\frac{1}{u}+\frac{n}{v}}{2}\left(\vartheta - \left(\frac{\frac{1}{u}m + \frac{1}{v}\sum_{k=1}^{n} x_k}{\frac{1}{u}+\frac{n}{v}}\right)\right)^2\right].$$

Diese Dichte ist bis auf einen konstanten Faktor gleich der Dichte von $\mathcal{N}\left(\frac{\frac{1}{u}m+\frac{1}{v}\sum_{k=1}^n x_k}{\frac{1}{u}+\frac{n}{v}}, \frac{1}{\frac{1}{u}+\frac{n}{v}}\right)$, also gilt

$$\nu_x = \mathcal{N}\left(\frac{\frac{1}{u}m+\frac{1}{v}\sum_{k=1}^n x_k}{\frac{1}{u}+\frac{n}{v}}, \frac{1}{\frac{1}{u}+\frac{n}{v}}\right).$$

Aus Satz 1.6.1 folgt schließlich

$$T(x) = \int_\Theta \tau(\vartheta)\,\mathrm{d}\nu_x(\vartheta) = \int_\Theta \vartheta\,\mathrm{d}\nu_x(\vartheta) = \frac{\frac{1}{u}m+\frac{1}{v}\sum_{k=1}^n x_k}{\frac{1}{u}+\frac{n}{v}} = sm + (1-s)\frac{1}{n}\sum_{k=1}^n x_k,$$

wobei $s = \frac{\frac{1}{u}}{\frac{1}{u}+\frac{n}{v}}$, also ist $T(x)$ eine Konvexkombination aus der Voreinschätzung m und dem empirischen Mittelwert der Beobachtungen. Je mehr Beobachtungen wir machen, desto kleiner wird s, da es sinnvoll ist, weniger auf die Voreinschätzung und mehr auf die Beobachtungen zu vertrauen. ◇

02.06.2022

Bemerkung 1.6.14 In diesem Fall ist T auch ein MAP-Schätzer. ○

Definition 1.6.15 (MAP-Schätzer)
Ein Schätzer T heißt Maximum-a-posteriori-Schätzer wenn

Maximum-a-posteriori-Schätzer

$$\frac{\mathrm{d}\nu_x}{\mathrm{d}\lambda}(T(x)) = \sup_{\vartheta\in\Theta}\frac{\mathrm{d}\nu_x}{\mathrm{d}\lambda}(\vartheta),$$

wobei λ das Lebesgue-Maß ist.

Beispiel 1.6.16 (Bayes-Ansätze in der Bildrestauration)
Seien S und I endliche Mengen und S eine Teilmenge der Pixel und I die Menge der Farben oder Graustufen sowie $\Theta = I^S$ die Menge aller möglichen Bilder. Für $x \in \Theta$ und $y \in \mathfrak{X} = \Theta$ sei $P(x,y)$ die Wahrscheinlichkeit y zu beobachten, wenn das Bild x ist. Hier ist es unsinnig, beispielsweise Maximum-Likelihood-Schätzer zu verwenden; ein Bayesscher Ansatz ist notwendig, da nicht alle Bilder gleichwahrscheinlich sind.

Ansatz. Sei μ eine a-priori-Verteilung auf Θ mit $\mu(\{x\}) > 0$ für alle $x \in \Theta$. Dann gilt (Achtung: x und y haben hier die Rollen getauscht)

$$\nu_y(x) = \frac{1}{Z_y}\mu(x)P(x,y) \qquad \text{mit } Z_y = \sum_{z\in\Theta}\mu(\{z\})P(z,y).$$

Mögliche Wahl von x nach Beobachtung y ist der MAP-Schätzer, der $\mu(\{x\})P(x,y)$ maximiert. (Ist μ die Gleichverteilung auf Θ, so ist der MAP-Schätzer identisch zu dem Maximum-Likelihood-Schätzer.)

Fragen. 1. Wie wählt man μ? 2. Wie findet man das Maximum von $\mu(\{x\})P(x,y)$?

Zu 1.: wir wählen $\mu(\{x\}) = \frac{1}{Z}\exp(-H(x))$, wobei H die „Energie von x“ ist. Diese Darstellung existiert immer, da wir $\mu(\{x\}) > 0$ vorausgesetzt haben. Das Maß μ heißt **Gibbs-Verteilung** zur Energiefunktion H, zum Beispiel, z.B. im Fall, dass I eine endliche Teilmenge reeller Zahlen ist. Eine mögliche Wahl ist

$$H(x) = \beta \sum_{(s,t)\text{ benachbarte Pixel}} (x_s - x_t)^2,$$

wobei $\beta > 0$ die inverse Temperatur ist. Dann folgt

$$\nu_y(x) = \frac{1}{Z_y} \exp\big(-H(x)\ln\big(P(x,y)\big)\big).$$

Es verbleibt nun die Maximierung des Exponenten. ◇

Beispiel 1.6.17 ($\mathbb{P}_\vartheta \sim \mathrm{Poi}(\vartheta)$, $\mu \sim \Gamma_{a,r} \implies \nu_x \sim \Gamma_{a+n,r+\sum_{k=1}^n x_k}$) 01.06.2022
Betrachte $\big(\mathbb{N}_0^n, \mathbb{P}(\mathbb{N}_0^n), (\mathrm{Poi}(\vartheta)^{\otimes n})_{\vartheta>0}\big)$ mit der a-priori-Verteilung $\mu \sim \Gamma_{a,r}$ für $a, r > 0$. Für $x \in \mathbb{N}_0$ hat die a-posteriori-Verteilung ν_x die folgende Dichte bezüglich des Lebesgue-Maßes

$$\begin{aligned}\frac{\rho(x,\vartheta)f_{a,r}(\vartheta)}{\int_0^\infty \rho(x,\theta)f_{a,r}(\theta)\,\mathrm{d}\theta} &= \frac{\left(\prod_{k=1}^n \cancel{\frac{1}{x!}}\vartheta^{x_k}e^{-\vartheta}\right)\cancel{\frac{a^r}{\Gamma(r)}}\vartheta^{r-1}e^{-a\vartheta}}{\int_0^\infty \left(\prod_{k=1}^n \cancel{\frac{1}{x!}}\theta^{x_k}e^{-\theta}\right)\cancel{\frac{a^r}{\Gamma(r)}}\theta^{r-1}e^{-a\theta}\,\mathrm{d}\theta}\\ &= \frac{\vartheta^{\sum_{k=1}^n x_k+r-1}e^{-(a+n)\vartheta}}{\int_0^\infty \theta^{\sum_{k=1}^n x_k+r-1}e^{-(a+n)\theta}\,\mathrm{d}\theta}\\ &= \frac{(a+n)^{r+\sum_{k=1}^n x_k}}{\Gamma\left(r+\sum_{k=1}^n x_k\right)}\vartheta^{\sum_{k=1}^n x_k+r-1}e^{-(a+n)\vartheta},\end{aligned}$$

welche die Dichte der $\Gamma_{a+n,r+\sum_{k=1}^n x_k}$-Verteilung ist.

Der Bayes-Schätzer ist $T_B^{(n)}(x) = \int_0^\infty \vartheta\,\mathrm{d}\nu_x(\vartheta) = \frac{r+\sum_{k=1}^n x_k}{a+n}$ genau der Erwartungswert von ν_x.

Nullsetzen der Ableitung nach ϑ der Dichte von ν_x ergibt

$$\frac{(a+n)^{r+\sum_{k=1}^n x_k}}{\Gamma\left(r+\sum_{k=1}^n x_k\right)}\left(r+\sum_{k=1}^n x_k - 1 - \vartheta(a+n)\right)\vartheta^{r+\sum_{k=1}^n x_k-2}e^{-(a+n)\vartheta} \overset{!}{=} 0$$

also ist der Maximum-a-posteriori-Schätzer $T_{\mathrm{MAP}}^{(n)}(x) = \frac{1}{a+n}\left(r+\sum_{k=1}^n x_k - 1\right)$.

Die Schätzfolgen $\big(T_{\mathrm{MAP}}^{(n)}\big)_{n\in\mathbb{N}}$ und $\big(T_B^{(n)}\big)_{n\in\mathbb{N}}$ sind asymptotisch erwartungstreu, jedoch sind die einzelnen Schätzer nicht erwartungstreu: für $n \in \mathbb{N}$ und $\vartheta \in \Theta$ gilt

$$\mathbb{E}_\vartheta\left[T_B^{(n)}\right] = \frac{1}{a+n}\left(\mathbb{E}\left[\sum_{k=1}^n X_k\right]+r\right) = \frac{n\vartheta+r}{a+n} \xrightarrow{n\to\infty} \vartheta$$

und analog

$$\mathbb{E}_\vartheta\left[T_{\mathrm{MAP}}^{(n)}\right] = \frac{n\vartheta+r-1}{a+n} \xrightarrow{n\to\infty} \vartheta.$$

Ferner folgt

$$\mathbb{V}_\vartheta\left[T_B^{(n)}\right] = \frac{1}{(a+n)^2}\mathbb{V}_\vartheta\left[\sum_{k=1}^n X_k + r\right] = \frac{1}{(a+n)^2}\sum_{k=1}^n \mathbb{V}_\vartheta[X_k] = \frac{n}{(a+n)^2}\vartheta \xrightarrow{n\to\infty} 0$$

und analog

$$\mathbb{V}_\vartheta\left[T_{\mathrm{MAP}}^{(n)}\right] = \frac{n}{(a+n)^2}\vartheta \xrightarrow{n\to\infty} 0.$$

Somit sind beide Schätzfolgen konsistent. ◇

2 Konfidenzbereiche

2.1 Grundlagen

Wir wollen wissen, wie genau $\tau\colon \Theta \to \Sigma$ von einem Schätzer $T\colon \mathfrak{X} \to \Sigma$ geschätzt wird. 02.06.2022

Definition 2.1.1 (Konfidenzbereich)
Seien $\left(\mathfrak{X}, \mathcal{F}, (\mathbb{P}_\vartheta)_{\vartheta\in\Theta}\right)$ ein statistisches Modell, $\tau\colon \Theta \to \Sigma$ eine Kenngröße und $\alpha \in (0,1)$. Eine Abbildung $C\colon \mathfrak{X} \to \mathcal{P}(\Sigma)$ ist ein Konfidenzbereich für τ zum Irrtumsniveau α (oder zum Sicherheitsniveau $1-\alpha$), wenn

Konfidenzbereich

$$\inf_{\vartheta\in\Theta} \mathbb{P}_\vartheta[x \in \mathfrak{X} : \tau(\vartheta) \in C(x)] \geqslant 1-\alpha,$$

wobei wir annehmen, dass $\{x \in \mathfrak{X} : s \in C(x)\} \in \mathcal{F}$ für alle $s \in \Sigma$ gilt.

Ist $\Sigma \subset \mathbb{R}$, und $C(x)$ ein Intervall für jedes $x \in \mathfrak{X}$, so ist C ein Konfidenzintervall.

Bemerkung 2.1.2 Die Funktion C heißt auch Bereichsschätzer im Gegensatz zum Punktschätzer T. ○

Bemerkung 2.1.3 Man kann immer $C(x) = \Sigma$ wählen, aber das ist uninteressant, stattdessen will man $C(x)$ möglichst „klein" wählen. ○

Bemerkung 2.1.4 Es ist *nicht* verboten, $C(x) = \varnothing$ für $x \in \mathfrak{X}$ zu wählen (sogar auf nicht-Nullmengen). ○

Bemerkung 2.1.5 Typische Werte für α sind $0,05$ oder $0,1$. ○

Warnung. Hat man ein konkretes $x_0 \in \mathfrak{X}$ beobachtet, so sollte man *nicht* sagen „$\tau(\vartheta)$ liegt mit der Mindestwahrscheinlichkeit $1-\alpha$ in $C(x_0)$".

Beispiel 2.1.6 (Verschobene Exponentialverteilung)
Betrachte das statistische Modell $\left(\mathbb{R}, \mathcal{B}(\mathbb{R}), (\mathbb{P}_\vartheta)_{\vartheta\in\mathbb{R}}\right)$, wobei $\mathbb{P}_\vartheta$ die Dichte

$$\rho(x,\vartheta) = e^{-x+\vartheta}\,\mathbb{1}_{[\vartheta,\infty)}(x)$$

bezüglich des Lebesgue-Maßes hat.

- Konfidenzintervalle der Form $(-\infty, b]$. Wir wollen ein $b > 0$ finden, sodass

  $$\mathbb{P}_\vartheta\big[x \in \mathbb{R} : \vartheta \in (-\infty, x-b]\big] \geqslant 1-\alpha$$

  gilt.

  Für $\vartheta \in \mathbb{R}$ gilt

  $$\begin{aligned}\mathbb{P}_\vartheta\big[x \in \mathbb{R} : \vartheta \in (-\infty, x-b]\big] &= \mathbb{P}_\vartheta[x - b \geqslant \vartheta] = \mathbb{P}_\vartheta[x \geqslant \vartheta + b] \\ &= \int_{\vartheta+b}^{\infty} e^{-x+\vartheta}\,\mathbb{1}_{[\vartheta,\infty)}(x)\,\mathrm{d}x = \int_b^\infty e^{-y}\,\mathrm{d}y = e^{-b}.\end{aligned}$$

  Wir wählen deshalb $b := -\ln(1-\alpha)$.

- Konfidenzintervall der Form $[a, x]$.

Wir wollen ein $a > 0$ finden, sodass

$$\mathbb{P}_\vartheta\big[x : \vartheta \in [x-a, x]\big] \geqslant 1-\alpha$$

gilt.

Für $\vartheta \in \mathbb{R}$ gilt

$$\begin{aligned}\mathbb{P}_\vartheta\big[x : \vartheta \in [x-a, x]\big] &= \mathbb{P}_\vartheta\big[x - a \leqslant \vartheta \leqslant x\big] = \mathbb{P}_\vartheta\big[\vartheta \leqslant x \leqslant \vartheta + a\big] \\ &= \int_\vartheta^{\vartheta+a} e^{-x+\vartheta}\, \mathbb{1}_{[\vartheta,\infty)}(x)\,\mathrm{d}x = \int_0^a e^{-x}\,\mathrm{d}x = 1 - e^{-a},\end{aligned}$$

also wählen wir $a := -\ln(\alpha)$.

- **Konfidenzintervall minimaler Länge.** Wir wollen $a, b \geqslant 0$ finden, sodass $\mathbb{P}_\vartheta\big[x : \vartheta \in [x-a, x-b]\big] \geqslant 1-\alpha$ und $a-b$ minimal ist.

  Für $\vartheta \in \mathbb{R}$ gilt

$$\begin{aligned}\mathbb{P}_\vartheta\big[x : \vartheta \in [x-a, x-b]\big] &= \mathbb{P}_\vartheta\big[x : x-a \leqslant \vartheta \leqslant x-b]\big] \\ &= \mathbb{P}_\vartheta\big[x : \vartheta + b \leqslant x \leqslant \vartheta + a]\big] \\ &= \int_{\vartheta+b}^{\vartheta+a} e^{-x+\vartheta}\, \mathbb{1}_{[\vartheta,\infty)}(x)\,\mathrm{d}x = \int_b^a e^{-y}\,\mathrm{d}y = e^{-b} - e^{-a}.\end{aligned}$$

  Daher setzen wir $1-\alpha = e^{-b} - e^{-a}$ und erhalten $a = -\ln(e^{-b} - 1 + \alpha)$ unter der Bedingung $e^{-b} > 1 - \alpha$.

  Die Länge des Konfidenzintervalls ist

$$\ell(b) := a - b = -\ln(e^{-b} - 1 + \alpha) - b.$$

  Für $b < -\ln(1-\alpha)$ gilt

$$\ell'(b) = \frac{e^{-b}}{\underbrace{e^{-b} - 1 + \alpha}_{>1}} - 1 > 0.$$

  Also ist die Länge minimal für $b = 0$ und das Konfidenzintervall $[x + \ln(\alpha), x]$. ◇

Beispiel 2.1.7

Gegeben sei das statistische Modell $\big(\mathbb{R}, \mathcal{B}(\mathbb{R}), (\mathbb{P}_\vartheta)_{\vartheta\in\mathbb{R}}\big)$, wobei für die festen Parameter $a, b > 0$ die Dichte des Wahrscheinlichkeitsmaßes $\mathbb{P}_\vartheta$

$$\rho(\vartheta, x) = \frac{ab}{a+b}\Big(e^{-a|x-\vartheta|}\, \mathbb{1}_{\{x<\vartheta\}}(x,\vartheta) + e^{-b|x-\vartheta|}\, \mathbb{1}_{\{x\geqslant\vartheta\}}(x,\vartheta)\Big)$$

ist. Es ist Hausaufgabe 7.2, ein Konfidenzintervall für ϑ zum Niveau $\alpha \in (0,1)$ mit minimaler Länge zu konstruieren. ◇

2.2 Konstruktionsmethode im Standardmodell mit $\tau = \mathrm{id}$

Seien $\big(\mathfrak{X}, \mathcal{F}, (\mathbb{P}_\vartheta)_{\vartheta\in\Theta}\big)$ ein Standardmodell mit Likelihoodfunktion ρ, $\tau := \mathrm{id}_\Theta$ und $\alpha \in (0,1)$. Für $\vartheta \in \Theta$ setzen wir $C_\vartheta := \{x \in \mathfrak{X} : \rho(x,\vartheta) \geqslant c_\vartheta\}$, wobei die „kritische Größe“ $c_\vartheta \geqslant 0$ so gewählt ist, dass $\mathbb{P}_\vartheta[C_\vartheta] \geqslant 1-\alpha$. (Es gilt $C_\vartheta \in \mathcal{F}$ als Urbild der messbaren Abbildung $\rho(\cdot,\vartheta)$.)

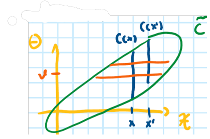


Abb. 7: Die Idee ist, für alle ϑ die Schnitte zu konstruieren und damit die x-Schnitte $C(x)$ zu erhalten.

Definiere nun $\tilde{C} := \{(x,\vartheta) : x \in C_\vartheta\}$ und dann

$$C(x) := \{\vartheta \in \Theta : x \in C_\vartheta\} = \{\vartheta \in \Theta : (x,\vartheta) \in \tilde{C}\}.$$

Dann gilt

$$\mathbb{P}_\vartheta[x \in \mathfrak{X} : \vartheta \in C(x)] = \mathbb{P}_\vartheta[x \in \mathfrak{X} : (x,\vartheta) \in \tilde{C}] = \mathbb{P}_\vartheta[C_\vartheta] \geqslant 1-\alpha$$

für alle $\vartheta \in \Theta$. Somit ist C ein Konfidenzbereich.

Beispiel 2.2.1 (Schadstoffausstoß von Kraftwerken [2, Bsp. 8.3])
Von $N := 10$ Kraftwerken werden bei $n := 4$ zufällig ausgewählten der Schadstoffausstoß gemessen. Gesucht ist ein Konfidenzbereich für die Anzahl ϑ der Kraftwerke, mit zu hohem Ausstoß. Wir wählen $\Theta := \{0,\dots,N\}$ und $\mathfrak{X} := \{0,\dots,n\}$ sowie $\alpha := \frac{1}{5}$. Die Likelihoodfunktion ist

$$\rho(x,\vartheta) = \frac{\binom{\vartheta}{x}\binom{N-\vartheta}{n-x}}{\binom{N}{n}} \sim \mathrm{Hyp}_{N,\vartheta,n}\,.$$

ϑ \ x	0	1	2	3	4
6	1	24	90	80	15
5	5	50	100	50	5
4	15	80	90	24	1
3	35	105	63	7	0
2	70	112	28	0	0
1	126	84	0	0	0
0	210	0	0	0	0

Tabelle 1: Die Tabelle wird symmetrisch ($\rho(x,\vartheta) = \rho(n-x, N-\vartheta)$) fortgeführt, sie zeigt die Werte $\binom{N}{n}\rho(x,\vartheta)$ für $x \in \mathfrak{X}$ und $\vartheta \in \Theta$.

Um C_ϑ zu erhalten, summieren wir in absteigender Reihenfolge der Größe nach die Einträge der ϑ-Zeile, bis $(1-\alpha)\binom{N}{n} = \frac{4}{5}\cdot 210 = 168$ erreicht wird, z.B. ist für $\vartheta = 6$ der größte Wert in der Zeile, 90, kleiner als 168, aber $90 + 80 > 168$, also markieren wir diese beiden Einträge blau. Also ist $C(x)$ die Menge aller $\vartheta \in \Theta$, sodass $\rho(x,\vartheta)$ blau markiert ist. Somit sind $C(0) = \{0,1,2\}$, $C(1) = \{1,\dots,5\}$, $C(2) = \{3,\dots,7\}$, $C(3) = \{5,\dots,9\}$ und $C(4) = \{8,9,10\}$. ◇

07.06.2022

Beispiel 2.2.2 (Konfidenzintervall für den E-Wert im Gauss-Modell)
Betrachte das Gaussche Produktmodell $\big(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), (\mathcal{N}_{m,v}^{\otimes n})_{m\in\mathbb{R},\ v>0}\big)$ für $n \geqslant 2$.

Gesucht ist ein Konfidenzintervall für den $\tau(\vartheta) := \tau = (m, v)$. Für $\vartheta = (m, v)$ bezeichnen wir $m(\vartheta) = m$.

Wir modifizieren die Konstruktion in Beispiel 2.2.1 dahingehend, was wir für $m \in \mathbb{R}$ wollen wir $C_m \in \mathcal{B}(\mathbb{R}^n)$ wählen mit $\mathbb{P}_\vartheta(C_{m(\vartheta)}) \geqslant 1-\alpha$ (möglichst "="). Dann sei $C(x) := \{m \in \mathbb{R} : x \in C_m\}$ für $x \in \mathbb{R}^n$. Dann gilt

$$\mathbb{P}_\vartheta[\{x \in \mathbb{R}^n : m(\vartheta) \in C(x)\}] = \mathbb{P}_\vartheta[\{x : x \in C_{m(\vartheta)}] \geqslant 1-\alpha.$$

Wir definieren die Statistik $T_m(x) := \big(M(x) - m\big)\sqrt{\frac{n}{V^*(x)}}$ für $m \in \mathbb{R}$ und $x \in \mathbb{R}^n$, wobei $M(x) := \frac{1}{n}\sum_{k=1}^n x_k$ der empirische Mittelwert und $V^*(x) := \frac{1}{n-1}\sum_{k=1}^n \big(x_k - M(x)\big)^2$ die empirische Varianz sind.

Behauptung. Die Verteilung von $T_{m(\vartheta)}(x)$ unter $\mathbb{P}_\vartheta$ hängt nicht von ϑ ab.

Beweis. Sind $X_1, \dots, X_n$ unabhängig $\mathcal{N}_{m,v}$-verteilt, so haben sie die Darstellung $X_k = m + \sqrt{v}Y_k$, wobei $Y_1, \dots, Y_n$ $N(0,1)$-verteilt sind. Es folgt

$$\begin{aligned}
T_m(x_1, \dots, x_n) &= \big(M(x) - m\big)\sqrt{\frac{n}{V^*(x)}} \\
&= \left(\cancel{m} + \frac{\sqrt{v}}{n}\sum_{k=1}^n Y_k - \cancel{m}\right)\sqrt{\frac{n(n-1)}{\sum_{k=1}^n \left(\sqrt{v}Y_k + \cancel{m} - \left(\cancel{m} + \frac{\sqrt{v}}{n}\sum_{k=1}^n Y_k\right)\right)^2}} \\
&= \left(\frac{1}{\sqrt{n}}\sum_{k=1}^n Y_k\right)\sqrt{\frac{n-1}{\sum_{k=1}^n \left(Y_k - \frac{1}{n}\sum_{k=1}^n Y_k\right)^2}} = T_0(Y_1, \dots, Y_n).
\end{aligned}$$

Das heißt die Verteilung Q von $T_{m(\vartheta)}(X)$ unter $\mathbb{P}_\vartheta$ hängt nicht von ϑ ab. □

Q heißt t_{n-1}-Verteilung (oder t-Verteilung mit $n-1$ Freiheitsgraden). Es gilt $Q(A) = Q(-A)$ und Q hat eine Dichte (cf. später), welche auf $[0, \infty)$ fallend ist.

Konstruktion von C_m. Zu $\alpha \in (0,1)$ wähle ein Intervall $I \subset \mathbb{R}$ der Form $[-t, t]$ mit $Q([-t, t]) = 1-\alpha$ (dies können wir nur machen, da wir eine Dichte haben). Sei $C_m := T_m^{-1}([-t, t])$. Dann folgt für $\vartheta \in \Theta := \mathbb{R} \times \mathbb{R}_{>0}$

$$\mathbb{P}_\vartheta[C_{m(\vartheta)}] = \mathbb{P}_\vartheta[|T_{m(\vartheta)}| \leqslant t] = Q([-t, t]) = 1-\alpha.$$

Also ist $C(x) := \{m \in \mathbb{R} : x \in C_m\}$ ein Konfidenzintervall für $m(\vartheta)$ zu α.

Für $x \in \mathbb{R}^n$ gilt

$$\begin{aligned}
C(x) &= \{m \in \mathbb{R} : x \in C_m\} = \{m \in \mathbb{R} : |T_m(x)| \leqslant t\} \\
&= \left[M(x) - t\sqrt{\frac{V^*(x)}{n}}, M(x) + t\sqrt{\frac{V^*(x)}{n}}\right].
\end{aligned}$$

◇

Beispiel 2.2.3 (Konfidenzbereiche für die Normalverteilung) 08.06.2022
Wir nehmen n unabhängige Messwerte, welche normalverteilt mit unbekannten Erwartungswert n sind.

1 Wir sind sicher, dass wir die Varianz $v > 0$ kennen. Welchen Konfidenzbereich für m zum Niveau $\alpha \in (0,1)$ wählen wir?

2 Die Varianz ist doch unbekannt. Wie groß ist das Niveau für den in a) gewählten Bereich?

1 Wir betrachten das Stichprobenmittel $M := \frac{1}{n}\sum_{k=1}^n X_k$ und transformieren es zu T_m, sodass die Verteilung von T_m nicht von m oder v abhängt.

Wir wählen $T_m := \sqrt{\frac{n}{v}}(M-m) \sim \mathcal{N}_{0,1}$. Sei $q_{\frac{\alpha}{2}}$ das $\frac{\alpha}{2}$-Fraktil der Standardnormalverteilung (das heißt $\mathcal{N}_{0,1}\left([q_{\frac{\alpha}{2}},\infty)\right) = \frac{\alpha}{2}$. Dann gilt

$$\mathbb{P}_m\left(T_m \in \left[-q_{\frac{\alpha}{2}}, q_{\frac{\alpha}{2}}\right]\right) = 1-\alpha$$

und

$$\begin{aligned}
1-\alpha &= \mathbb{P}_m\left(-q_{\frac{\alpha}{2}} \leqslant \sqrt{\frac{n}{v}}(M-m) \leqslant -q_{\frac{\alpha}{2}}\right) \\
&\leqslant \mathbb{P}_m\left(M - \sqrt{\frac{v}{n}}q_{\frac{\alpha}{2}} \leqslant m \leqslant M + \sqrt{\frac{v}{n}}q_{\frac{\alpha}{2}}\right) \\
&= \mathbb{P}_m\left(m \in \left[M - \sqrt{\frac{v}{n}}q_{\frac{\alpha}{2}}, M + \sqrt{\frac{v}{n}}q_{\frac{\alpha}{2}}\right]\right).
\end{aligned}$$

Wir wählen den Konfidenzbereich

$$C_\alpha(x) = \left[M - \sqrt{\frac{v}{n}}q_{\frac{\alpha}{2}}, M + \sqrt{\frac{v}{n}}q_{\frac{\alpha}{2}}\right].$$

2 Bei unbekannter Varianz können wir T_m nicht wie in a) wählen. In den Vorlesung wurde $T_m := \sqrt{\frac{n}{V^*}}(M-m)$ gewählt, wobei V^* die korrigierte Stichprobenvarianz ist. Es gilt $T_m \sim t_{n-1}$. Sei $t_{\frac{\alpha}{2}}$ das $\frac{\alpha}{2}$-Fraktil der t_{n-1}-Verteilung. Dann ist

$$C_b(x) = \left[M_n - \sqrt{\frac{v^*}{n}}t_{\frac{\alpha}{2}}, M_n - \sqrt{\frac{v^*}{n}}t_{\frac{\alpha}{2}}\right]$$

ein Konfidenzinterval zum Nivea α.

Für den Konfidenzbereich aus 1 gilt dann

$$\begin{aligned}
C_\alpha(x) &= \left[M_n - \sqrt{\frac{v}{n}}q_{\frac{\alpha}{2}}, M_n + \sqrt{\frac{v}{n}}q_{\frac{\alpha}{2}}\right] \\
&= \left[M_n - \left(\sqrt{\frac{v}{v^*}}q_{\frac{\alpha}{2}}\right)\sqrt{\frac{v^*}{n}}, M_n + \left(\sqrt{\frac{v}{v^*}}q_{\frac{\alpha}{2}}\right)\sqrt{\frac{v^*}{n}}\right]
\end{aligned}$$

und C_α ist ein Konfidenzbereich zum Niveau

$$\tilde{\alpha} = 2t_{n-1}\left(\left[q_{\frac{\alpha}{2}}\sqrt{\frac{v}{v^*}}, \infty\right)\right).$$

◇

Beispiel 2.2.4 (Schlafmittel)
Es sollen zwei Schlafmittel miteinander verglichen werden. Es werden $n := 10$ Patienten beide Schlafmittel in unterschiedlichen Nächten gegeben und die Differenz der Schlafdauer gemessen. Alle Patienten mit den zweiten Schlafmittel haben mindestens genau so lange

Patient	1	2	3	4	5	6	7	8	9	10
Differenz	1.2	2.4	1.3	1.3	0.0	1.0	1.8	0.8	4.6	1.4

geschlafen wie mit dem ersten. Wir nehmen an, dass die Messwerte normalverteilt sind. Das Stichprobenmittel und die korrigierte Stichprobenvarianz sind

$$M = \frac{1}{10} \sum_{k=1}^{10} X_k = 1.58 \quad \text{und} \quad V^* = \frac{1}{9} \sum_{k=1}^{10} (X_k - M)^2 = 1.513.$$

Mit dem Irrtumsniveau $\alpha = 0.05$ erhalten wir für die t_9-Verteilung das $\frac{\alpha}{2}$-Fraktil $t = 2.262$. Ein Konfidenzbereich zum Nivea α ist

$$C(x) = \left[M - t\sqrt{\frac{V^*}{10}}, M + t\sqrt{\frac{V^*}{10}} \right] = [1.58 - 0.88, 1.58 + 0.88] = [0.7, 2.46].$$

◇

Beispiel 2.2.5 (Konfidenzbereich im Binomialmodell)
Seien $\mathfrak{X} := \{0, \ldots, n\}$, $\Theta := (0, 1)$ sowie $\mathbb{P}_\vartheta := B_{n,\vartheta}$.

1. Methode: Markov-Ungleichung. Der beste Schätzer für ϑ ist $T(x) := \frac{x}{n}$. Eine natürliche Wahl ist $\left[\frac{x}{n} - \varepsilon, \frac{x}{n} + \varepsilon\right]$ für ein geeignetes $\varepsilon > 0$. Mit der Markov-Ungleichung folgt

$$\mathbb{P}_\vartheta \left(\vartheta \notin C(x)\right) = \mathbb{P}_\vartheta \left(\left| \frac{x}{n} - \vartheta \right| < \varepsilon \right) \leqslant \frac{1}{\varepsilon^2} \mathbb{V}_\vartheta \left[\frac{x}{n} \right] = \frac{\vartheta(1-\vartheta)}{n\varepsilon^2} \leqslant \frac{1}{4n\varepsilon^2} \overset{!}{\leqslant} \alpha$$

für alle $\vartheta \in (0, 1)$. Wir wählen $\varepsilon = \frac{1}{2\sqrt{n\alpha}}$. Für $n = 1000$ und $\alpha = 0.025$ ist also $\varepsilon = 0.1$.

2. Methode: Normalapproximation. Wir nehmen an, dass n so groß ist, dass wir den zentralen Grenzwertsatz anwenden können. Wie zuvor wählen wir $C(x) := \left[\frac{x}{n} - \varepsilon, \frac{x}{n} + \varepsilon\right]$. Dann gilt

$$\begin{aligned}
\mathbb{P}_\vartheta[\vartheta \in C(x)] = \mathbb{P}_\vartheta \left(\left| \frac{x}{n} - \vartheta \right| < \varepsilon \right) &= \mathbb{P}_\vartheta \left(\left| \frac{x - n\vartheta}{\sqrt{n\vartheta(1-\vartheta)}} \right| < \varepsilon \sqrt{\frac{n}{\vartheta(1-\vartheta)}} \right) \\
&\approx \Phi \left(\varepsilon \sqrt{\frac{n}{\vartheta(1-\vartheta)}} \right) - \Phi \left(-\varepsilon \sqrt{\frac{n}{\vartheta(1-\vartheta)}} \right) \\
&= 2\Phi \left(\varepsilon \sqrt{\frac{n}{\vartheta(1-\vartheta)}} \right) - 1 \geqslant 2\Phi\left(2\varepsilon\sqrt{n}\right) - 1,
\end{aligned}$$

wobei Φ die Verteilungsfunktion der Standardnormalverteilung ist.

Sei $\alpha = 0.025$. Wir wählen einen Sicherheitszuschlag von 0.02 für den Approximationsfehler und erhalten

$$2\Phi\left(2\varepsilon\sqrt{n}\right) - 1 \geqslant 0.975 + 0.02 \iff \varepsilon \geqslant \frac{1}{2\sqrt{n}} \Phi^{-1}(0.995) = \frac{1}{2\sqrt{n}} \cdot 2.575.$$

Wählen wir $n = 1000$, erhalten wir $\varepsilon \geqslant 0.0407$.

Verglichen mit der ersten Methoden ist das Konfidenzintervall hier nur halb so groß. ◇

3 Rund um die Normalverteilung: χ^2, F- und t-Verteilungen

Literatur zur mehrdimensionalen Normalverteilung: [4, 7.12-7.19].

Definition 3.0.1 (Gammaverteilung)
Für $\alpha, r > 0$ ist das Wahrscheinlichkeitsmaß $\Gamma_{\alpha,r}$ auf $\big((0,\infty), \mathcal{B}\left((0,\infty)\right)\big)$ mit der Dichte

$$\gamma_{\alpha,r}(x) := \frac{\alpha^r}{\Gamma(r)} x^{r-1} e^{-\alpha x}, \qquad x > 0,$$

die Gammaverteilung mit Skalenparameter α und Formparameter r, wobei

$$\Gamma(r) := \int_0^\infty y^{r-1} e^{-y}\,\mathrm{d}y.$$

Gammaverteilung

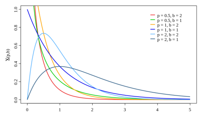


Abb. 8: Die Dichte $\gamma_{a,r}$ für verschiedene $a, r > 0$. Von Sinner1 - Eigenes Werk, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=3546723.

Die Gammaverteilung mit $r = 1$ ist die Exponentialverteilung.

Definition 3.0.2 (Betaverteilung)
Für $a, b > 0$ ist das Wahrscheinlichkeitsmaß $\mathscr{B}_{a,b}$ auf $\big((0,1), \mathcal{B}\left((0,1)\right)\big)$ mit der Dichte

$$\beta_{a,b}(s) := \frac{1}{B(a,b)} s^{a-1}(1-s)^{b-1}, \qquad s \in (0,1),$$

die Betaverteilung zu a, b, wobei

$$B(a,b) := \int_0^1 s^{a-1}(1-s)^{b-1}\,\mathrm{d}s.$$

Betaverteilung

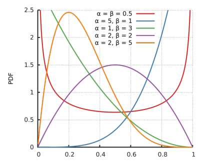


Abb. 9: Die Dichte $\beta a, b$ für verschiedene $a, b > 0$.

Lemma 3.0.3 ($X \sim \mathcal{N}(0,1) \implies X^2 \sim \Gamma_{\frac{1}{2},\frac{1}{2}}$)
Ist X $\mathcal{N}(0,1)$-verteilt, so ist X^2 $\Gamma_{\frac{1}{2},\frac{1}{2}}$-verteilt.

Beweis. Seien $\varphi_{0,1}(x) := \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$ für $x \in \mathbb{R}$ die Dichte von X und $T\colon (0,\infty) \to (0,\infty)$, $x \mapsto x^2$ ein Diffeomorphismus. Also hat $|X|$ auf $(0,\infty)$ die Dichte $2\varphi_{0,1}$. Nach dem Transformationssatz hat $X^2 = T(|X|)$ die Dichte

$$\begin{aligned}\rho_T(y) &= 2\varphi_{0,1}\big(T^{-1}(y)\big)\left|\frac{\mathrm{d}}{\mathrm{d}y}T^{-1}(y)\right| = 2\varphi_{0,1}(\sqrt{y})\left|\frac{\mathrm{d}}{\mathrm{d}y}\sqrt{y}\right| \\ &= 2\varphi_{0,1}(\sqrt{y})\left|\frac{1}{2\sqrt{y}}\right| = \frac{1}{\sqrt{y}}\varphi_{0,1}(\sqrt{y}) = \frac{1}{\sqrt{2\pi}} y^{-\frac{1}{2}} e^{-\frac{1}{2}y},\end{aligned}$$

und das ist die Dichte einer $\Gamma_{\frac{1}{2},\frac{1}{2}}$-Verteilung. □

Lemma 3.0.4 ($X \sim \Gamma_{\alpha,r}$, $Y \sim \Gamma_{\alpha,s} \implies X + Y \sim \Gamma_{\alpha,r+s}$, $\frac{X}{X+Y} \sim \mathscr{B}_{r,s}$)
Sind $\alpha, r, s > 0$ sowie X und Y unabhängig mit den Verteilungen $\Gamma_{\alpha,r}$ und $\Gamma_{\alpha,s}$, so sind $X+Y$ und $\frac{X}{X+Y}$ unabhängig mit den Verteilung $\Gamma_{\alpha,r+s}$ resp. $\mathscr{B}_{r,s}$.

Beweis. Die gemeinsame Verteilung von X und Y auf $(0,\infty)^2$ hat die Dichte

$$\rho(x,y) = \gamma_{\alpha,r}(x)\gamma_{\alpha,s}(y) = \frac{\alpha^{r+s}}{\Gamma(r)\Gamma(s)} x^{r-1} y^{s-1} e^{-\alpha(x+y)}.$$

Wir definieren die Transformation $T(x,y) := \left(x+y, \frac{x}{x+y}\right)$ für $x,y>0$. Dann ist T ein Diffeomorphismus von $(0,\infty)^2$ nach $(0,\infty)\times(0,1)$, denn die Abbildung ist offensichtlich glatt und das Gleichungssystem $x+y=u$ und $\frac{x}{x+y}=v$ für $u\in(0,\infty)$ und $v\in(0,1)$ eindeutig lösbar: $x=uv$ und $y=(u(1-v)$ und somit $T^{-1}(u,v) = \big(uv, u(1-v)\big)$ (offensichtlich glatt).

Es gilt

$$\mathrm{D}T^{-1}(u,v) = \begin{pmatrix} v & u \\ 1-v & -u \end{pmatrix}$$

und somit $\det\big(DT^{-1}(u,v)\big) = -uv-(1-v)u = -u$.

Also ist die Dichte von $\left(X+Y, \frac{X}{X+Y}\right) = T(X,Y)$

$$\begin{aligned}
\rho(u,v) &= \rho(T^{-1}(u,v))\left|\det\left(\mathrm{D}T^{-1}(u,v)\right)\right| = \rho\big(uv, u(1-v)\big)u \\
&= \frac{\alpha^{r+s}}{\Gamma(r)\Gamma(s)}(uv)^{r-1}\big(u(1-v)\big)^{s-1} e^{-\alpha(uv+u(1-v))}u \\
&= \frac{\alpha^{r+s}}{\Gamma(r)} u^{r+s-1} e^{-\alpha u} \frac{1}{\Gamma(s)} v^{r-1}(1-v)^{s-1} = \gamma_{\alpha,r+s}(u)\beta_{r,s}(v).
\end{aligned}$$

Somit sind $X+Y$ und $\frac{X}{X+Y}$ unabhängig verteilt mit den entsprechenden Verteilungen. □

Korollar 3.0.5
Für $\alpha, r, s > 0$ gilt $\gamma_{\alpha,r} * \gamma_{\alpha,s} = \gamma_{\alpha,r+s}$.

Korollar 3.0.6 (Summe quadrierter Normalverteilungen)
Sind $X_1,\dots,X_n$ unabhängig $\mathcal{N}(0,1)$-verteilt, so ist $\sum_{k=1}^n X_k^2$ $\Gamma\left(\frac{1}{2},\frac{n}{2}\right)$-verteilt.

DEFINITION 3.0.7 (χ^2-VERTEILUNG)
Die χ_n^2-Verteilung (mit n *Freiheiten*) ist die $\Gamma\left(\frac{1}{2},\frac{n}{2}\right)$-Verteilung.

χ_n^2-Verteilung

Bemerkung 3.0.8 Die χ_n^2-Verteilung hat die Dichte $\frac{1}{\Gamma\left(\frac{n}{2}\right)2^{\frac{n}{2}}} x^{\frac{n}{2}-1} e^{-\frac{1}{2}x}$ für $x>0$. ○

09.06.2022

SATZ 3.0.1: FISHER-VERTEILUNG

Seien $X_1,\dots,X_m$ sowie $Y_1,\dots,Y_n$ unabhängig $\mathcal{N}(0,1)$-verteilt. Dann hat die Zufallsvariable $F_{m,n} := \frac{1}{m}\sum_{k=1}^m X_k^2 / \frac{1}{n}\sum_{j=1}^n Y_j^2$ die Dichte

$$f_{m,n}(x) := \frac{m^{\frac{m}{2}} n^{\frac{n}{2}}}{B\left(\frac{m}{2},\frac{n}{2}\right)} \frac{x^{\frac{m}{2}-1}}{(n+mx)^{\frac{m+n}{2}}}, \qquad x>0.$$

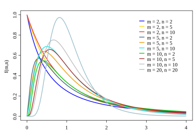


Abb. 10: Die Dichte der FISHER-Verteilung für verschiedene $m,n\in\mathbb{N}$.

Beweis. Nach Korollar 3.0.6 ist die Zufallsvariable $X := \sum_{k=1}^m X_k^2$ $\Gamma_{\frac{1}{2},\frac{m}{2}}$-verteilt und die Zufallsvariable $Y := \sum_{j=1}^n Y_j^2$ $\Gamma_{\frac{1}{2},\frac{n}{2}}$-verteilt. Weiter sind X und Y unabhängig. Nach Lemma 3.0.4 ist $Z := \frac{X}{X+Y}$ $\mathcal{B}_{\frac{m}{2},\frac{n}{2}}$-verteilt. Es ist

$$F_{m,n} = \frac{n}{m}\frac{X}{Y} = \frac{n}{m}\frac{Z}{1-Z} =: T(Z),$$

wobei $T\colon (0,1) \to (0,\infty)$, $z \mapsto \frac{n}{m}\frac{z}{1-z}$ (strikt monoton wachsend und somit) ein Diffeomorphismus mit $T^{-1}(y) = \frac{my}{n+my}$ für $y > 0$ ist. Also hat $F_{m,n}$ die Dichte von

$$\beta_{\frac{m}{2},\frac{n}{2}}\left(T^{-1}(y)\right) \cdot \frac{mn}{(n+my)^2} = f_{m,n}(y), \qquad y > 0,$$

wobei $\frac{mn}{(n+my)^2}$ die Funktionaldeterminante ist. □

Definition 3.0.9 (Fisher-Verteilung)
Die Verteilung $\mathcal{F}_{m,n}$ von $F_{m,n}$ heißt Fisher-Verteilung mit m und n Freiheitsgraden.

Fisher-Verteilung

Bemerkung 3.0.10 Es gilt $\mathcal{F}_{m,n} = \mathcal{B}_{m,n} \circ T^{-1}$ (Fisher-Verteilung ist Bildmaß der Beta-Verteilung unter T), wobei T wie oben ist. ○

Korollar 3.0.11
Seien $X, Y_1, \dots, Y_n$ unabhängig $\mathcal{N}(0,1)$-verteilt. Dann hat die Zufallsvariable $Z := \frac{X}{\sqrt{\frac{1}{n}\sum_{j=1}^n Y_j^2}}$ die Dichte

$$\tau_n(x) = \frac{1}{\sqrt{n}B\left(\frac{1}{2},\frac{n}{2}\right)}\left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{2}}, \qquad x \in \mathbb{R}.$$

Beweis. Nach Satz 3.0.1 hat Z^2 die Dichte $f_{1,n}(x)$ für $x > 0$. Nach dem Transformationssatz (mit $\Phi\colon (0,\infty) \to (0,\infty)$, $z \mapsto \sqrt{z}$ und $(\Phi^{-1})'(z) = 2z$) hat $|Z|$ die Dichte $f_{1,n}(y^2) \cdot 2y$ für $y > 0$. Aufgrund der Symmetrie hat Z die Dichte $f_{1,n}(y^2)|y|$ für $y \in \mathbb{R}$. □

Definition 3.0.12 (Student t-Verteilung)
Das Wahrscheinlichkeitsmaß t_n auf $\left(\mathbb{R}, \mathcal{B}(\mathbb{R})\right)$ mit Dichte τ_n heißt Student t-Verteilung mit n Freiheitsgraden.

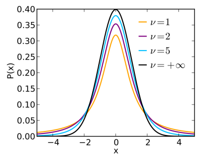


Abb. 11: Die Dichte der t_ν-Verteilung für verschiedene ν. Von Skbkekas - Eigenes Werk, CC BY 3.0, https://commons.wikimedia.org/w/index.php?curid=9546828.

Bemerkung 3.0.13 Im Moment ist unklar, ob die Verteilung mit der in Kapitel zwei definierten übereinstimmt. ○

Bemerkung 3.0.14 Die Funktion $\tau_1(x) := \frac{1}{\pi}\frac{1}{1+x^2}$ für $x \in \mathbb{R}$ ist die Dichte der Standard-Cauchy-Verteilung. ○

Bemerkung 3.0.15 ($\lim_{n\to\infty} t_n$) Für $n \to \infty$ konvergiert die t_n-Verteilung schwach gegen $\mathcal{N}(0,1)$, was aus Anwendung des schwachen Gesetzes der großen Zahlen auf den Nenner oder aus der gleichmäßigen Konvergenz der Dichten folgt. ○

Satz 3.0.2: Student (1908) [2, Satz 9.17]

Im Modell $\left(\mathbb{R}^n, \mathcal{B}(\mathbb{R})^n, (\mathcal{N}_{m,v}^{\otimes n})_{m\in\mathbb{R},v>0}\right)$ für $n \geqslant 2$ gilt für $\vartheta := (m,v) \in \mathbb{R}\times(0,\infty) =: \Theta$ bezüglich $\mathbb{P}_\vartheta := \mathcal{N}_{m,v}^{\otimes n}$ für den empirischen Mittelwert $M(X) := \frac{1}{n}\sum_{k=1}^n X_k$ sowie die korrigierte empirische Varianz $V^*(X) := \frac{1}{n-1}\sum_{j=1}^n \left(X_j - M(X)\right)^2$

1. M und V^* sind unabhängig.
2. M ist $\mathcal{N}\left(m, \frac{v}{n}\right)$-verteilt und $\frac{n-1}{v}V^*$ ist χ^2_{n-1}-verteilt.
3. $T_m := \sqrt{\frac{n}{V^*}}(M-m)$ ist t_{n-1}-verteilt (hier ist der Zusammenhang mit Kapitel zwei).

Beweis. Ohne Beschränkung der Allgemeinheit seien $m = 0$ und $v = 1$, andernfalls können wir X_k durch $Y_k := \frac{1}{\sqrt{v}}\big(X_k - m\big)$ ersetzen und beachten, dass sich m in der Definition von V^* weghebt (d.h. $M(Y) = \frac{1}{\sqrt{v}}(M(X) - m)$, $V^*(Y) = \frac{1}{v}V^*(X)$), und dass sich $\sqrt{v}$ in der Definition von T_m wegkürzt (d.h. $T_m(Y) = T_m(X)$).

1. Sei C eine orthogonale $n \times n$-Matrix mit erster Zeile $\frac{1}{\sqrt{n}}(1, \ldots, 1)$. Dann gilt $C \cdot (1, \ldots, 1)^\mathsf{T} = (\sqrt{n}, 0, \ldots, 0)^\mathsf{T}$. Sei nun $Y := CX$, wobei $X \in \mathbb{R}^n$ $\mathcal{N}(0, I_n)$-verteilt ist. Dann ist (cf. [4]) Y auch $\mathcal{N}(0, I_n)$-verteilt. Insbesondere sind die Komponenten $Y_1, \ldots, Y_n$ unabhängig $\mathcal{N}(0,1)$-verteilt. Es gilt $M(X) = \frac{1}{\sqrt{n}}\sum_{k=1}^n \frac{1}{\sqrt{n}}X_k = \frac{1}{\sqrt{n}}Y_1$ und somit

$$\begin{aligned}
(n-1)V^*(X) &= \sum_{k=1}^n \big(X_k - M(X)\big)^2 = \sum_{k=1}^n X_k^2 - 2M(X)X_k + M(X)^2 \\
&= \sum_{k=1}^n X_k^2 + nM(X)^2 - 2M(X)\sum_{k=1}^n X_k = \sum_{k=1}^n X_k^2 - nM(X)^2 \qquad (32)\\
&= \sum_{j=1}^n Y_j^2 - Y_1^2 = \sum_{j=2}^n Y_j^2,
\end{aligned}$$

da $\sum_{k=1}^n X_k^2 = \|X\|_2^2 = \|CX\|_2^2 = \|Y\|_2^2 = \sum_{k=1}^n Y_k^2$ wegen der Orthogonalität von C gilt. Also folgt die Unabhängigkeit.

2. Der erste Teil ist klar und der zweite folgt aus (32).

3. Es gilt
$$\sqrt{\frac{n}{V^*}}(M-m) = \frac{\sqrt{\frac{n}{v}}(M-m)}{\sqrt{\frac{1}{n-1}\frac{n-1}{v}V^*}} \overset{1}{\underset{2}{\sim}} \frac{\mathcal{N}(0,1)}{\sqrt{\frac{1}{n-1}\chi^2_{n-1}}} \overset{\text{Def. 3.0.12}}{=} t_{n-1}.$$
$\square$

4 Hypothesentests

4.1 Grundlagen

Definition 4.1.1 (Test, Nullhypothese, Alternative, Gütefunktion) Seien $\big(\mathfrak{X}, \mathcal{F}, (\mathbb{P}_\vartheta)_{\vartheta\in\Theta}\big)$ ein statistisches Modell und $\Theta = \Theta_0 \sqcup \Theta_1$ mit $\Theta_0 \neq \varnothing \neq \Theta_1$. Wir bezeichnen Θ_0 als Nullhypothese und Θ_1 als Alternative. Nullhypothese

- Eine Statistik $\varphi\colon \mathfrak{X} \to [0,1]$ heißt Test von Θ_0 gegen Θ_1 und nicht-randomisiert, wenn $\varphi(\mathfrak{X}) \subset \{0,1\}$ und sonst randomisiert. Ist φ nicht-randomisiert, heißt die Menge $\{x \in \mathfrak{X} : \varphi(x) = 1\}$ Ablehnungsbereich (oder Verwerfungsbereich) der Nullhypothese. Test Ablehnungsbereich
- Ferner heißt $\sup_{\vartheta\in\Theta_0} \mathbb{E}_\vartheta[\varphi]$ Fehler erster Art von φ und φ heißt (zufälliger) Test zum (Irrtums)niveau $\alpha \in (0,1)$, wenn $\sup_{\vartheta\in\Theta_0} \mathbb{E}_\vartheta[\varphi] \leqslant \alpha$. Fehler erster Art
- Die Gütefunktion des Tests φ ist $G_\varphi\colon \Theta \to [0,1]$, $\vartheta \mapsto \mathbb{E}_\vartheta[\varphi]$. Für $\vartheta \in \Theta_1$ heißt $G_\varphi(\vartheta)$ (die Wahrscheinlichkeit, die Alternative zu akzeptieren) Macht von φ an der Stelle ϑ und $\beta_\varphi(\vartheta) := 1 - G_\varphi(\vartheta)$ Fehler zweiter Art an der Stelle $\vartheta \in \Theta_1$. Gütefunktion Fehler zweiter Art

14.06.2022

Bemerkung 4.1.2 Ist $\varphi(x) = 1$, so heißt das, dass die Alternative akzeptiert wird und $\varphi(x) = 0$ heißt, dass die „Nullhypothese akzeptiert wird“ (besser: die Nullhypothese wird nicht verworfen). Ist $\varphi(x) = 0$ (und der Test noch so gut und n sehr groß) sollte man *nicht* sagen „die Nullhypothese ist mit großer Sicherheit korrekt“.

Ist $\varphi(x) \in (0,1)$, so heißt das, dass man sich mit Wahrscheinlichkeit $\varphi(x)$ für die Alternative entscheidet. ○

In vielen Fällen besteht Θ_0 nur aus einem Element (z.B. „Ein Würfel ist fair“). Oft konstruiert man zu einem gegeben Irrtumsniveau einen Test und wählt dann (wenn möglich) den mit der größten Macht.

Bemerkung 4.1.3 Oft entspricht Θ_0 dem „Normalfall“ und Θ_1 einer ungewöhnlichen Abweichung. ○

Bemerkung 4.1.4 (Tests als Schätzer) Einen nicht-randomisierten Test φ kann man als Schätzer für die Kenngröße $\tau := \mathbb{1}_{\Theta_1}$ auf Θ interpretieren. ○

Ist dieser Schätzer erwatungstreu?

Beispiel 4.1.5 (Ist ein Würfel fair?)
Ein Würfel wir n mal unabhängig geworfen. Wir wollen herausfinden, ob er fair ist. Seien dafür $\mathfrak{X} := \{1,\ldots,6\}^n$, $\mathcal{F} := \mathcal{P}(\mathfrak{X})$ und Θ die Menge aller Wahrscheinlichkeitsmaße auf $\{1,\ldots,6\}$. (Wir identifizieren diese Maße mit ihren Zähldichten.) Seien $\vartheta_0 := \frac{1}{6}(1,\ldots,1)$, und $\Theta_0 := \{\vartheta_0\}$ sowie $\Theta_1 := \Theta\backslash\Theta_0$ und $\mathbb{P}_\vartheta := \vartheta^{\otimes n}$.

Ein „vernünftiger“ Test φ sieht so aus: wir setzen $\varphi(x) = 1$ wenn die relativen Häufigkeiten, mit denen $1,\ldots,6$ auftreten, nicht nahe bei ϑ_0 liegen, und $\varphi(x) = 0$ sonst. ◇

Bemerkung 4.1.6 Man muss immer das Modell sowie α und φ festlegen, *bevor* man Beobachtungen durchführt. ○

Beispiel 4.1.7 (Qualitätskontrolle (cf. [2, (10.1)]))
Es werden $N := 10.000$ Orangen an einen Händler geliefert, welcher die Lieferung nur akzeptiert, wenn höchstens fünf Prozent der Orangen faul sind. Der Händler entnimmt eine Stichprobe mit $n := 50$ Orangen, von welchen $x \in \{0, \ldots, n\}$ faul sind.

Wir wählen $\mathfrak{X} := \{0, \ldots, n\}$, $\Theta := \{0, \ldots, N\}$, $\Theta_0 := \{0, \ldots, \frac{1}{20}N\}$, $\Theta_1 := \{\frac{1}{20}N + 1, \ldots, N\}$ und

$$\mathbb{P}_\vartheta(x) = \frac{\binom{\vartheta}{x}\binom{N-\vartheta}{n-x}}{\binom{N}{n}}.$$

Es ist naheliegend, die Nullhypothese abzulehnen, wenn mehr als $c \in \mathbb{N}_0$ Orangen in der Stichprobe faul sind und sonst nicht abzulehnen.

Der Händler wählt c so, dass der Fehler erster Art höchstens α ist. ◇

Beispiel 4.1.8 (Münzwurf-Glücksspieler)
Einem Glücksspieler wird vorgeworfen beim Münzspiel zu betrügen und anstatt einer fairen Münze eine Münze mit Erfolgswahrscheinlichkeit $p = 0.75$ zu verwenden. Diese Anschuldigung soll nun mit n Münzwürfen überprüft werden. ◇

Allgemein	**Beispiel (Münzwurf-Glücksspieler)**
Statistisches Modell $\big(\mathfrak{X}, \mathcal{F}, (\mathbb{P}_\vartheta)_{\vartheta\in\Theta}\big)$	$\mathfrak{X} = \{0, \ldots, n\}$, $\mathcal{F} = \mathbb{P}(\mathfrak{X})$, $\Theta = \{\frac{1}{2}, \frac{3}{4}\}$, $\mathbb{P}_\vartheta(\{k\}) = \binom{n}{k}\vartheta^k(1-\vartheta)^{n-k}$
Nullhypothese und Alternative: Zerlege $\Theta = \Theta_0 \sqcup \Theta_1$. **Nullhypothese:** $H_0 : \vartheta \in \Theta_0$. Bedeutung: Der Parameter ϑ ist für mich akzeptabel. Wir sind im (gewünschten) Normalfall. **Alternative:** $H_1 : \vartheta \in \Theta_1$. Bedeutung: Der Parameter ϑ ist für mich problematisch. Es liegt eine Abweichung vom Normalfall vor.	Wir wählen die Zerlegung nach dem Prinzip „im Zweifel für den Angeklagten“. Der Normalfall ist hier, das kein Betrug begangen wurde, das heißt $\Theta_0 := \{\frac{1}{2}\}$ und $\Theta_1 := \{\frac{3}{4}\}$. **Nullhypothese:** $\vartheta = \frac{1}{2}$. Der Spieler hat nicht betrogen und mit fairer Münze gespielt. **Alternative:** $\vartheta = \frac{3}{4}$. Der Spieler hat betrogen.
Fehlerarten. *Fehler erster Art (gravierender / peinlicher Fehler):* Entscheidung für Alternative, obwohl Nullhypothese stimmt. *Fehler zweiter Art:* Entscheidung für Nullhypothese obwohl die Alternative stimmt.	*Fehler erster Art:* Der Spieler wird wegen Betrugs bestraft, obwohl er unschuldig ist. *Fehler zweiter Art:* Der Spieler wird nicht bestraft, obwohl er betrogen hat.
Entscheidungsregel: Test $\varphi\colon \mathfrak{X} \to [0,1]$. Bei Beobachtung $x \in \mathfrak{X}$ entscheidet man nach folgender Regel: ist $\varphi(x) = 0$, wählt man die Nullhypothese, ist $\varphi(x) = 1$, die Alternative und ist $\varphi(x) \in (0,1)$ wählen wir mit Wahrscheinlichkeit $\varphi(x)$ die Alternative.	Wähle einen Schwellenwert $c \in \mathfrak{X}$ und setze $\varphi(x) := \mathbb{1}_{x>c}$. Der Spieler wird nicht bestraft, wenn weniger als c von n Würfen erfolgreich sind, aber bestraft wenn dem nicht so ist. Hierbei hängen c und $\varphi(x)$ vom Irrtumsniveau ab.

Tabelle 2: Aufbau eines Hypothesentests.

Definition 4.1.9 (Bester Test)
Ein Test φ von Θ_0 gegen Θ_1 heißt (gleichmäßig) bester Test (engl.: UMP - *uniformly most powerful*) zum Niveau α, wenn der Fehler erster Art höchstens α ist und für jeden Test ψ mit Niveau α gilt: bester Test

$$G_\varphi(\vartheta) \geqslant G_\psi(\vartheta) \qquad \forall \vartheta \in \Theta_1$$

Definition 4.1.10 (Unverfälschter Test)
Ein Test φ heißt unverfälscht zum Niveau α, wenn $G_\varphi(\vartheta_0) \leqslant \alpha \leqslant G_\varphi(\vartheta_1)$ für alle $\vartheta_0 \in \Theta_0$ und $\vartheta_1 \in \Theta_1$. unverfälscht

Bemerkung 4.1.11 (Beziehungen zwischen Tests und Konfidenzbereichen)
Sei $\left(\mathfrak{X}, \mathcal{F}, (\mathbb{P}_\vartheta)_{\vartheta \in \Theta}\right)$ ein statistisches Modell. Ist $C\colon \mathfrak{X} \to \mathcal{P}(\Theta)$ ein Konfidenzbereich zum Irrtumsniveau $\alpha \in (0,1)$ und $\vartheta_0 \in \Theta$, so ist $\{x \in \mathfrak{X} : \vartheta_0 \notin C(x)\}$ der Ablehnungsbereich eines Test von $H_0 : \vartheta = \vartheta_0$ gegen $H_1 : \vartheta \neq \vartheta_0$ zum Niveau α. Ist umgekehrt für jedes $\vartheta_0 \in \Theta$ ein nicht-randomisierter Test für $H_0 : \vartheta = \vartheta_0$ gegen $H_1 : \vartheta \neq \vartheta_0$ zum Niveau α gegeben, so lässt sich daraus ein Konfidenzbereich zum Irrtumsniveau α gewinnen (HA 9.1, [2, Aufg. 10.1]).○

4.2 Neyman-Pearson-Lemma

Definition 4.2.1 (Neyman-Pearson Test)
Seien $\Theta := \{0,1\}$ mit $\Theta_k := \{k\}$ für $k \in \{0,1\}$ sowie $\mathbb{P}_0$ und $\mathbb{P}_1$ Wahrscheinlichkeitsmaße auf $(\mathfrak{X}, \mathcal{F})$ mit $\mathbb{P}_0 \neq \mathbb{P}_1$. Seien μ ein σ-endliches Maß auf $(\mathfrak{X}, \mathcal{F})$ mit $\mathbb{P}_0, \mathbb{P}_1 \ll \mu$ (z.B. $\mu = \frac{1}{2}(\mathbb{P}_0 + \mathbb{P}_1)$) und ρ die zugehörigen Likelihoodfunktion sowie

$$R(x) := \begin{cases} \frac{\rho(x,1)}{\rho(x,0)}, & \text{wenn } \rho(x,0) > 0, \\ \infty, & \text{sonst} \end{cases}$$

der Likelihood-Quotient.

Likelihood-Quotient

Ein Test φ ist ein Neyman-Pearson Test (NPT), wenn ein $c \in [0,\infty)$ existiert mit

$$\varphi(x) \begin{cases} = 1, & R(x) > c, \\ = 0, & R(x) < c, \\ \in [0,1], & R(x) = c. \end{cases}$$

Satz 4.2.1: Neyman-Pearson Lemma (1932)

Je größer $R(x)$ ist, desto eher wird man sich für die Alternative $\mathbb{P}_1$ entscheiden.

Sei $\alpha \in (0,1)$.
1. Es existiert ein Neyman-Pearson Test φ mit $\mathbb{E}_0[\varphi] = \alpha$.
2. Jeder Neyman-Pearson Test mit $\mathbb{E}_0[\varphi] = \alpha$ ist ein bester Test zum Niveau α.

Beweis. 1. Wähle ein $c \in [0,\infty)$, sodass $\mathbb{P}_0[R(x) > c] \leqslant \alpha$ und $\mathbb{P}_0[R(x) \geqslant c] \geqslant \alpha$. Dann ist c das α-Fraktil von $\mathbb{P}_0 \circ R^{-1}$. Dieses c existiert, denn unter der Nullhypothese ist $\rho(x,0)$ fast sicher positiv, das heißt $\rho(x,0) > 0$ gilt $\mathbb{P}_0$-fast sicher.

Fall 1: $\mathbb{P}_0[R(x) = c] = 0$. Dann folgt $\mathbb{P}_0[R(x) > c] = \alpha$ und für $\varphi := \mathbb{1}_{\{R>c\}}$ ist ein Neyman-Pearson Test.

Fall 2: $\mathbb{P}_0[R(x) = c] > 0$. Dann gilt $\gamma := \frac{\alpha - \mathbb{P}_0[R(x)>c]}{\mathbb{P}_0[R(x)=c]} \in (0,1]$, da $\mathbb{P}_0[R(x) > c] + \mathbb{P}_0[R(x) = c] = \mathbb{P}_0[R(x) \geqslant c] \geqslant \alpha$. Wir definieren

$$\varphi(x) := \begin{cases} 1, & R(x) > c, \\ 0, & R(x) < c, \\ \gamma, & R(x) = c. \end{cases}$$

Dann ist φ ein Neyman-Pearson Test und $\mathbb{E}_0[\varphi] = 1 \cdot \mathbb{P}_0[R(X) > c] + \gamma \cdot \mathbb{P}_0[R(X) = c] = \alpha$.

2. Sei φ ein Neyman-Pearson Test mit $\mathbb{E}_0[\varphi] = \alpha$ mit Schwellenwert $c \in [0,\infty)$ und ψ

ein beliebiger Test zum Niveau α. Dann gilt

$$\begin{aligned}\mathbb{E}_1[\varphi] - \mathbb{E}_1[\psi] &= \int_{\mathfrak{X}} \varphi(x)\,\mathrm{d}\mathbb{P}_1(x) - \int_{\mathfrak{X}} \psi(x)\,\mathrm{d}\mathbb{P}_1(x) \\ &= \int_{\mathfrak{X}} \varphi(x) - \psi(x)\,\mathrm{d}\mathbb{P}_1(x) = \int_{\mathfrak{X}} \big(\varphi(x) - \psi(x)\big)\rho(x,1)\,\mathrm{d}\mu(x).\end{aligned}$$

Gilt $\varphi(x) - \psi(x) > 0$ für ein $x \in \mathfrak{X}$, dann ist $\varphi(x) > 0$ und somit $R(x) \geqslant c$, also $\rho(x,1) \geqslant c\rho(x,0)$ (auch wenn $R(x) = \infty$). Ist $\varphi(x) - \psi(x) < 0$ für $x \in \mathfrak{X}$, dann ist $\varphi(x) < 1$ und somit $R(x) \leqslant c$, also $\rho(x,1) \leqslant c\rho(x,0)$.

Also folgt

$$\begin{aligned}\int_{\mathfrak{X}} \big(\varphi(x) - \psi(x)\big)\rho(x,1)\,\mathrm{d}\mu(x) &\geqslant c \int_{\mathfrak{X}} \big(\varphi(x) - \psi(x)\big)\rho(x,0)\,\mathrm{d}\mu(x) \\ &= \underbrace{c}_{\geqslant 0} \underbrace{\big(\underbrace{\mathbb{E}_0[\varphi]}_{=\alpha} - \underbrace{\mathbb{E}_0[\psi]}_{\leqslant\alpha}\big)}_{\geqslant 0} \geqslant 0.\end{aligned}$$

□

Beispiel 4.2.2 (Münzwurf-Glücksspieler) 15.06.2022

Die Dichte bezüglich des Zählmaßes auf $\{0, \ldots, n\}$ ist $\rho(x,\vartheta) = \binom{n}{x}\vartheta^x(1-\vartheta)^{n-x}$. Dann ist

$$R(x) = \frac{\rho\left(x, \frac{3}{4}\right)}{\rho\left(x, \frac{1}{2}\right)} = \frac{\cancel{\binom{n}{x}} 3^x \frac{1}{4^n}}{\cancel{\binom{n}{x}} \frac{1}{2^n}} = 3^x 2^{-n}.$$

Für $n = 10$ wollen wir den besten Test zum Niveau $\alpha = \frac{1}{100}$ bestimmen.

Für $c \in [0, \infty)$ gilt

$$\mathbb{P}_0[R(x) > c] = \mathbb{P}_0[3^x 2^{-10} > c] = \mathbb{P}_0[3^x > 2^{10}c] = \mathbb{P}_0\left[x > \frac{1}{\ln(3)}\ln(2^{10}c)\right].$$

Wir wählen also c so, dass $\frac{1}{\ln(3)}\ln(2^{10}c)$ ein α-Fraktil von $\mathbb{P}_0$ ist.

Es gilt

$$\mathbb{P}_0(X = k) = \binom{10}{k} 2^{-10}$$

und somit

$$\mathbb{P}_0(X = 10) = \frac{1}{1024} \approx 0.001 \qquad \text{und} \qquad \mathbb{P}_0(X = 9) = \frac{5}{512} \approx 0.010$$

und daher $\mathbb{P}_0(X \in \{9, 10\}) \approx 0.011$. Somit ist 9 das α-Fraktil von $\mathbb{P}_0$ (das heißt $\mathbb{P}_0(X > 9) < \alpha < \mathbb{P}_0(X \geqslant 9)$). Wir wählen $c \in [0, \infty)$, sodass $\frac{1}{\ln(3)}\ln(2^{10}c) = 9$ ist (das heißt $c = 2^{-10}3^9 = \frac{19683}{1024} \approx 19.2$) und

$$\gamma = \frac{\alpha - \mathbb{P}_0(R > c)}{\mathbb{P}_0(R = c)} = \frac{\alpha - \mathbb{P}_0(X > 9)}{\mathbb{P}_0(X = 9)} = \frac{\frac{1}{100} - \frac{5}{512}}{\frac{1}{1024}} = \frac{6}{25}.$$

Somit ist der beste Test

$$\varphi(x) = \begin{cases} 1, & \text{wenn } R(x) > c, \\ \frac{6}{25}, & \text{wenn } R(x) = c, \\ 0, & \text{wenn } R(x) < c, \end{cases} = \begin{cases} 1, & \text{wenn } X > 9, \\ \frac{6}{25}, & \text{wenn } X = 9, \\ 0, & \text{wenn } X < 9, \end{cases}$$

◇

16.06.2022

Satz 4.2.2: Lemma von Stein (1952)

Im unendlichen Produktmodell $\left(\mathfrak{X}, \mathcal{F}, (\mathbb{P}_\vartheta)_{\vartheta\in\{0,1\}}\right) = \left(E^{\mathbb{N}}, \mathfrak{E}^{\otimes n}, (Q_\vartheta^{\otimes n})_{\vartheta\in\{0,1\}}\right)$ mit $Q_0 \neq Q_1$ betrachten wir für jedes $n \in \mathbb{N}$ den Neyman-Pearson Test φ_n mit $\mathbb{E}_0[\varphi_n] = \alpha$, wobei $\alpha \in (0,1)$ fest ist. Dann gilt

$$\lim_{n\to\infty} \frac{1}{n} \ln\left(1 - \mathbb{E}_1[\varphi_n]\right) = -H(Q_0, Q_1)$$

Insbesondere konvergiert $\mathbb{E}_1[\varphi_n]$ exponentiell gegen 1.

Beweis. Siehe [2, 10.4]. □

Wir sehen nun, das die obigen Aussage nicht mehr richtig ist, wenn wir zweiseitige Testprobleme betrachten (und dafür brauchen wir mindestens drei Wahrscheinlichkeitsmaße).

Beispiel 4.2.3 (Münzwurf)
Seien $\Theta = \left\{\frac{1}{4}, \frac{1}{2}, \frac{3}{4}\right\}$, $\mathfrak{X} = \{0,1\}$ und $\mathbb{P}_\vartheta(\{1\}) = \vartheta$ sowie $\mathbb{P}_\vartheta(\{0\}) = 1 - \vartheta$.

Fall (a): $\Theta_0 = \{\frac{1}{2}\}$ **und** $\Theta_1 := \left\{\frac{1}{4}, \frac{3}{4}\right\}$**.** Wir zeigen, dass es *keinen* besten Test zum Niveau $\alpha \in (0,1)$ gibt. Sei $\varphi\colon \mathfrak{X} \to [0,1]$ ein Test zum Niveau α. Dann gilt $\alpha \geqslant \mathbb{E}_0[\varphi] = \frac{1}{2}\varphi(0) + \frac{1}{2}\varphi(1)$. Ohne Beschränkung der Allgemeinheit sei $\varphi(1) \geqslant \varphi(0)$. Wenn φ ein bester Test zum Niveau α ist, dann wähle $\psi\colon \mathfrak{X} \to [0,1]$ mit $\psi(0) > \psi(1)$ und $\psi(0) + \psi(1) = 2\alpha$. Es folgt

$$G_\psi\left(\frac{1}{4}\right) = \mathbb{E}_{\frac{1}{4}}[\psi] = \frac{1}{4}\psi(1) + \frac{3}{4}\psi(0) = \underbrace{\frac{1}{4}\left(\psi(1) + \psi(0)\right)}_{=\frac{\alpha}{2}} + \underbrace{\frac{1}{2}\psi(0)}_{>\frac{\alpha}{2}} > \alpha$$

sowie

$$G_\varphi\left(\frac{1}{4}\right) = \mathbb{E}_{\frac{1}{4}}[\varphi] = \frac{1}{4}\varphi(1) + \frac{3}{4}\varphi(0) = \underbrace{\frac{1}{4}\left(\varphi(1) + \varphi(0)\right)}_{\leqslant\frac{\alpha}{2}} + \underbrace{\frac{1}{2}\varphi(0)}_{\leqslant\frac{\alpha}{2}} \leqslant \alpha.$$

Es folgt $G_\varphi(\frac{1}{4}) < G_\psi(\frac{1}{4})$, es gibt also keinen besten Test.

Fall (b): $\Theta_0 = \{\frac{1}{4}\}$ **und** $\Theta_1 := \left\{\frac{1}{2}, \frac{3}{4}\right\}$**.** Sei φ ein Test mit $\mathbb{E}_{\frac{1}{4}}[\varphi] = \frac{1}{4}\varphi(1) + \frac{3}{4}\varphi(0) = \alpha$. Wähle $\varphi(1)$ maximal unter den Nebenbedingungen $\frac{1}{4}\varphi(1) + \frac{3}{4}\varphi(0) = \alpha$ und $\varphi(1), \varphi(0) \in [0,1]$.

Wir zeigen, dass φ ein bester Test ist. Für $p \in [\frac{1}{4}, 1]$ (insbesondere für $p \in \{\frac{1}{2}, \frac{3}{4}\}$ gilt für jeden Test ψ zum Niveau α

$$\begin{aligned}\mathbb{E}_p[\varphi] &= p \cdot \varphi(1) + (1-p)\varphi(0) = p\big(4\alpha - 3\varphi(0)\big) + (1-p)\varphi(0) \\ &= 4\alpha p + (1-4p)\varphi(0) \geqslant 4\alpha p + \underbrace{(1-4p)}_{\leqslant 0}\psi(0) \geqslant \mathbb{E}_p[\psi].\end{aligned}$$

Also ist φ ein bester Test. ◇

4.3 Beste einseitige Tests

Definition 4.3.1 (Wachsende Likelihoodquotienten)
Ein Standardmodell $\left(\mathfrak{X}, \mathcal{F}, (\mathbb{P}_\vartheta)_{\vartheta\in\Theta}\right)$ mit $\Theta \subset \mathbb{R}$ hat wachsende Likelihoodquotienten bezüglich einer Statistik $T\colon \mathfrak{X} \to \mathbb{R}$, wenn für alle $\vartheta_1 < \vartheta_2$ und $R_{\vartheta_2:\vartheta_1}(x) := \frac{\rho(x,\vartheta_2)}{\rho(x,\vartheta_1)}$ gilt, dass $R_{\vartheta_2:\vartheta_1}(x) = f_{\vartheta_2:\vartheta_1}(T(x))$ für eine streng wachsende Funktion $f_{\vartheta_2:\vartheta_1}\colon \mathbb{R} \to \mathbb{R}$ gilt.

wachsende Likelihoodquotienten

Beispiel 4.3.2 (Exponentielles Modell hat wachsende Likelihoodquotienten)
Jedes exponentielle Modell bezüglich T hat wachsende Likelihoodquotienten bezüglich T oder $-T$, denn

$$\begin{aligned} R_{\vartheta_2:\vartheta_1}(x) = \frac{\rho(x,\vartheta_2)}{\rho(x,\vartheta_1)} &= \frac{\exp\left(a(\vartheta_2)T(x) - b(\vartheta_2)\right)\cancel{h(x)}}{\exp\left(a(\vartheta_1)T(x) - b(\vartheta_1)\right)\cancel{h(x)}} \\ &= \exp\left(\big(a(\vartheta_2) - a(\vartheta_1)\big)T(x) + b(\vartheta_1) - b(\vartheta_2)\right). \end{aligned}$$

Also ist für $\vartheta_2 > \vartheta_1$ $R_{\vartheta_2:\vartheta_1} = f_{\vartheta_2:\vartheta_1}(T(x))$ mit $y \mapsto f_{\vartheta_2:\vartheta_1}(y)$ streng wachsend, wenn a streng wachsend ist. Ist a hingegen streng fallend, ersetze a durch $-a$ und T durch $-T$. $\diamond$

Satz 4.3.1: Wachsende Likelihoodquotienten $\implies \exists$ bester Test für einseitiges Testproblem

Sei $\left(\mathfrak{X}, \mathcal{F}, (\mathbb{P}_\vartheta)_{\vartheta\in\Theta}\right)$ ein Standardmodell mit $\Theta \subset \mathbb{R}$ und wachsenden Likelihoodquotienten. Dann existiert ein bester Test zum Niveau $\alpha \in (0,1)$ für das einseitige Testproblem $H_0 : \vartheta \leqslant \vartheta_0 \in \Theta$ gegen $H_1 : \vartheta > \vartheta_0$. Dieser hat die Gestalt

$$\varphi(x) = \begin{cases} 1, & \text{wenn } T(x) > c, \\ \gamma, & \text{wenn } T(x) = c, \\ 0, & \text{wenn } T(x) < c, \end{cases}$$

wobei sich c und γ aus der Bedingung $G_\varphi(\vartheta_0) = \alpha$ ergeben. Die Gütefunktion von φ ist monoton wachsend.

Beweis. Sei $\alpha \in (0,1)$. Dann lassen sich c und γ durch $\mathbb{E}_{\vartheta_0}[\varphi] = \mathbb{P}_{\vartheta_0}[T(x) > c] + \gamma\mathbb{P}_{\vartheta_0}(T(x) = c) \stackrel{!}{=} \alpha$ bestimmen.

Sei $\vartheta_1 > \vartheta_0$. Wir zeigen, dass φ ein Neyman-Pearson Test zu ϑ_0 und ϑ_1 zum Niveau α und damit bester Test zwischen ϑ_0 und ϑ_1 ist. Da $R_{\vartheta_1:\vartheta_0}$ eine streng wachsende Funktion von $T(x)$ ist, gilt

$$\varphi(x) = \begin{cases} 1, & \text{wenn } R_{\vartheta_1:\vartheta_0} > f_{\vartheta_1:\vartheta_0}(c), \\ \gamma, & \text{wenn } R_{\vartheta_1:\vartheta_0} = f_{\vartheta_1:\vartheta_0}(c), \\ 0, & \text{wenn } R_{\vartheta_1:\vartheta_0} < f_{\vartheta_1:\vartheta_0}(c), \end{cases}$$

also ist φ ein Neyman-Pearson Test.

Ebenso ist φ für das eben berechnete c und γ ein Neyman-Pearson Test zwischen ϑ und $\overline{\vartheta}$ für $\vartheta < \overline{\vartheta}$ zum Niveau

$$\mathbb{E}_\vartheta[\varphi] = \mathbb{P}[T(x) > c] + \gamma\mathbb{P}_\vartheta[T(x) = c] =: \beta$$

Damit ist φ auch bester Test zwischen ϑ und $\overline{\vartheta}$, also insbesondere besser als der konstante Test $\psi \equiv \beta$ (welcher Niveau β hat). Also gilt

$$G_\varphi(\overline{\vartheta}) \geqslant G_\psi(\overline{\vartheta}) = \beta = G_\varphi(\vartheta)$$

und damit ist $\vartheta \mapsto G_\varphi(\vartheta)$ monoton wachsend. Weiter gilt

$$\sup_{\vartheta \leqslant \vartheta_0} \mathbb{E}_\vartheta[\varphi] = \sup_{\vartheta \leqslant \vartheta_0} G_\varphi(\vartheta) = G_\varphi(\vartheta_0) = \alpha,$$

also hat φ das Niveau α für H_0 gegen H_1.

Sei nun ψ ein weiterer Test von H_0 gegen H_1 zum Niveau α und $\vartheta_1 > \vartheta_0$ beliebig. Dann ist ψ insbesondere ein Test von ϑ_0 gegen ϑ_1 zum Niveau α.
Da φ wie oben gezeigt in dieser Situation bester Test zum Niveau α ist, gilt insbesondere $\mathbb{E}_{\vartheta_1}[\varphi] \geqslant \mathbb{E}_{\vartheta_1}[\psi]$. Da $\vartheta_1 > \vartheta_0$ beliebig war, ist φ sogar bester Test von H_0 gegen H_1 zum Niveau α. □

Bemerkung 4.3.3 Der Satz 4.3.1 gilt analog für $H_0 : \vartheta \geqslant \vartheta_0$ gegen $H_1 : \vartheta < \vartheta_0$. In diesem Fall muss $R_{\vartheta_2:\vartheta_1}$ streng monoton steigend in T für $\vartheta_2 < \vartheta_1$ sein. ○

Beispiel 4.3.4 (Einseitiger Gauss-Test mit bekanntem Erwartungswert)
Sei $\left(\mathbb{R}^n, \mathcal{B}(\mathbb{R})^{\otimes n}, (\mathcal{N}_{m,\vartheta}^{\otimes n})_{\vartheta>0}\right)$ das Gauss-Produktmodell mit bekanntem Mittelwert $m \in \mathbb{R}$. Seien $H_0 : \vartheta \geqslant \vartheta_0$ und $H_1 : \vartheta < \vartheta_0$. Dann ist die Likelihoodfunktion ρ_1 für $n = 1$

$$\rho_1(x,\vartheta) = \frac{1}{\sqrt{2\pi\vartheta}} \exp\left(-\frac{1}{2\vartheta}(x-m)^2\right) = \exp\left(-\frac{1}{2\vartheta}(x-m)^2 - \frac{1}{2}\ln(2\pi\vartheta)\right).$$

Sei $T_1(x) := (x-m)^2$. Dann ist das Modell für $n = 1$ exponentiell bezüglich T_1 und somit ist das Produktmodell ebenfalls exponentiell bezüglich $T(x) := \frac{1}{n}\sum_{k=1}^n T_1(x_k)$. Somit ist Satz 4.3.1 anwendbar und der beste Test φ zum Niveau $\alpha \in (0,1)$ hat den Verwerfungsbereich

$$\left\{\sum_{k=1}^n (x_k - m)^2 < nc\right\},$$

wobei c so ist, dass $\mathbb{P}_{\vartheta_0}\left(\sum_{k=1}^n (X_k-m)^2 < nc\right) \overset{!}{=} \alpha$. Mit $X_k - m = \sqrt{\vartheta_0} Y_k$, wobei $Y_1, \dots, Y_n \sim \mathcal{N}(0,1)$ unabhängig unter $\mathbb{P}$ sind, gilt

$$\mathbb{P}_{\vartheta_0}\left(\sum_{k=1}^n (X_k - m)^2 < nc\right) = \mathbb{P}\left(\sum_{k=1}^n Y_k^2 < \frac{nc}{\vartheta_0}\right) = F\left(\frac{nc}{\vartheta_0}\right),$$

wobei F die Verteilungsfunktion einer χ_n^2-verteilten Zufallsvariable ist. Somit ist $c = n\vartheta_0 F^{-1}(\alpha)$. ◇

4.4 Likelihood-Quotiententest

21.06.2022

Definition 4.4.1 (Likelihood-Quotiententest)
In einem Standardmodell sei

$$R(x) := \frac{\sup_{\vartheta \in \Theta_1} \rho(x, \vartheta)}{\sup_{\vartheta \in \Theta_0} \rho(x, \vartheta)}.$$

Ein Test

$$\varphi(x) = \begin{cases} 1, & \text{wenn } R(x) > a, \\ 0, & \text{wenn } R(x) < a. \end{cases}$$

mit $a \geqslant 0$ heißt Likelihood-Quotiententest (LQ-Test).

Ist R groß, spricht viel für Θ_1, da es dann ein $\vartheta_1 \in \Theta_1$ gibt, sodass $\rho(x, \vartheta_1) \gg \rho(x, \vartheta)$ für alle $\vartheta \in \Theta_0$ gilt.

Bemerkung 4.4.2 LQ-Tests sind nicht immer beste Tests, auch wenn welche existieren (analog zu Maximum-Likelihood-Schätzern, welche nicht immer beste Schätzer sind), aber oft „gut“. ○

4.5 Einseitige Gauß-Tests

22.06.2022

Beispiel 4.5.1 (Einseitiger Gauss-Test mit bekannter Varianz)
Betrachte $\big(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), (\mathcal{N}_{m,v}^{\otimes n})_{m\in\mathbb{R}}\big)$ mit bekannter Varianz $v > 0$. Für $m_0 \in \mathbb{R}$ testen wir $H_0 : m \leqslant m_0$ gegen $H_1 : m > m_0$ zum Niveau $\alpha \in (0,1)$.

Für $m_2 > m_1$ ist der Likelihoodquotient

$$R_{m_2,m_1}(x) = \frac{\rho(x,m_2)}{\rho(x,m_1)} = \frac{\exp\left(-\frac{1}{2v}\sum_{k=1}^n (x_k - m_2)^2\right)}{\exp\left(-\frac{1}{2v}\sum_{k=1}^n (x_k - m_1)^2\right)}$$
$$= \exp\left(\frac{1}{v}(m_2 - m_1)\sum_{k=1}^n x_k - n\frac{m_2^2}{2v} + n\frac{m_1^2}{2v}\right)$$

und somit R eine wachsende Funktion von $M(x) := \frac{1}{n}\sum_{k=1}^n x_k$. Somit hat der beste Test die Form $\varphi(x) = \mathbb{1}_{M(x)>c}$, wobei $c \in \mathbb{R}$ so gewählt ist, dass

$$\alpha = \mathbb{E}_{m_0}[\varphi] = \mathbb{P}_{m_0}\big[M(X) > c\big]$$
$$= \mathbb{P}_{m_0}\left[\frac{M(X) - m_0}{\sqrt{\frac{v}{n}}} > \sqrt{\frac{n}{v}}(c - m_0)\right] = 1 - \Phi\left(\sqrt{\frac{n}{v}}(c - m_0)\right),$$

wobei Φ die Verteilungsfunktion der Standardnormalverteilung ist. Wir lösen nach c auf und erhalten den Test

$$\varphi(x) = \mathbb{1}_{\{M(x) > m_0 + \sqrt{\frac{v}{n}}\Phi^{-1}(1-\alpha)\}}.$$

◇

Beispiel 4.5.2 (Einseitiger Gauss-Test mit bekanntem Erwartungswert)
Betrachte $\big(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), (\mathcal{N}_{m,v}^{\otimes n})_{v>0}\big)$ mit bekanntem Erwartungswert $m \in \mathbb{R}$. Wir haben bereits gezeigt, dass

$$\varphi(x) = \mathbb{1}_{\{\sum_{k=1}^n (X_k - m)^2 < v\chi^2_{n,\alpha}\}},$$

wobei $\chi^2_{n,\alpha}$ das α-Quantil der χ^2_n-Verteilung ist, ein gleichmäßig bester Test zum Niveau α von $H_0 : v \geqslant v_0$ gegen $H_1 : v < v_0$ ist. ◇

Bemerkung 4.5.3 In den vorigen Beispielen war ein Parameter der Normalverteilung bekannt. Bei der Durchführung von Experiment sind aber üblicherweise beide Parameter unbekannt. Hierfür liefert Satz 4.3.1 jedoch keinen besten Test, da nicht $\Theta \subset \mathbb{R}$ gilt. Eine natürliche Modifikation der Tests ist, den unbekannten Parameter, welcher nicht getestet wird, durch seinen erwartungstreuen Schätzer zu ersetzen. Eine weitere Konstruktionsmöglichkeit für diese Tests ist der verallgemeinerte Likelihood-Quotient

$$R(x) = \frac{\sup_{\vartheta\in\Theta_1} \rho(x,\vartheta)}{\sup_{\vartheta\in\Theta_0} \rho(x,\vartheta)}$$

mit dem Likelihood-Quotienten-Test

$$\varphi(x) = \begin{cases} 1, & \text{wenn } R(x) > c, \\ 0, & \text{wenn } R(x) > c. \end{cases}$$

Die Optimalität dieser Tests zu beweisen kann aufwendig sein (cf. [2, 10.4]). Wir beschränken uns auf die Konstruktion und die Aussagen. ○

Beispiel 4.5.4 (χ^2-Test für die Varianz)
Betrachte $\big(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), (\mathcal{N}_{m,v}^{\otimes n})_{m\in\mathbb{R}, v>0}\big)$ und $\Theta_0 := \{(m,v) \in \mathbb{R}\times(0,\infty) : v \leqslant v_0\}$ sowie $\Theta_1 := \{(m,v) \in \mathbb{R}\times(0,\infty) : v > v_0\}$. Der verallgemeinerte Likelihood-Quotient ist

$$R(x) = \frac{\sup_{m\in\mathbb{R}, v>v_0} \rho(x,(m,v))}{\sup_{m\in\mathbb{R}, v\leqslant v_0} \rho(x,(m,v))} = \frac{\sup_{m\in\mathbb{R}, v>v_0} v^{-\frac{n}{2}} \exp\Big(-\sum_{k=1}^n \frac{(X_k-m)^2}{2v}\Big)}{\sup_{m\in\mathbb{R}, v\leqslant v_0} v^{-\frac{n}{2}} \exp\Big(-\sum_{k=1}^n \frac{(X_k-m)^2}{2v}\Big)}.$$

Für $M(x) = \frac{1}{n}\sum_{k=1}^n x_k$ gilt

$$\sup_{m\in\mathbb{R}} \exp\left(-\sum_{k=1}^n \frac{(x_k-m)^2}{2v}\right) = \exp\left(-\sum_{k=1}^n \frac{\big(x_k - M(x)\big)^2}{2v}\right).$$

Wir setzen $V(x) = \frac{1}{n}\sum_{k=1}^n \big(x_k - M(x)\big)^2$. Dann gilt

$$\begin{aligned} R(x) &= \frac{\sup_{v>v_0} v^{-\frac{n}{2}} \exp\Big(-n\frac{V(x)}{2v}\Big)}{\sup_{v\leqslant v_0} v^{-\frac{n}{2}} \exp\Big(-n\frac{V(x)}{2v}\Big)} \\ &= \begin{cases} \exp\Big(\frac{n}{2}\Big(\frac{V(x)}{v_0} - \ln\Big(\frac{V(x)}{v_0}\Big) - 1\Big)\Big), & \text{wenn } V(x) > v_0, \\ \exp\Big(-\frac{n}{2}\Big(\frac{V(x)}{v_0} - \ln\Big(\frac{V(x)}{v_0}\Big) - 1\Big)\Big), & \text{wenn } V(x) \leqslant v_0. \end{cases} \end{aligned}$$

Somit ist $R(x)$ strikt monoton wachsend in $V(x)$ und der Likelihood-Quotiententest hat die Form $\varphi(x) = \mathbb{1}_{\{V(x)>c\}}$. Wir haben bereits gezeigt, dass $\frac{1}{v}V(X) \sim \chi^2_{n-1}$. Es kann gezeigt werden, dass

$$\varphi(X) = \mathbb{1}_{\left\{\sum_{k=1}^n \left(x_k - M(x)\right)^2 > v_0 \chi^2_{n-1;1-\alpha}\right\}},$$

wobei $\chi^2_{n-1,1-\alpha}$ das α-Fraktil der χ^2_{n-1}-Verteilung ist, ein gleichmäßig bester Test ist. $\diamond$

Bemerkung 4.5.5 Für den rechtsseitigen Test mit $\Theta_0 = \{(m,v) \in \mathbb{R}\times(0,\infty) : v \geqslant v_0\}$ und $\Theta_1 := \{(m,v) \in \mathbb{R}\times(0,\infty) : v < v_0\}$ existiert allerdings kein gleichmäßig bester Test. Hierfür muss man sich auf den unverfälschten Test einschränken. $\circ$

Beispiel 4.5.6 (t-Test für den Erwartungswert)
Betrachte $\big(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), (\mathcal{N}_{m,v}^{\otimes n})_{m\in\mathbb{R}, v>0}\big)$ und $\Theta_0 := \{(m,v) \in \mathbb{R}\times(0,\infty) : m \leqslant m_0\}$ sowie $\Theta_1 := \{(m,v) \in \mathbb{R}\times(0,\infty) : m > m_0\}$. Der verallgemeinerte Likelihood-Quotient ist

$$R(x) = \begin{cases} \Big(\frac{V}{\tilde{V}}\Big)^{-\frac{n}{2}}, & \text{wenn } M(x) \leqslant m_0, \\ \Big(\frac{V}{\tilde{V}}\Big)^{\frac{n}{2}}, & \text{wenn } M(x) > m_0, \end{cases}$$

wobei $\tilde{V}(x) := \frac{1}{n}\sum_{k=1}^n (x_k - m)^2$ und $V(x) := \frac{1}{n}\sum_{k=1}^n \big(x_k - M(x)\big)^2$.

Außerdem gilt

$$\frac{\tilde{V}}{V} = 1 + \frac{1}{n-1} T_{m_0}^2 \qquad \text{mit } T_{m_0} = \sqrt{\frac{n}{V^*}}\big(M(x) - m_0\big).$$

Dann ist R strikt monoton wachsend in T_{m_0} und der Likelihood-Quotienten-Test hat die Form

$$\varphi(x) = \mathbb{1}_{\{T_{m_0}(x)>c\}}.$$

Wir wissen bereits, dass T_{m_0} t_{n-1}-verteilt ist. Es kann gezeigt werden, dass

$$\varphi(x) = \mathbb{1}_{\{T_{m_0}(x) > t_{n-1,1-\alpha}\}}$$

der gleichmäßig beste Test zum Niveau α innerhalb der Klasse der unverfälschten Tests ist, wobei $t_{n-1,1-\alpha}$ ist das α-Fraktil der t_{n-1}-Verteilung ist. ◇

5 Asymptotische Tests und Rangtest

5.1 Anpassungstests

Wie testet man, ob ein Würfel fair ist? Wir betrachten allgemeiner $E = \{1, \ldots, s\}$ für $s \geqslant 2$. Dann ist $\Theta := \{\vartheta = \big(\vartheta(k)\big)_{k\in E} \in [0,1]^s : \sum_{k=1}^s \vartheta(k) = 1\}$ die Menge aller Wahrscheinlichkeitsmaße auf E.

Das zugehörige unendliche Produktmodell ist $\big(E^{\otimes \mathbb{N}}, \mathscr{P}(E)^{\otimes \mathbb{N}}, (\vartheta^{\otimes \mathbb{N}})_{\vartheta\in\Theta}\big)$. Für gegebenes $\rho \in \Theta$ mit $\rho(k) > 0$ für alle $k \in E$ (z.B. $\rho = \big(\frac{1}{6}, \ldots, \frac{1}{6}\big)$ beim Würfel) soll die Hypothese $H_0 : \vartheta = \rho$ gegen die Alternative $H_1 : \vartheta \neq \rho$ getestet werden. Seien $\Theta_0 := \{\rho\}$ und $\Theta_1 := \Theta \backslash \{\rho\}$.

Können wir diese Prozedur auch anwenden, wenn $\rho(k) = 0$ gilt?

Seien X_n das (E-wertige) Ergebnis des n-ten Wurfes und $h_n(k) := \#\{j \leqslant n : X_j = k\}$ die absoluten Häufigkeiten und $L_n(k) := \frac{1}{n}h_n(k)$ die relativen Häufigkeiten. Dann ist L_n ein zufälliges Wahrscheinlichkeitsmaß.

Wir betrachten für festes $n \in \mathbb{N}$ den LQ-Test

$$R_n(x) = R_n(x_1, \ldots, x_n) = \frac{\sup_{\vartheta\in\Theta_1} \prod_{k=1}^s \vartheta(k)^{h_n(k)}}{\prod_{k=1}^s \rho(k)^{h_n(k)}} = \frac{\sup_{\vartheta\in\Theta} \prod_{k=1}^s \vartheta(k)^{h_n(k)}}{\prod_{k=1}^s \rho(k)^{h_n(k)}},$$

wobei die letzte Gleichung aus der Stetigkeit des Zählers in ϑ und der Dichtheit $\Theta_1 \subset \Theta$ folgt.

Der Zähler ist gleich $\prod_{k=1}^s L_n(k)^{h_n(k)}$ (das ist leicht zu überprüfen[6], cf. [2, Bsp. 7.7]). Daher gilt

$$R_n(x) = \prod_{k=1}^s \left(\frac{L_n(k)}{\rho(k)}\right)^{h_n(k)}$$

und somit

$$\ln\left(R_n(x)\right) = n\sum_{k=1}^s L_n(k)\ln\left(\frac{L_n(k)}{\rho(k)}\right) = nH(L_n; \rho).$$

Somit hat der LQ-Test die Form

$$\varphi_n = \begin{cases} 1, & \text{wenn } nH(L_n;\rho) > c, \\ 0, & \text{wenn } nH(L_n;\rho) < c. \end{cases}$$

für ein $c = c_n \in \mathbb{R}$.

Wie kann man zu gegebenen $\alpha \in (0,1)$ und $n \in \mathbb{N}$ das c so berechnen, dass φ_n Niveau α hat?

Nota bene. Es ist $(h_n(k))_{k=1}^s \in \{0, \ldots, n\}^s$ unter $\mathbb{P}_\rho$ multinomial verteilt: für $k_1, \ldots, k_s \in \{0, \ldots, n\}$ mit $\sum_{j=1}^s k_j = n$ gilt

$$\mathbb{P}_\rho\big((h_n(j))_{j=1}^s = (k_j)_{j=1}^s\big) = \frac{n!}{k_1! \cdot \ldots \cdot k_s!}\prod_{j=1}^s (\rho(j))^{k_j}.$$

[6]Ist $\vartheta(k) = 0$ für ein $k \in E$, so ist die zu maximierende Funktion gleich Null. Wir können also annehmen, dass stets $\vartheta(k) > 0$ gilt. Wir maximieren stattdessen $\ln\big(\prod_{k=1}^s \vartheta(k)^{h_n(k)}\big) = n\sum_{k=1}^s L_n(k)\ln\big(\vartheta(k)\big)$ unter der Nebenbedingung $\sum_{k=1}^s \vartheta(k) = 1$. Die Lagrange-Funktion ist $\mathcal{L}(\vartheta, \lambda) := \sum_{k=1}^s \big(nL_n(k)\ln\big(\vartheta(k)\big) + \lambda\vartheta(k)\big) - \lambda$. Nullsetzen der ϑ-Ableitung von $\mathcal{L}$ ergibt $\frac{1}{n}\lambda = \frac{L_n(k)}{\vartheta(k)}$ für alle $k \in E$. Also ist $\frac{L_n(k)}{\vartheta(k)}$ konstant, wir nennen diese Konstante $C_n > 0$. Mit der Nebenbedingung folgt $1 = \sum_{k=1}^s \vartheta(k) = \frac{1}{C_n}\sum_{k=1}^s L_n(k) = \frac{1}{C_n}$ und somit $C_n = 1$, also $\vartheta(j) = L_n(j)$.

Damit kann man prinzipiell das c zu α berechnen. Für große n ist dies aber mit sehr viel Aufwand verbunden.

Lemma 5.1.1 (Asymptotik für $nH(L_n;\rho)$)
Sei

$$D_{n,\rho} := \sum_{k=1}^{s} \frac{\big(h_n(k) - n\rho(k)\big)^2}{n\rho(k)} = n \sum_{k=1}^{s} \rho(k) \left(\frac{L_n(k)}{\rho(k)} - 1 \right)^2 .$$

Dann gilt

$$nH(L_n;\rho) - \frac{1}{2} D_{n,\rho} \xrightarrow{n\to\infty} 0 \quad \textit{in } \mathbb{P}_\rho\textit{-Wahrscheinlichkeit.}$$

Beweis. Es gilt

$$\begin{aligned} nH(L_n;\rho) - \frac{1}{2}D_{n,\rho} &= n \sum_{k=1}^{s} \rho(k) \left(\frac{L_n(k)}{\rho(k)} \ln\left(\frac{L_n(k)}{\rho(k)}\right) - \frac{1}{2}\left(\frac{L_n(k)}{\rho(k)} - 1\right)^2 \right) \\ &\overset{(\star)}{=} n \sum_{k=1}^{s} \rho(k) \left(1 - \frac{L_n(k)}{\rho(k)} + \frac{L_n(k)}{\rho(k)} \ln\left(\frac{L_n(k)}{\rho(k)}\right) - \frac{1}{2}\left(\frac{L_n(k)}{\rho(k)} - 1\right)^2 \right) \\ &=: n \sum_{k=1}^{s} \rho(k) g(Y_{k,n}) \end{aligned}$$

wobei wir in $(\star)$ verwenden, dass $\sum_{k=1}^{s} \rho(k) = \sum_{k=1}^{s} L_n(k) = 1$ gilt sowie $Y_{k,n} := \frac{L_n(k)}{\rho(k)}$ und

$$g(x) = 1 - x + x\ln(x) - \frac{1}{2}(x-1)^2.$$

Dann ist g auf $(0,\infty)$ glatt mit $g(0) = \frac{1}{2}$, $g(1) = 0$, $g'(1) = g''(1) = 0$. Also ist

$$g(1+x) = O(|x|^3) \qquad x \searrow 0.$$

Es gilt

$$\begin{aligned} \mathbb{E}_\rho\left(n(Y_{k,n}-1)^2\right) &= n\,\mathbb{E}_\rho\left[\left(\frac{L_n(k)}{\rho(k)} - 1\right)^2\right] = \frac{n}{\rho(k)^2 n^2}\,\mathbb{E}_\rho\left(h_n(k) - n\rho(k)\right)^2 \\ &= \frac{1}{\rho(k)^2 n} n \underbrace{\rho(k)\big(1-\rho(k)\big)}_{=\mathbb{V}\left[\mathrm{Bin}\left(n,\rho(k)\right)\right]} = \frac{1}{\rho(k)}\big(1-\rho(k)\big) \leqslant \frac{1}{\rho(k)}, \end{aligned} \tag{$\star\star$}$$

da $h_n(k) \sim \mathrm{Bin}\big(n,\rho(k)\big)$. Also folgt für $C > 0$ mit der Chebychev-Ungleichung (C)

$$\begin{aligned} \mathbb{P}_\rho\left(|Y_{k,n} - 1| \leqslant \frac{C}{\sqrt{n}}\right) &= 1 - \mathbb{P}_\rho\left(|Y_{k,n}-1| > \frac{C}{\sqrt{n}}\right) \overset{(C)}{\geqslant} 1 - \frac{\mathbb{E}\left[|Y_{k,n}-1|^2\right]}{\left(\frac{C}{\sqrt{n}}\right)^2} \\ &= 1 - \frac{n\,\mathbb{E}\left[|Y_{k,n}-1|^2\right]}{C^2} \overset{(\star\star)}{\geqslant} 1 - \frac{1}{\rho(k)C^2} \xrightarrow{C\to\infty} 1 \end{aligned}$$

gleichmäßig in k und n.

Also gilt auf der Menge $|Y_{k,n} - 1| \leqslant \frac{C}{\sqrt{n}}$ für alle $j \in \{1,\dots,s\}$

$$\left| nH(L_n;\rho) - \frac{1}{2}D_{n,\rho} \right| \leqslant n\tilde{C}\left(\frac{C}{\sqrt{n}}\right)^3 = \tilde{C}C^3\frac{1}{\sqrt{n}} \xrightarrow{n\to\infty} 0.$$

Seien nun $\delta > 0$ und $\varepsilon > 0$ beliebig. Nach oben gezeigter Konvergenz können wir C so wählen, dass $\mathbb{P}_\rho\Big(|Y_{k,n}-1| \leqslant \frac{C}{\sqrt{n}}\Big) > 1-\delta$ für jedes $n \in \mathbb{N}$ ist (da obige Konvergenz unabhängig von n war). Nun wählen wir $N \in \mathbb{N}$ so groß, dass $\tilde{C}C^3\frac{1}{\sqrt{n}} < \varepsilon$ für alle $n > N$. Es gilt nun

$$\begin{aligned}\mathbb{P}_\rho\Big(\Big|nH(L_n;\rho) - \frac{1}{2}D_{n,\rho}\Big| \leqslant \varepsilon\Big) &\geqslant \mathbb{P}_\rho\Big(\Big\{\Big|nH(L_n;\rho) - \frac{1}{2}D_{n,\rho}\Big| \leqslant \varepsilon\Big\} \cap \Big\{|Y_{k,n}-1| \leqslant \frac{C}{\sqrt{n}}\Big\}\Big)\\ &= \mathbb{P}_\rho\Big(|Y_{k,n}-1| \leqslant \frac{C}{\sqrt{n}}\Big) > 1-\delta\end{aligned}$$

und es folgt die Konvergenz in Wahrscheinlichkeit. □

Lemma 5.1.2 (Verteilung von $D_{n,\rho}$ unter H_0)
Die Verteilung von $D_{n,\rho}$ unter H_0 konvergiert für $n \to \infty$ gegen eine χ^2_{s-1}-Verteilung.

Beweis. Der Zufallsvektor $v_n := \Big(\frac{h_n(1)-n\rho(1)}{\sqrt{\rho(1)}}, \dots, \frac{h_n(s)-n\rho(s)}{\sqrt{\rho(s)}}\Big)^\mathsf{T}$ ist von der Form

$$v_n = \sum_{j=1}^n Z_j \quad \text{mit} \quad Z_j := \left(\frac{\mathbb{1}_{\{X_j=1\}} - \rho(1)}{\sqrt{\rho(1)}}, \dots, \frac{\mathbb{1}_{\{X_j=s\}} - \rho(1)}{\sqrt{\rho(s)}}\right).$$

Da die X_j unabhängig identisch verteilt sind, sind es auch die Z_j. Nach dem mehrdimensionalen zentralen Grenzwertsatz konvergiert $\frac{1}{\sqrt{n}}v_n$ in Verteilung (oder: schwach) gegen $\mathcal{N}(0,\Sigma)$, wobei für $k,\ell \in E$

$$\Sigma_{k,\ell} = \operatorname{Cov}\big((Z_1)_k,(Z_1)_\ell\big) = \begin{cases} \mathbb{V}(Z_{1,k}) = \dfrac{1}{\rho(k)}\rho(k)\big(1-\rho(k)\big) = 1-\rho(k), & \text{wenn } k=\ell,\\ = \dfrac{-\rho_k\rho_\ell - \rho_k\rho_\ell + \rho_k\rho_\ell}{\sqrt{\rho(k)\rho(\ell)}} = -\sqrt{\rho(k)\rho(\ell)} & \text{wenn } k \neq \ell \end{cases}$$

Beachte, dass

$$\left\|\frac{1}{\sqrt{n}}V_n\right\|^2 = D_{n,\rho}$$

gilt. Die Matrix Σ ist symmetrisch und positiv semidefinit und somit diagonalisierbar mit nicht-negativen Eigenwerten. Wir zeigen, dass 0 ein Eigenwert zum Eigenvektor $e_s = \sqrt{\rho(1)}\,\mathbb{1}^\mathsf{T}$ ist. Für alle $k \in E$ gilt

$$\sum_{\ell=1}^s \Sigma_{k,\ell}\sqrt{\rho(\ell)} = \Sigma_{k,k}\sqrt{\rho(k)} + \sum_{\substack{\ell=1\\ \ell\neq k}}^s \Sigma_{k,\ell}\sqrt{\rho(\ell)} = \big(1-\rho(k)\big)\sqrt{\rho(k)} - \sum_{\substack{\ell=1\\ \ell\neq k}}^s \sqrt{\rho(k)\rho(\ell)}\sqrt{\rho(\ell)} = 0.$$

23.06.2022

Ferner ist $\lambda = 1$ ein $s-1$-facher Eigenwert von Σ, denn es gilt

$$\Sigma - I_s = \begin{pmatrix} -\rho(1) & -\sqrt{\rho(1)\rho(2)} & \dots & -\sqrt{\rho(1)\rho(s)}\\ -\sqrt{\rho(1)\rho(2)} & -\rho(2) & \dots & \vdots\\ \vdots & \dots & \ddots & \vdots\\ -\sqrt{\rho(1)\rho(s)} & \dots & \dots & -\rho(s)\end{pmatrix}.$$

Der Rang von $\Sigma - I_s$ ist eins; dividiert man die k-te Zeile durch $\sqrt{\rho(k)}$, so ist sie gleich $-(\sqrt{\rho(1)}, \dots, \sqrt{\rho(s)})^\mathsf{T}$.

Somit lässt sich Σ als
$$\Sigma = U \begin{pmatrix} I_{s-1} & 0 \\ 0 & 0 \end{pmatrix} U^\mathsf{T},$$
wobei U orthogonal ist. Die letzte Spalte von U ist e_s.

Der Vektor $T_n := U^\mathsf{T} V_n$ hat den letzten Eintrag 0, da $\sum_{k=1}^s h_n(k) = \sum_{k=1}^s n\rho(k) = n$ gilt.

Weil U orthogonal ist, gilt $\frac{1}{n}\|T_n\|^2 = \frac{1}{n}\|V_n\|^2 = D_{n,\rho}$. Da $\frac{1}{\sqrt{n}}V_n$ in Verteilung gegen $\mathcal{N}(0,\Sigma)$ konvergiert, konvergiert $\frac{1}{\sqrt{n}}T_n$ gegen eine Normalverteilung mit Erwartungsvektor $(0,\dots,0)^\mathsf{T}$ und Kovarianzmatrix
$$\begin{aligned}\mathbb{E}\left[\frac{1}{\sqrt{n}}T_n\left(\frac{1}{\sqrt{n}}T_n\right)^\mathsf{T}\right] &= \mathbb{E}\left[\frac{1}{n}U^\mathsf{T} V_n V_n^\mathsf{T} U\right] = U^\mathsf{T}\,\mathbb{E}\left[\frac{1}{n}V_n V_n^\mathsf{T}\right]U \\ &= U^\mathsf{T}\Sigma U = \operatorname{diag}(1,\dots,1,0).\end{aligned}$$

Somit konvergiert die Verteilung des Vektors $\frac{1}{\sqrt{n}}\tilde{T}_n \in \mathbb{R}^{s-1}$ (das Tilde bedeutet, dass wir den letzten Eintrag weglassen) gegen $\mathcal{N}(0, I_{s-1})$.

Also konvergiert $D_{n,\rho} = \frac{1}{n}\|T_n\|^2 = \frac{1}{n}\|\tilde{T}_n\|^2$ in Verteilung gegen eine χ^2_{s-1}-Verteilung. □

Bemerkung 5.1.3 Damit kann man für vorgegebenes Niveau $\alpha \in (0,1)$ die Konstante c aus dem LQ-Test approximativ so bestimmen, dass das Niveau ungefähr gleich α ist. ○

Bemerkung 5.1.4 Der obige Test heißt χ^2-Anpassungstest (*engl:* goodness of fit). ○

χ^2-Anpassungstest

Beispiel 5.1.5 (Urnenexperiment mit Zurücklegen / Mendels Erbsen) Betrachte $E := \{1,\dots,s\}$ und den Beobachtungsraum $X := E^n$. Dies modelliert s verschiedene Kugeln in einer Urne und n Ziehungen. Wir wählen
$$\Theta := \left\{\vartheta = (\vartheta(j))_{j=1}^s \in [0,1]^s : \sum_{j=1}^s \vartheta(j) = 1\right\},$$
die Menge aller möglichen Verteilungen für den ersten Zug, wobei $\vartheta(k)$ die Wahrscheinlichkeit darstellt, die k-te Kugel zu ziehen.

Wir haben eine Vermutung $\rho \in \Theta$ für die Verteilung und wollen $H_0 : \vartheta = \rho$ gegen $H_1 : \vartheta \neq \rho$ testen. Für $k \in \{1,\dots,s\}$, zählen wir, wie oft wir die k-te Kugel gezogen haben; $h_n(k) := |\{j \in \{1,\dots,n\} : X_j = k\}|$. Wir definieren
$$D_{n,\rho} := \sum_{k=1}^s \frac{\big(h_n(k) - n\rho(k)\big)^2}{n\rho(k)}.$$

Dann ist $\varphi := \mathbb{1}_{\{D_{n,\rho} > c\}}$ für ein $c \in \mathbb{R}$ der χ^2-Anpassungstest. Für große n wählen wir $c := \chi^2_{s-1,1-\alpha}$, also das $(1-\alpha)$-Fraktil der χ^2-Verteilung mit $s-1$ Freiheitsgraden.

Betrachte zum Beispiel Mendels Erbsen, ein Experiment zur Vererbungslehrer, wobei die Form (rund (A) oder kantig (a)) und die Farbe (gelb (B) oder grün (b)) von Erbsen beobachtet wird.

Die Nullhypothese ist, dass rund und gelb *dominante Merkmale* sind und das Häufigkeitsverhältnis 9:3:3:1 ist.

Wir definiere also $E := \{AB, Ab, aB, ab\}$ und $\mathfrak{X} := \mathbb{E}^n$ sowie $\Theta := \{\vartheta = (\vartheta(k))_{k\in E} : \vartheta_{AB} + \vartheta_{Ab} + \vartheta_{aB} + \vartheta_{ab} = 1\}$ und $\rho := \left(\frac{9}{16}, \frac{3}{16}, \frac{3}{16}, \frac{1}{16}\right)$.

Für $\alpha = 0.1$ ist $\chi^2_{3,0.9} = 6.3$. Der zugehörige χ^2-Anpassungstest ist $\mathbb{1}_{\{D_{n,\rho} > 6.3\}}$.

Wir haben die folgenden $n := 556$ Beobachtungen gesammelt

AB	A b	a B	a b
315	108	101	32

Dann ist

$$D_{n,\rho} = \frac{\left(315 - 556 \cdot \frac{9}{16}\right)^2}{556 \cdot \frac{9}{16}} + \frac{\left(108 - 556 \cdot \frac{3}{16}\right)^2}{556 \cdot \frac{3}{16}} + \frac{\left(101 - 556 \cdot \frac{3}{16}\right)^2}{556 \cdot \frac{3}{16}} + \frac{\left(32 - 556 \cdot \frac{1}{16}\right)^2}{556 \cdot \frac{1}{16}}$$
$$= 0.47.$$

Da $D_{n,\rho} = 0.47 < 6.3 = \chi^2_{3,0.9}$ ist, lehnen wir die Nullhypothese nicht ab. ◇

Frage. Angenommen, wir wählten $E = \mathbb{R}$ anstatt $E = \{1, \dots, s\}$ und F sei eine vorgegebene Verteilungsfunktion auf $\mathbb{R}$. Wir wollen im Produktmodell die Hypothese F gegen alle anderen Verteilungsfunktionen auf $\mathbb{R}$ testen. Wir können den LQ-Test nicht verwenden, da dieses Modell kein Standardmodell ist (es existiert kein dominierendes Maß μ_0, sodass alle Wahrscheinlichkeitsmaße absolut stetig bezüglich μ_0 sind). Die Nullhypothese ist, dass F die Verteilungsfunktion ist.

Bemerkung 5.1.6 Eine Möglichkeit, welche wir jedoch nicht verfolgen werden, ist es, $\mathbb{R}$ in disjunkte Intervalle zu zerlegen („Gruppenbildung“) und dann den χ^2-Test anzuwenden. ○

Bemerkung 5.1.7 Dieses Problem ist nicht parametrisch. ○

Die folgende Aussage führt uns zu einem vernünftigen Test.

Satz 5.1.1: Gliwenko-Cantelli

Seien $X_1, X_2, \dots$ unabhängig identisch verteilte reelle Zufallsvariablen mit Verteilungsfunktion F und $F_n(t) := \frac{1}{n}\sum_{k=1}^n \mathbb{1}_{\{X_k \leqslant t\}}$ für $t \in \mathbb{R}$ die empirische Verteilungsfunktion. Dann konvergiert F_n fast sicher gleichmäßig gegen F.

Beweis. (1) Punktweise Konvergenz ist klar, denn für festes $t \in \mathbb{R}$ gilt nach dem starken Gesetz der großen Zahlen

$$F_n(t) = \frac{1}{n}\sum_{k=1}^n \mathbb{1}_{\{X_k \leqslant t\}} \to \mathbb{E}\left[\mathbb{1}_{\{X_1 \leqslant t\}}\right] = \mathbb{P}(X_1 \leqslant t) = F(t) \qquad \text{fast sicher.}$$

Es ist nicht klar, dass es einen globale Nullmenge gibt, auf der $F_n \to F$ punktweise fast sicher außerhalb jener Nullmenge gilt.

(2) Gleichmäßig Konvergenz. Für $t \in \mathbb{R}$ gilt

$$\lim_{s \nearrow t} F_n(s) = \frac{1}{n}\sum_{k=1}^n \mathbb{1}_{X_k < t} \xrightarrow[\text{GGZ}]{n\to\infty} \mathbb{E}[X_1 < t] = \mathbb{P}(X_1 < t) = \lim_{s \nearrow t} F(s) \qquad \text{fast sicher.}$$

Seien $N \in \mathbb{N}$ fest sowie $x_j := \inf\left\{t \in (-\infty, \infty) : F(t) \geqslant \frac{j}{N}\right\}$ für $j \in \{0, \dots, N\}$. Dann gilt

$$-\infty = x_0 \leqslant x_1 \leqslant \dots \leqslant x_N \leqslant \infty.$$

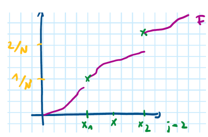


Abb. 12: Die Verteilungsfunktion F induziert eine Partition von $\mathbb{R}$ über die $(x_j)_{j=0}^N$.

Dann gilt wegen 1

$$R_n := \max_{j\in\{1,\dots,N\}} \left(|F_n(x_j) - F(x_j)| + |F_n(x_j-) - F(x_j-)|\right) \xrightarrow{n\to\infty} 0 \quad \text{fast sicher,}$$

wobei $G(x-) := \lim_{y\nearrow x} G(y)$ den einseitigen Grenzwert bezeichnet.

Für $x \in (x_{j-1}, x_j)$ gilt, da Verteilungsfunktionen monoton wachsend sind,

$$F_n(x) \leqslant F_n(x_j-) \leqslant F(x_j-) + R_n \leqslant F(x) + R_n + \frac{1}{N}$$
$$F_n(x) \geqslant F_n(x_{j-1}) \geqslant F(x_{j-1}) - R_n \geqslant F(x) - R_n - \frac{1}{N}$$

und somit

$$|F_n(x) - F(x)| \leqslant R_n + \frac{1}{N} \qquad \forall x \in (x_{j-1}, x).$$

Es folgt

$$\sup_{x\in\mathbb{R}} |F_n(x) - F(x)| \leqslant R_n + \frac{1}{N} \xrightarrow{n\to\infty} \frac{1}{N}$$

und somit

$$\limsup_{n\to\infty} \|F_n - F\|_\infty \leqslant \frac{1}{N} \qquad \text{fast sicher.}$$

Da $N \in \mathbb{N}$ beliebig war, folgt die Aussage. □

Warnung. Sind F und F_n Verteilungsfunktionen und $F_n(t) \xrightarrow{n\to\infty} F(t)$ für alle $t \in \mathbb{R}$ gilt, so folgt *nicht* $\|F_n - F\|_\infty \to 0$.

Definition 5.1.8 (Kolmogorov-Smirnoff Test)
Ein Kolmogorov-Smirnoff Test (oder: KS-Anpassungstest) ist ein Hypothesentest für die Nullhypothese F im n-fachen Produktmodell, welcher die folgende Form hat.

$$\varphi(x) = \begin{cases} 1, & \text{wenn } \|F_n - F\|_\infty > c, \\ 0, & \text{wenn } \|F_n - F\|_\infty < c. \end{cases}$$

Bemerkung 5.1.9 (Berechnung von $\|F_n - F\|_\infty$)
Sind die $X_1, \dots, X_n$ sind aufsteigend sortiert und F stetig, gilt

$$\|F_n - F\|_\infty = \max_{k\in\{1,\dots,n\}} \left(F(X_k) - \frac{k-1}{n}, \frac{k}{n} - F(X_k) \right),$$

das heißt, der betragsmäßig größte Abstand zwischen F und F_n ist an einer der Sprungstellen der empirischen Verteilungfunktion, welche die Werte $\left(\frac{j}{n}\right)_{j=0}^n$ annimmt (siehe auch das Lemma unten).

Die Fraktile für c hängen nicht von F ab und können in einer Tabelle nachgeschlagen werden. ○

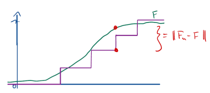


Abb. 13: Eine Verteilungsfunktion und eine empirische Verteilungsfunktion.

Lemma 5.1.10
Ist F stetig, so hängt die Verteilung von $\|F_n - F\|_\infty$ nicht von F ab.

Beweis. Sei $\Delta_n := \sup_{t\in\mathbb{R}} |F_n(t) - F(t)|$ für $n \in \mathbb{N}$. Unter der Nullhypothese, dass F stetig ist, sind fast sicher alle $x_1, \ldots, x_n$ verschieden. Sei $X_{k:n}$ der k-kleinste Wert, das heißt $X_{1:n} < \ldots < X_{n:n}$ mit $X_{k:n} \in \{X_1, \ldots, X_n\}$ für $k \in \{1, \ldots, n\}$.

Es gilt

$$\Delta_n = \max_{k\in\{1,\ldots,n\}} \max\left(\frac{k}{n} - F(X_{k:n}), F(X_{k:n}) - \frac{k-1}{n}\right).$$

Seien $U_k := F(X_k)$ für $k \in \{1, \ldots, n\}$. Die sind unabhängig identisch gleichverteilt mit Verteilung (da F stetig ist) $\mathcal{U}([0,1])$ und es gilt

$$\Delta_n = \max_{k\in\{1,\ldots,n\}} \max\left(\frac{k}{n} - U_{k:n}, U_{k:n} - \frac{k-1}{n}\right).$$

Also hängt die Verteilung von Δ_n nicht von F ab. □

Beispiel 5.1.11 (Benzinverbrauch)
Eine zufällige Auswahl von zehn Autos eines bestimmten Typs ergab folgenden Verbrauch in ℓ/100km:

10.8	11.3	10.4	9.8	10.0
10.6	11.0	10.5	9.5	11.2

Wir testen zum Niveau $\alpha = 0.05$, ob der Benzinverbrauch $\mathcal{N}_{10,1}$ ist.

Zunächst sortieren wir die Beobachtungen aufsteigend.

k	1	2	3	4	5	6	7	8	9	10
X_k	9.5	9.8	10.0	10.4	10.5	10.6	10.8	11.0	11.2	11.3

Die Verteilungsfunktion von $\mathcal{N}_{10,1}$ ist $F(x) = \Phi(X - 10)$, wobei Φ die Verteilungsfunktion einer Standardnormalverteilung ist. Dann gilt

k	1	2	3	4	5	6	7	8	9	10
$F(X_k)$	0.31	0.42	0.5	0.66	0.69	0.73	0.79	0.84	0.89	0.90
$F(X_k) - \frac{k-1}{n}$	0.31	0.32	0.3	0.36	0.29	0.23	0.19	0.14	0.09	0.00
$\frac{k}{n} - F(X_k)$	−0.21	−0.22	−0.2	−0.26	−0.19	−0.13	−0.09	−0.04	0.01	0.10

Daher ist $\|F_n - F\|_\infty = 0.36$. Aus einer Tabelle erhalten wir $c = 0.41$, da $\|F_n - F\|_\infty < c$ gilt, lehnen wir die Nullhypothese nicht ab. ◇

28.06.2022

Somit ist die Verteilung von Δ_n gleich der von

$$G(t) = \begin{cases} 0, & t \leqslant 0, \\ t, & 0 \leqslant t \leqslant 1, \\ 1, & t \geqslant 1 \end{cases}$$

statt F und $G_n(t) := \frac{1}{n}\sum_{k=1}^n \mathbb{1}_{U_{k:n}\leqslant t}$ anstatt $F_n(t)$.

Was ist die asymptotische Verteilung von Δ_n für $n \to \infty$?

Ohne Beweis. Sei $V_n(t) := \sqrt{n}\big(G_n(t) - G(t)\big)$ für $t \in [0,1]$ ein stochastischer Prozess. Es ist klar, dass für festes $t \in [0,1]$ $\mathbb{E}[V_n(t)] = 0$ sowie $\mathbb{V}[V_n(t)] = (\sqrt{n})^2 \frac{1}{n^2} nt(1-t) = t(1-$

t) ($\mathbb{1}_{U_{k:n} \leqslant t}$ eine Ber(t)-verteilte Zufallsvariable) gilt. Nach dem zentralen Grenzwertsatz konvergiert $V_n(t)$ in Verteilung (haben wir nicht definiert) gegen $\mathcal{N}(0, t(1-t))$. Unklar ist, ob der Prozess V_n gegen einen einfacheren stochastischen Prozess in Verteilung konvergiert. Man kann zeigen (cf. [3]), dass V_n in Verteilung im Raum $D[0,1]$ gegen eine BROWNsche Brücke konvergiert, das heißt gegen einen GAUSS-Prozess auf $[0,1]$ mit einer Darstellung $V(t) = W(t) - tW(1)$ für $t \in [0,1]$, wobei $W(t)$ für $t \geqslant 0$ eine Standard BROWNsche Bewegung ist.

Daraus folgt, dass die Verteilung von $\sqrt{n}\Delta_n$ in Verteilung gegen $\sup_{t\in[0,1]} |V(t)|$ konvergiert. Die Verteilungsfunktion dieser Größe ist

$$\mathbb{P}\left[\sup_{t\in[0,1]} |V(t)| \leqslant x\right] = \begin{cases} 0, & x \leqslant 0, \\ 1 - 2\sum_{k=1}^{\infty}(-1)^{k+1} e^{-2k^2x^2}, & x > 0. \end{cases}$$

Damit kann man für vorgegebenes α die Konstante c approximativ so bestimmen, dass der obige Test Niveau α hat.

Bemerkung 5.1.12 Der KS-Test lässt sich so modifizieren, dass damit das „Zweistichprobenproblem“ behandelt werden kann: seien $X_1, X_2, \dots$ u.i.v. reellwertig mit Verteilungsfunktion F und empirischer Verteilungsfunktion F_n sowie $Y_1, Y_2, \dots$ u.i.v. reellwertige Zufallsvariablen mit Verteilungsfunktion $\tilde{F}$ und empirischer Verteilungsfunktion $\tilde{F}_n$. Die Verteilungsfunktionen F und $\tilde{F}$ sind unbekannt. Wir wollen die Nullhypothese $F = \tilde{F}$ gegen die Alternative $F \neq \tilde{F}$ testen. Man wird bei n bzw. m Proben $X_1, \dots, X_n$ sowie $Y_1, \dots, Y_m$ die Nullhypothese ablehnen, wenn $\Delta_{n,m} = \|F_n - \tilde{F}_m\|_\infty$ (die Größe ist verteilungsunabhängig wenn F und $\tilde{F}$ stetig sind) groß ist (Literatur: Wikipedia, [3]). ○

5.2 χ^2-Test auf Unabhängigkeit

Das folgende Beispiel kommt aus [2, Bsp. 11.17].

Beispiel 5.2.1 (χ^2-Test auf Unabhängigkeit)
Wir betrachten zwei „Merkmale“ mit je *endlich* vielen verschiedenen Ausprägungen, z.B. Schulbildung (ungelernt, ..., Uniabschluss) und die Frage, wie stark sich die gefragten Personen durch Umweltschadstoffe beeinträchtigt fühlen.

Beeinträchtigung \ Schulbildung	ungelernt				Uni	Σ
gar nicht	212	434	169	79	45	939
etwas	85	245	146	93	69	628
ziemlich	38	85	74	56	48	301
sehr	20	35	30	21	20	126
Σ	355	799	419	249	182	2004

Tabelle 3: Kontingenztafel zur „Umweltfrage“.

Sind die beiden Merkmale unabhängig?

Allgemein: Seien $A = \{1, \dots, a\}$, $B := \{1, \dots, b\}$ die möglichen Ausprägungen der beiden Merkmale. Für $E := A \times B$ betrachten wir das Produktmodell $\left(E^{\otimes \mathbb{N}}, \mathcal{P}(E)^{\otimes \mathbb{N}}, (\vartheta^{\otimes \mathbb{N}})_{\vartheta \in \Theta}\right)$, wobei $\Theta := \{\vartheta = \left(\vartheta(i,j)\right)_{(i,j) \in E} \in (0,1)^E : \sum_{(i,j) \in E} \vartheta(i,j) = 1\}$. Für $\vartheta \in \Theta$ seien $\vartheta^A = (\vartheta^A(i))_{i \in A}$ mit $\vartheta^A(i) := \sum_{j=1}^b \vartheta(i,j)$ und ϑ^B analog definiert die Randverteilungen von $\vartheta \in \Theta$.

Seien $\Theta_0 := \{\vartheta \in \Theta : \vartheta = \vartheta^A \otimes \vartheta^B\}$ sowie $\Theta_1 := \Theta \backslash \Theta_0$.

Nach n unabhängigen E-wertigen Beobachtungen $X_1, \dots, X_n$ ergibt sich eine „Kontingenztafel“ mit den Einträgen $h_n(i,j) := |\{k \in \{1, \dots, n\} : X_k = (i,j)\}|$ (absolute Häufigkeiten). Die normierten Werte sind $L_n(i,j) := \frac{1}{n} h_n(i,j)$. Ferner definieren wir $L_n^A(i) := \sum_{j=1}^b L_n(i,j)$ und analog für h_n^A sowie analog für B.

Wir werden Θ_0 ablehnen, wenn L_n nicht nah bei $L_n^A \otimes L_n^B$ liegt. Da wir ein Standardmodell vorliegen haben, können wir einen LQ-Test anwenden. Der Likelihoodquotient ist (wegen Stetigkeit in ϑ können wir im Zähler wieder Θ_1 durch Θ ersetzen)

$$\begin{aligned}
R_n &= \frac{\sup_{\vartheta \in \Theta} \prod_{(i,j) \in E} \vartheta(i,j)^{h_n(i,j)}}{\sup_{\gamma \otimes \beta \in \Theta_0} \prod_{(i,j) \in E} (\gamma(i)\beta(j))^{h_n(i,j)}} = \frac{\sup_{\vartheta \in \Theta} \prod_{(i,j) \in E} \vartheta(i,j)^{h_n(i,j)}}{\sup_\gamma \prod_{i \in A} \gamma(i)^{h_n^A(i)} \sup_\beta \prod_{j \in B} \beta(j)^{h_n^B(j)}} \\
&= \frac{\prod_{(i,j) \in E} L_n(i,j)^{h_n(i,j)}}{\prod_{i \in A} L_n^A(i)^{h_n^A(i)} \prod_{j \in B} L_n^B(j)^{h_n^B(j)}} = \prod_{(i,j) \in E} \left(\frac{L_n(i,j)}{L_n^A(i) L_n^B(j)} \right)^{n L_n(i,j)} \\
&= \exp\left(n H\left(L_n; L_n^A \otimes L_n^B\right)\right).
\end{aligned}$$

Ein kleines Problem ist, dass die rechte Seite auch für $\vartheta \in \Theta_0$ von ϑ abhängt. Um ein Niveau $\alpha \in (0,1)$ zu garantieren, müsste man sicherstellen, dass $\sup_{\beta,\gamma} \mathbb{P}_{\gamma \otimes \beta}(R_n > c) \leqslant \alpha$ gilt!

Man betrachtet wie beim χ^2-Anpassungstest die Asymptotik für $n \to \infty$:

$$\tilde{D}_n := n \sum_{i,j} L_n^A(i) L_n^B(j) \left(\frac{L_n(i,j)}{L_n^A(i) L_n^B(j)} - 1 \right)^2 .$$

[2, Satz 11.18] zeigt, dass unter $\rho = \gamma \otimes \beta$ der Wert $nH\big(L_n; L_n^A \otimes L_n^B\big)$ „nah" bei $\tilde{D}_n$ liegt und dass $\tilde{D}_n$ in Verteilung gegen eine $\chi^2_{(a-1)(b-1)}$-Verteilung konvergiert (das Schätzen der Zähldichte γ „verbraucht" $a-1$ Freiheitsgrade, das Schätzen von β $b-1$. Die Gesamtzahl $ab-1$ der Freiheitsgrade verringert sich daher um $(a-1)+(b-1)$ und es bleiben nur $(a-1)(b-1)$ Freiheitsgrade übrig). ◇

Bemerkung 5.2.2 Beim „Umweltbespiel" wird die Nullhypothese bei $\alpha = 0.01$ deutlich abgelehnt [2, S. 307]. ○

5.3 Vorzeichen- und Rangtests

30.06.2022

Gute Quellen sind [2, Kap. 11.4] sowie [1].

Definition 5.3.1 (α-Quantil, Median, $\mu(Q)$)
Seien Q ein Wahrscheinlichkeitsmaß auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ und $\alpha \in (0,1)$. Ein $q \in \mathbb{R}$ ist ein α-Quantil von Q wenn $Q((-\infty, q)) \leqslant \alpha \leqslant Q((-\infty, q])$ ist. Ein $\frac{1}{2}$-Quantil ist der Median von Q. Ferner bezeichne $\mu(Q)$ die Menge aller Mediane von Q.

α-Quantil

Definition 5.3.2 (Fraktil)
In der obigen Situation ist das α-Fraktil das $(1-\alpha)$-Quantil.

Bemerkung 5.3.3 Die Menge alle α-Quantile von Q ist ein kompaktes nichtleeres Intervall. ○

Notation. Seien $\mu_{\max}(Q) := \max(\mu(Q))$ und $\mu_{\min}(Q)$ analog definiert.

Beispiel 5.3.4
Beim Werfen eines fairen Würfels ist $\mu(Q) = [3,4]$. ◇

Definition 5.3.5 (Ordnungsstatistik)
Seien $X_1, X_2, \ldots, X_n$ unabhängige reellwertige Zufallsvariablen mit stetigen Verteilungsfunktionen. (Dann sind fast sicher alle $X_1, X_2, \ldots, X_n$ verschieden.) Die Ordnungsstatistik der $X_1, \ldots, X_n$ sind definiert durch $X_{1:n} < X_{2:n} < \ldots < X_{n:n}$ mit $X_{k:n} \in \{X_1, \ldots, X_n\}$ für $k \in \{1, \ldots, n\}$.

Ordnungsstatistik

Notation. Für $\alpha \in (0,1)$ sei $b_n(\alpha)$ das größte α-Quantil der Binomialverteilung $B_{n,\frac{1}{2}}$.

Wir wollen testen, ob im Produktmodell der Median (welcher viel stabiler ist als der Mittelwert) gleich einer vorgegebenen Zahl μ_0 oder größere als μ_0 (einseitiger Hypothesentest) bzw. gleich μ_0 oder ungleich μ_0 ist (zweiseitiger Hypothesentest).

Bemerkung 5.3.6 Hat die Zufallsvariable X eine „stetige Verteilung" (das heißt die Verteilungsfunktion ist stetig), so gilt

$$\begin{aligned}\mathbb{P}(X \in \mu(Q)) &= \mathbb{P}(\mu_{\min}(Q) \leqslant X \leqslant \mu_{\max}(Q)) \\ &= \mathbb{P}(X \leqslant \mu_{\max}(Q)) - \mathbb{P}(X \leqslant \mu_{\min}(Q)) = \frac{1}{2} - \frac{1}{2} = 0.\end{aligned}$$

○

Den folgenden Satz kann man auch (in ungenauerer Fassung) in [2, Satz 8.19] finden.

Satz 5.3.1: Konfidenzintervall für Median im Produktmodell

Seien $n \in \mathbb{N}$, $\alpha \in (0,1)$ sowie $k = b_n(\frac{\alpha}{2})$. Dann ist $[X_{k:n}, X_{n-k+1:n}]$ im Produktmodell Q^n mit stetigem Q ein Konfidenzintervall für $\mu(Q)$, das heißt $Q^{\otimes n}\big(\mu(Q) \subset [X_{k:n}, X_{n-k+1:n}]\big) \geqslant 1 - \alpha$.

Beweis. Es gilt

$$Q^{\otimes n}\big(X_{k:n} > \mu_{\min}(Q)\big) = Q^{\otimes n}\left(\sum_{j=1}^{n} \mathbb{1}_{\{X_j \leqslant \mu_{\min}(Q)\}} < k\right) = B_{n,\frac{1}{2}}(\{0,\dots,k-1\}) \leqslant \frac{\alpha}{2}$$

und ebenso

$$Q^{\otimes n}\big(X_{n-k+1:n} < \mu_{\max}(Q)\big) \leqslant \dots \leqslant \frac{\alpha}{2},$$

also folgt

$$Q^{\otimes n}\big(\mu(Q) \subset [X_{k:n}, X_{n-k+1:n}]\big) \geqslant 1 - \frac{\alpha}{2} - \frac{\alpha}{2} = 1 - \alpha.$$

□

Beispiel 5.3.7 (Vorzeichentest (*engl.* signed test))
Zwei Dünger sollen verglichen werden. Sind beide gleich gut? Eine Alternative ist „Dünger 2 und Dünger 1 sind verschieden“, eine andere ist „Dünger 2 ist besser als Dünger 1“.

Versuchsanordnung. Jedes von n Feldern wird in zwei gleichgroße Teile geteilt und dort Dünger 1 bzw. Dünger 2 verwendet. Die Erträge der Feldern seien $(Y_1, Z_1), \dots, (Y_n, Z_n)$, wobei Y_k der Ertrag auf Feld k mit Dünger 1 und Z_k der Ertrag auf Feld k mit Dünger 2 sind. Da die Felder unterschiedlich sind (z.B. Bodenbeschaffenheit, Mikroklima, ...), nehmen wir *nicht* an, dass die Paare $\big((Y_k, Z_k)\big)_{k=1}^n$ identisch verteilt sind. Da globale Ereignisse (Hochwasser, Dürre, Sturm, ...) auftreten können, welche alle Felder betreffen, wollen wir *nicht* annehmen, dass die Paare $\big((Y_k, Z_k)\big)_{k=1}^n$ unabhängig sind.

Als Nullhypothese wählen wir, dass nicht nur Y_k und Z_k für jedes k die selbe Verteilung haben, sondern sogar (Y_k, Z_k) die selbe Verteilung hat wie (Z_k, Y_k). Weitere Annahmen (unter der Nullhypothese) seien, dass die $X_k := Y_k - Z_k$ stetige Verteilungsfunktionen F_k für $k \in \{1, \dots, n\}$ haben und das

$$I_k := \begin{cases} 1, & X_k > 0 \\ 0, & X_k = 0 \\ -1, & X_k < 0 \end{cases}$$

für $k \in \{1, \dots, n\}$ unabhängig sind. Unter der Nullhypothese hat X_k dieselbe Verteilung wie $-X_k$, somit ist die Menge der Mediane von X_k ein symmetrisches Intervall um die Null für alle $k \in \{1, \dots, n\}$.

Wir lehnen die Nullhypothese ab, wenn im zweiseitigen Fall die Anzahl der Einsen unter den $I_1, \dots, I_n$ stark von $\frac{n}{2}$ abweicht (das heißt $|\sum_{k=1}^n I_k|$ „groß“). Die entspricht einem Test auf Fairness einer Münze beim n-fachen unabhängigen Werfen.

Allgemeiner: Seien $X_1, X_2, \dots, X_n$ reelle Zufallsvariablen. Wir wollen testen, ob alle Verteilungen den Median μ_0 (vorgegebene Zahl) haben (oben ist $\mu = 0$). Die einzigen Annahmen sind, dass alle Verteilungsfunktionen F_k von X_k (müssen nicht gleich sein) stetig sind und $I_k := \begin{cases} 1, & X_k > \mu_0, \\ -1, & X_k < \mu_0 \end{cases}$ für $k \in \{1, \dots, n\}$ unter der Nullhypothese unabhängig sind mit $\mathbb{P}(I_k = 1) = \mathbb{P}(I_k = -1) = \frac{1}{2}$ für alle $k \in \{1, \dots, n\}$. Die Nullhypothese wird abgelehnt, wenn $|\sum_{k=1}^n I_k|$ „verdächtig“ groß ist. Ein Test zum Niveau α lässt sich durch Quantile $b_n(\alpha)$ von $B_{n,\frac{1}{2}}$ bestimmen. ◇

Wie kann man den Schwellenwert bestimmen?

Rangtests

Ist Dünger 2 besser Dünger 1 oder sind beide gleich gut? Dünger 1 wird auf k Feldern verwendet und Dünger 2 auf ℓ Feldern. Die Erträge seien $X_1, \dots, X_k$ sowie $X_{k+1}, \dots, X_n$ mit $n := k + \ell$. Wir nehmen an, dass $X_1, \dots, X_n$ unabhängig sind und dass $X_1, \dots, X_k$ die Verteilung P haben und $X_{k+1}, \dots, X_n$ die Verteilung Q haben, wobei P und Q unbekannt aber stetig sind. Ist Dünger 2 besser, so sind $X_{k+1}, \dots, X_n$ „eher" größer als $X_1, \dots, X_k$, das heißt $P \geqslant Q$.

Definition 5.3.8 (Stochastisch kleiner)
Seien P und Q Wahrscheinlichkeitsmaße auf $\big(\mathbb{R}, \mathcal{B}(\mathbb{R})\big)$. Dann ist P stochastisch kleiner als Q, und wir schreiben $P \preceq Q$, wenn $P((c,\infty)) \leqslant Q((c,\infty))$ für alle $c \in \mathbb{R}$ (oder äquivalent: $P((-\infty, c]) \geqslant Q((-\infty, c])$ für alle $c \in \mathbb{R}$). Wir schreiben $P \prec Q$, wenn $P \preceq Q$ und $P \neq Q$ ist.

Beispiel 5.3.9 ([2, 11.23])
Ist $m < m'$, so sind $\mathcal{N}(m, v) \prec \mathcal{N}(m', v)$ und $\Gamma_{a,m} \prec \Gamma_{a,m'}$ für $a > 0$. ◇

Wir testen die Hypothese $H_0 : P = Q$ gegen die Alternative $H_1 : P \prec Q$.

05.07.2022

Definition 5.3.10 (Rangstatistik)
Als Rangstatistik der Folge $X_1, \dots, X_n$ bezeichnet man die Zufallsvariablen $R_1, \dots, R_n$ mit $R_m := \#\big\{j \in \{1, \dots, n\} : X_j \leqslant X_m\big\} \geqslant 1$ für $m \in \{1, \dots, n\}$.

Rangstatistik

Beispiel 5.3.11 (Rangstatistik)
Sind $X_2 < X_1 < X_4 < X_3$, so sind $(R_1, R_2, R_3, R_4) = (2, 1, 4, 3)$. ◇

Bemerkung 5.3.12 (Rang- vs. Ordnungsstatistik) Es gilt $X_k = X_{R_k:n}$. ○

Bemerkung 5.3.13 Die Werte $\{R_1, \dots, R_n\}$ sind eine Permutation von $\{1, \dots, n\}$. ○

Welche Funktionen der Ränge der $X_1, \dots, X_k$ können wir wählen, um zu entscheiden, ob diese hinreichend „klein" sind? Wir wählen die Rangsummen.

Definition 5.3.14 (Rangsummen W_P, W_Q)
Es seien $W := W_P := R_1 + \ldots + R_k$ sowie $W_Q := R_{k+1} + \ldots + R_n$.

Nach Bemerkung 5.3.13 gilt $W_P + W_Q = \sum_{j=1}^n j = \frac{1}{2}n(n+1)$. Wir lehnen H_0 ab, wenn W „klein" ist und sonst nicht.

Definition 5.3.15 (U-Statistik)
Die U-Statistik ist

U-Statistik

$$\boxed{U = U_{k,\ell} := \sum_{m=1}^{k} \sum_{j=k+1}^{n} \mathbb{1}_{\{X_m > X_j\}}}$$

Lemma 5.3.16 (Rangsumme und U-Statistik)
Es gilt $W = U + \frac{1}{2}k(k+1)$.

Beweis. Ordnet man die ersten $X_1, \dots, X_k$ so um, dass $X_1 < \dots < X_k$, dann ändern sich weder W noch U. Seien also $X_1 < \dots < X_k$. Für die Ränge $R_1 < \dots < R_k$ gilt

$$R_m = m + \#\{j \in \{k+1, \dots, n\} : X_j < X_k\} \tag{33}$$

für $m \in \{1, \dots, k\}$ und somit

$$W = \sum_{m=1}^{k} R_m \overset{(33)}{=} \sum_{m=1}^{k} m + \underbrace{\sum_{m=1}^{k} \#\{j \in \{k+1, \dots, n\} : X_j < X_k\}}_{=\sum_{m=1}^{k}\sum_{j=k+1}^{n} \mathbb{1}_{\{X_m > X_j\}}} = \frac{1}{2}k(k+1) + U.$$

□

Definition 5.3.17 (Mann-Whitney-U-Test)
Ein Test der Nullhypothese $H_0 : P = Q$ gegen $H_1 : P \prec Q$ mit Ablehnungsbereich $\{U < c\} = \{W < \frac{1}{2}k(k+1) + c\}$ mit $c \in \{0, \dots, k\ell\}$ heißt Mann-Whitney-U-Test oder Wilcoxon-Zweistichproben-Rangsummentest.

Bemerkung 5.3.18 (Randfall $k = \ell = 1$) Ist $k = \ell = 1$, dann ist $U = 1$ genau dann wenn $X_1 > X_2$ und $U = 0$ genau dann wenn $X_1 < X_2$. Unter H_0 ist die Wahrscheinlichkeit $P(U < 1) = P(U = 0) = \frac{1}{2}$ und $P(U < 0) = 0$ sowie $P(U < 2) = 1$.

Sei $\alpha \in (0, \frac{1}{2})$. Dann gilt $P(U < c) \leqslant \alpha$ nur, wenn $c = 0$ ist. ○

Wie kann man im Allgemeinen c zu $\alpha \in (0,1)$ wählen, damit der Mann-Whitney-U-Test Niveau α hat?

Satz 5.3.2: U-Verteilung unter H_0

Für stetige P und $m \in \{0, \dots, k\ell\}$ gilt $P^{\otimes n}(U = m) = \frac{1}{\binom{n}{k}} N(m; k, \ell)$, wobei $N(m; k, \ell)$ die Anzahl der Partitionen $\sum_{j=1}^{k} m_j = m$, von m in k aufsteigend geordnete Zahlen $m_1 \leqslant m_2 \leqslant \dots \leqslant m_k$ aus $\{0, \dots, \ell\}$ ist.

Bemerkung 5.3.19 Die Verteilung von U hängt insbesondere nicht von P ab, wenn P stetig ist. ○

Beweis. (Idee) Sei $R := \{R_1, \dots, R_k\}$ ist eine zufällige k-elementige Teilmenge von $\{1, \dots, n\}$ gleichverteilt (da P stetig ist) auf allen $\binom{n}{k}$ k-elementigen Teilmengen von $\{1, \dots, n\}$. Für ein $r = (r_1, \dots, r_k)$ mit $r_1 < \dots < r_k$ und $r_j \in \{1, \dots, n\}$ sei $m(r) := \big(m_1(r), \dots, m_k(r)\big)$ mit $m_j(r) := r_j - j$ für $j \in \{1, \dots, k\}$. Dann ist $r \mapsto m(r)$ eine Bijektion zwischen der obigen Menge der r und der Menge aller $\{0, \dots, \ell\}$-wertigen aufsteigenden Folgen $m_1 \leqslant m_2 \leqslant \dots \leqslant m_k$.

Weiter ist $m_j(r)$ die Anzahl alle Elemente in $\{1, \dots, n\} \backslash \{r_1, \dots, r_k\}$, die kleiner als j sind. Also ist $\sum_{j=1}^{k} m_j(R) = U$, das heißt $\sum_{j=1} m_j(R) = m$ genau dann wenn $U = m$, und das ist die Behauptung. □

Beispiel 5.3.20 (Zum Beweis)
Es bedeute 0, dass eine Realisierung zu der ersten Stichprobe stammt und 1, das sie aus der zweiten Stichprobe kommt. Unser Ergebnis sei 10001101. Dann sind $k = \ell = 4$ und $n = 8$ sowie $m_1(r) = 0$ und $m_2(r) = 2$, da die zweite Null im Ergebnis an vierter Stelle steht. ◇

Bemerkung 5.3.21 Also lautet der MANN-WHITNEY-U-Test zum Niveau $\alpha \in (0,1)$: berechne $\frac{1}{\binom{n}{k}} N(m; k, \ell)$ für $m \in \{0, 1, \ldots\}$ solange, bis die Summe größer als α ist. Wenn das bei m_0 der Fall ist, dann wähle $c = m_0$. ○

Für große k und ℓ liefert der folgende Satz einen Test zum approximativen Niveau α:

SATZ 5.3.3: HOEFFDING: $U^*_{k,\ell} \to \mathcal{N}(0,1)$

Seien $X_1, \ldots, X_n$ unabhängig und identisch verteilt mit stetiger Verteilung P. Dann konvergiert $U^*_{k,\ell} := v_{k,\ell}^{-\frac{1}{2}} \left(U_{k,\ell} - \frac{1}{2} k\ell \right)$ (der Erwartungswert von $U_{k,\ell}$ ist $\frac{1}{2}k\ell$) in Verteilung gegen $\mathcal{N}(0,1)$ für $k, \ell \to \infty$, wobei $v_{k,\ell} := \frac{1}{12} k\ell(k + \ell + 1)$.

Beweis. [2, Satz 12.28]. □

Beispiel 5.3.22
Um die Verlängerung der Reaktionszeit durch ein bestimmtes Medikament zu untersuchen, wurden 20 Personen einem Reaktionstest unterzogen, von denen 10 zuvor das Medikament eingenommen hatten und die anderen 10 eine Kontrollgruppe bildeten. Es ergaben sich folgende Reaktionszeiten (in Sekunden):

behandelte Gruppe	0.83	0.66	0.94	0.78	0.81	0.60	0.88	0.90	0.79	0.86
Kontrollgruppe	0.64	0.70	0.69	0.80	0.71	0.82	0.62	0.91	0.59	0.63

Es ist Hausaufgabe 12.2 mit einem MANN-WHITNEY-U-Test zum Niveau $\alpha := 0.05$ die Hypothese zu testen, dass die Reaktionszeit durch das Medikament nicht beeinflusst wird, gegen die Alternative einer verlängerten Reaktionszeit, und zwar (a) exakt, (b) unter Verwendung der Normalapproximation. ◇

Bemerkung 5.3.23 (Zusammenhang U-Test ↔ KS Zweistichprobentest)
Wir haben gesehen, dass der KOLMOGOROV-SMIRNOFF-Zweistichprobentest $\Delta_{k,\ell} := \|F_k - \tilde{F}_\ell\|_\infty$, wobei F_k resp. $\tilde{F}_\ell$ die empirischen Verteilungsfunktionen von $X_1, \ldots, X_k$ resp. $X_{k+1}, \ldots, X_n$ sind, benutzt wird. Dieses $\Delta_{k,\ell}$ hängt nur von den Ränge von $X_1, \ldots, X_n$ ab (die Differenz erreicht ihr Maximum an den Sprungstellen...). Wir modifizieren den KOLMOGOROV-SMIRNOFF-Zweistichprobentest, sodass er besser zur Alternative $P < Q$ passt: wähle $\overline{\Delta_{k,\ell}} := \sup_{x \in \mathbb{R}} F_k(x) - \tilde{F}_\ell(x)$, welche wieder eine Funktion der Ränge von $X_1, \ldots, X_n$ ist.

Beispiel 5.3.24
Es ist nicht so, dass $\overline{\Delta_{k,\ell}}$ eine Funktion der Rangsumme W ist. Seien zum Beispiel $k = \ell = 3$ sowie

X_1	X_2	X_3	X_4	X_5	X_6	W	$\overline{\Delta_{3,3}}$
1	3	5	2	4	6	9	$\frac{1}{3}$
1	2	6	3	4	5	9	$\frac{2}{3}$

Dann ist W für beide paare von Stichproben (erste und zweite Zeile) gleich, jedoch ist $\overline{\Delta_{3,3}}$ unterschiedlich. $\diamond$

6 Lineare Modelle und Varianzanalyse

6.1 Einfache lineare Regression

07.07.2020

Exemplarisch betrachten wir die Wärmeausdehnung eines Metallstabs. Für vorgegebene Temperaturen $t_1, t_2, \dots, t_n$ wird jeweils die Länge des Metallstabs X_k gemessen.

Wir modellieren die Länge als $X_k = \gamma_0 + \gamma_1 t_k + \sqrt{v}\xi_k$ für $k \in \{1, \dots, n\}$, wobei γ_0 und γ_1 unbekannt sind und geschätzt werden sollen und $v > 0$ ein „Störparameter“ ist. Letztlich sind $\xi_1, \dots, \xi_n$ Zufallsvariablen, welchen den Messfehler beschreiben.

Wir müssen nicht annehmen, dass die ξ_k unabhängig, unkorreliert oder identisch verteilt sind, wir nehmen jedoch $\mathbb{E}[\xi_k] = 0$ und $\mathbb{V}[\xi_k] = 1$ an.

Mit $X := (X_1, \dots, X_n)$, $t := (t_1, \dots, t_n)$, $\xi := (\xi_1, \dots, \xi_n)$ und $\mathbb{1} := (1, \dots, 1) \in \mathbb{R}^n$ können wir das Modell knapp als $X = \gamma_0 \mathbb{1} + \gamma_1 t + \sqrt{v}\xi$ schreiben.

Wir wählen das Modell $\left(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), (\mathbb{P}_{\gamma,v})_{(\gamma,v)\in\mathbb{R}^2 \times (0,\infty)}\right)$ (*kein* Produktmodell!), wobei $\mathbb{P}_{\gamma,v}$ die Verteilung von $\gamma_0 \mathbb{1} + \gamma_1 t + \sqrt{v}\xi$ hat.

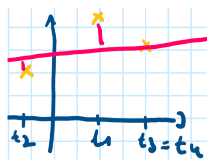


Abb. 14: Der KQ-Schätzer bestimmt die Regressionsgerade $\hat\gamma_0 + t\hat\gamma_1$, welche die Daten approximieren soll.

Definition 6.1.1 (Prinzip der kleinsten Quadrate, KQ-Schätzer)
Das Prinzip der kleinsten Quadrate („least squares“) wählt $\hat\gamma = (\hat\gamma_0, \hat\gamma_1) \in \mathbb{R}^2$ so, dass der mittlere quadratische Fehler

$$F_\gamma := \frac{1}{n}\sum_{k=1}^{n}(X_k - \gamma_0 - \gamma_1 t_k)^2$$

minimiert wird. Dieses $\hat\gamma$ heißt KQ-Schätzer (sofern er eindeutig ist).

Bemerkung 6.1.2 Sind alle t_k gleich, dann ist die Lösung $\hat\gamma$ nicht eindeutig, unter anderem, da dann t und $\mathbb{1}$ linear abhängig sind. Wir nehmen daher an, dass nicht alle t_k gleich sind. ○

Satz 6.1.1: KQ-Schätzer für einfache lineare Regression

Seien t und $\mathbb{1}$ linear unabhängig sowie $M(y) := \frac{1}{n}\sum_{k=1}^{n} y_k$ und $V(t) := \frac{1}{n}\sum_{k=1}^{n} t_k^2 - M(t)^2$ sowie $c(t,y) := \frac{1}{n}\langle t, y\rangle - M(t)M(y)$ für $t, y \in \mathbb{R}^n$. (Wegen $t \neq \lambda\mathbb{1}$ folgt $V(t) > 0$.) Der KQ-Schätzer des obigen Problems $X = \gamma_0 \mathbb{1} + \gamma_1 t + \sqrt{v}\xi$ ist

$$\hat\gamma_0 = M(x) - M(t)\frac{c(t,x)}{V(t)} \qquad \hat\gamma_1 = \frac{c(t,x)}{V(t)}$$

und ist erwartungstreu.

Beweis. Wir erhalten $\hat\gamma$ aus dem linearen Gleichungssystem $0 = \frac{\partial}{\partial\gamma_0}F_\gamma = \frac{\partial}{\partial\gamma_1}F_\gamma$. □

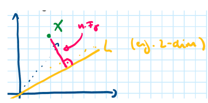


Abb. 15: Die Regressionsgerade $\hat\gamma_0 \mathbb{1} + \hat\gamma_1 t$ ist die orthogonale Projektion von X auf L.

Bemerkung 6.1.3 Dieses Modell (und damit die Aussage des Satzes) ist ein Spezialfall des linearen Modells im nächsten Abschnitt, cf. Beispiel 6.3.5. ○

Bemerkung 6.1.4 Die Menge $L := \{\gamma_0 \mathbb{1} + \gamma_1 t : \gamma_0, \gamma_1 \in \mathbb{R}\} \subset \mathbb{R}^n$ ist ein zweidimensionaler Teilraum, wenn $\mathbb{1}$ und t linear unabhängig sind und sonst ist die Dimension gleich eins. ○

Bemerkung 6.1.5 Ist L eindimensional, so existiert die orthogonale Projektion von x auf L und ist eindeutig, aber die Darstellung $\hat{\gamma}_0 \mathbb{1} + \hat{\gamma}_1 t$ ist *nicht* eindeutig. ○

Beispiel 6.1.6 (Motorleistung) Ein Motor erbrachte bei $n = 8$ Messungen die folgende Leistung (in kW) in Abhängigkeit vom Drehmoment (1000 U / min):

t_k	0.8	1.5	2.5	3.5	4.2	4.7	5.0	5.5
X_k	8.8	14.7	22.8	29.4	38.2	44.1	47.8	51.5

Wir berechnen die Regressionsgerade und schätzen die Leistung für das Drehmoment 4000 U / min.

Es sind

$$M(t) := \frac{1}{8}\sum_{k=1}^{8} t_k \approx 3.46 \qquad \text{und} \qquad M(X) := \frac{1}{8}\sum_{k=1}^{8} X_k \approx 32.16$$

sowie

$$V(t) := \frac{1}{8}\sum_{k=1}^{8} t_k^2 - M(t)^2 \approx 14.55 - (3.46)^2 \approx 2.56$$

sowie

$$c(t,x) := \frac{1}{n}\sum_{k=1}^{8} t_k X_k - M(t)M(x) \approx 134.87 - 3.46 \cdot 32.16 \approx 23.51.$$

Nach Satz 6.1.1 erhalten wir die erwartungstreuen Schätzer

$$\hat{\gamma}_0 := M(x) - M(t)\frac{c(t,x)}{V(t)} \approx 0.34 \qquad \hat{\gamma}_1 := \frac{c(t,x)}{V(t)} \approx 9.19,$$

also ist die Regressionsgerade $9.19t + 0.34$. Für das Drehmoment 4000 U / min schätzen wir also eine Leistung von 37.10 kW. ◇

6.2 Das lineare Modell

Definition 6.2.1 (Lineare Modell)
Seien $s, n \in \mathbb{N}$ mit $s \leqslant n$. Ein lineares Modell für n reellwertige Beobachtungen mit unbekannten Verschiebungsparameter $\gamma \in \mathbb{R}^s$ und $v > 0$ besteht aus lineares Modell

- einer „Design"-Matrix $A \in \mathbb{R}^{n,s}$ mit vollem Rang $s \leqslant n$,
- einem Zufallsvektor $\xi \in \mathbb{R}^n$ („Fehler") von standardisierten Zufallsvariablen (das heißt $\xi_k \in L^2(\mathbb{R}^n)$, $\mathbb{E}[\xi_k] = 0$ und $\mathbb{E}[\xi_k^2] = 1$ für alle $k \in \{1, \dots, n\}$),
- einem Beobachtungsvektor $X := A\gamma + \sqrt{v}\xi \in \mathbb{R}^n$.

Das Modell ist $\left(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), (\mathbb{P}_{(\gamma,v)})_{\gamma \in \mathbb{R}^s, v>0}\right)$, wobei $\mathbb{P}_{(\gamma,v)}$ die Verteilung von $A\gamma + \sqrt{v}\xi$ bezeichnet.

Bemerkung 6.2.2 Es gilt $\mathbb{E}_{(\gamma,v)}[X] = A\gamma$. ○

Wir wollen einen KQ-Schätzer $\hat{\gamma}$ zu X zu finden (welcher $\frac{1}{n}\sum_{k=1}^n (X_k - A\gamma)^2$ minimiert).

Beispiel 6.2.3 (Lineare Modelle) Die oben besprochene einfache lineare Regression ist ein lineares Modell mit $A = (\mathbb{1}, t) \in \mathbb{R}^{n,2}$, das heißt $s = 2$. Auch polynomielle Regression $X_k = \sum_{j=0}^d \gamma_j t_k^j + \sqrt{v}\xi_k$ für $k \in \{1, \dots, n\}$ ist ein lineares Modell mit $d = s - 1$ und $A := (1, t, t^2, \dots, t^d) \in \mathbb{R}^{n \times (d+1)}$, wobei $t^k := (t_1^k, \dots, t_n^k)^\intercal$ für $k \in \{1, \dots, d\}$ ist. Auch ein Modell mit mehreren Einflussgrößen $X_k = \gamma_0 \sin(\delta_k \cdot t_k) + \gamma_1 e^{u_k^2} + \gamma_2 \frac{1}{\delta_k^2 + t_k^2 + u_k^2 + 1} + \sqrt{v}\xi_k$ ist ein lineares Modell mit $A = \left(\sin(\delta_k \cdot t_k), e^{u_k^2}, \frac{1}{\delta_k^2 + t_k^2 + u_k^2 + 1}\right)_{k=1}^n$. ◇

Bemerkung 6.2.4 Für die Verteilung von $\xi_1, \dots, \xi_n$ kann man unterschiedliche Annahmen machen, z.B. $\mathcal{N}(0, E_n)$ oder $\xi_1 = \dots = \xi_n$ mit $\mathbb{P}(\xi = 1) = \mathbb{P}(\xi = -1) = \frac{1}{2}$. ○

Notation. Der Teilraum $L = L(A) := \mathrm{Bild}(A) = \{A\gamma : \gamma \in \mathbb{R}^n\} \subset \mathbb{R}^n$ ist $\mathrm{rang}(A) = s$-dimensional.

Wie vorher wollen wir X orthogonal auf L projizieren. Sei $\pi_L \in \mathbb{R}^{n,n}$ die zugehörigen Projektionsmatrix. Da $\mathcal{A}\colon \mathbb{R}^s \to L \subset \mathbb{R}^n$, $x \mapsto Ax$ bijektiv ist, existiert genau ein $\hat{\gamma} \in \mathbb{R}^s$ mit $A\hat{\gamma} = \pi_L X$.

Wie sieht π_L aus und können wir eine Formel für $\hat{\gamma}$ aufschreiben?

Lemma 6.2.5 (KQ-Schätzer für lineare Modelle)
Die Matrix $A^\intercal A \in \mathbb{R}^{s,s}$ ist invertierbar und $\pi_L = A(A^\intercal A)^{-1}A^\intercal$. Die einzige Lösung von $\pi_L X = A\hat{\gamma}$ ist der KQ-Schätzer $\hat{\gamma} = (A^\intercal A)^{-1}A^\intercal X$.

Beweis. (1) Angenommen, es gibt ein $c \in \mathbb{R}^s$ mit $A^\intercal Ac = 0$. Dann gilt $\|Ac\|^2 = (Ac)^\intercal Ac = c^\intercal A^\intercal Ac = 0$ und somit $Ac = 0$. Da A vollen Rang hat, folgt daraus $c = 0$.

(2) Für $X \in \mathbb{R}^n$ gilt

$$A(A^\intercal A)^{-1}A^\intercal X \in L$$

und für $y \in \mathbb{R}^s$ gilt

$$\langle X - \pi_L X, Ay \rangle = \langle X - A(A^\mathsf{T} A)^{-1} A^\mathsf{T} X, Ay \rangle = \langle A^\mathsf{T} X - \underbrace{A^\mathsf{T} A(A^\mathsf{T} A)^{-1}}_{=\mathrm{id}_{s,s}} A^\mathsf{T} X, y \rangle$$
$$= \langle \underbrace{A^\mathsf{T} X - A^\mathsf{T} X}_{=0}, y \rangle = 0$$

und somit $X - \pi_L X \perp L$. Also folgt $A\hat{\gamma} = A(A^\mathsf{T} A)^{-1} A^\mathsf{T} X = \pi_L X$. □

Satz 6.2.1: Erwartungstreue Schätzer im linearen Modell

In der obigen Situation gilt

1. $\hat{\gamma} = (A^\mathsf{T} A)^{-1} A^\mathsf{T} X$ ist ein erwartungstreuer Schätzer für γ.
2. Sei τ eine lineare reelle Funktion von γ, also $\tau(\gamma) = c^\mathsf{T}\gamma$ für ein $c \in \mathbb{R}^s$ und alle $\gamma \in \mathbb{R}^s$. Dann ist $T := \langle c, \hat{\gamma} \rangle$ ein erwartungstreuer Schätzer für τ und es gilt $T = a^\mathsf{T} X$ mit $a = A(A^\mathsf{T} A)^{-1} c$.
3. Sind zusätzlich die $\xi_1, \ldots, \xi_n$ unkorreliert, das heißt $\mathbb{E}_{(\gamma,v)}[\xi_i \xi_j] = 0$ für alle $i, j \in \{1, \ldots, n\}$ mit $i \neq j$, dann hat T minimale Varianz unter allen linearen erwartungstreuen Schätzern für τ („T is BLUE - best linear unbiased estimator").
4. Unter den Voraussetzungen von 3 und $n > s$ ist die (korrigierte) Stichprobenvarianz

$$V^* := \frac{1}{n-s} |X - \pi_L X|^2 = \frac{1}{n-s} \left(|X|^2 - |\pi_L X|^2 \right) = \frac{1}{n-s} |X - A\hat{\gamma}|^2$$

ein erwartungstreuer Schätzer für v.

Beweis. 1 Es gilt

$$\mathbb{E}_{(\gamma,v)}[\hat{\gamma}] = (A^\mathsf{T} A)^{-1} A^\mathsf{T} \, \mathbb{E}_{(\gamma,v)}[X] = (A^\mathsf{T} A)^{-1} A^\mathsf{T} A\gamma = \gamma.$$

2 Es gilt

$$\mathbb{E}_{(\gamma,v)}[T] = \mathbb{E}_{(\gamma,v)}[c^\mathsf{T} \hat{\gamma}] = c^\mathsf{T} \, \mathbb{E}_{(\gamma,v)}[\hat{\gamma}] = c^\mathsf{T} \gamma = \tau(\gamma)$$

sowie

$$T = c^\mathsf{T} \hat{\gamma} = c^\mathsf{T} (A^\mathsf{T} A)^{-1} A^\mathsf{T} X = \left(A(A^\mathsf{T} A)^{-1} c \right)^\mathsf{T} X = a^\mathsf{T} X.$$

12.07.2022

3 Sei $S\colon \left(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n)\right) \to \left(\mathbb{R}, \mathcal{B}(\mathbb{R})\right)$ ein linearer erwartungstreuer Schätzer für τ. Dann existiert ein $b \in \mathbb{R}^n$ mit $S = \langle b, X \rangle$ und es gilt

$$\langle b, A\gamma \rangle \overset{6.2.2}{=} \mathbb{E}_{(\gamma,v)}[b^\mathsf{T} X] = \mathbb{E}_{(\gamma,v)}[S] = \tau(\gamma)$$
$$= c^\mathsf{T} \gamma = \mathbb{E}_{(\gamma,v)}[T] = a^\mathsf{T} \, \mathbb{E}_{(\gamma,v)}[X] = \langle a, A\gamma \rangle$$

und somit $\langle b - a, A\gamma \rangle = 0$ für alle $\gamma \in \mathbb{R}^s$, also $b - a \perp L$. Da $a \in L$ folgt $a = \pi_L b$. Dann folgt mit dem Satz von Pythagoras $|a| \leqslant |b|$, genauer $|a| = |\pi_L b| \leqslant \|\pi_L\| |b| = |b|$, wobei $\| \cdot \|$ die Operatornorm ist (orthogonale Matrizen haben immer Einheitsoperatornorm).

Mit $\vartheta = (\gamma, v)$ gilt

$$\begin{aligned}\mathbb{V}_\vartheta[S] - \mathbb{V}_\vartheta[T] &= \mathbb{E}_\vartheta\left[|\langle b, X\rangle - \mathbb{E}_\vartheta[\langle b, X\rangle]|^2\right] - \mathbb{E}_\vartheta\left[|\langle a, X\rangle - \mathbb{E}_\vartheta[\langle a, X\rangle]|^2\right] \\ &\overset{\mathbb{E}_\vartheta[X]=A\gamma}{=} \mathbb{E}_\vartheta\left[\langle b, X - A\gamma\rangle^2 - \langle a, X - A\gamma\rangle^2\right] \\ &= \mathbb{E}_\vartheta\left[b^\top (X - A\gamma)(X - A\gamma)^\top b - a^\top (X - A\gamma)(X - A\gamma)^\top a\right] \\ &\overset{X - A\gamma = \sqrt{v}\xi}{=} v\big(b^\top \underbrace{\mathbb{E}_\vartheta[\xi\xi^\top]}_{=\mathrm{id}_n} b - a^\top \mathbb{E}_\vartheta[\xi\xi^\top] a\big) = v\big(|b|^2 - |a|^2\big) \geqslant 0\end{aligned}$$

(Selber überlegen: allgemeinere Bedingung als Unkorreliertheit unter der die Aussage erhalten bleibt.)

4 **Darstellung der Stichprobenvarianz.** Die Formel $|X - \pi_L X|^2 = |X|^2 - |\pi_L X|^2$ folgt direkt aus dem Satz von PYTHAGORAS. Mit Lemma 6.2.5 folgt $|X - \pi_L X|^2 = |X - A\hat{\gamma}|^2$.

Erwartungstreue. Seien $v_1, \dots, v_n$ eine Orthonormalbasis von $\mathbb{R}^n$, sodass $v_1, \dots, v_s$ eine Orthonormalbasis von L ist. Sei Γ die (orthonormale) Matrix mit den Spalten $v_1, \dots, v_n$. Dann bildet Γ den Raum $\{x \in \mathbb{R}^n : x_{s+1} = \dots = x_n = 0\}$ bijektiv auf L ab. Die darstellende (bezüglich der Basis $v_1, \dots, v_n$) Matrix der orthogonalen Projektion auf L ist $M_s := \begin{pmatrix} \mathrm{id}_s & 0 \\ 0 & 0 \end{pmatrix} \in \mathbb{R}^{n,n}$. Daher gilt $\pi_L = \Gamma M_s \Gamma^\top$. Sei $\eta := \Gamma^\top \xi \in \mathbb{R}^n$. Dann gilt

$$\begin{aligned}(n-s)V^* = |X - \pi_L X|^2 &= \big|\cancel{A\gamma} + \sqrt{v}\xi - \pi_L\big(\cancel{A\gamma} + \sqrt{v}\xi\big)\big|^2 \overset{(\star)}{=} v\,|\xi - \pi_L \xi|^2 \\ &\overset{(\dagger)}{=} v\,\big|\Gamma^\top(\xi - \pi_L\xi)\big|^2 = v|\eta - M_s\eta|^2 = v \sum_{k=s+1}^{n} \eta_k^2,\end{aligned} \tag{$\diamond$}$$

wobei wir in $(\star)$ benutzen, dass $A\gamma \in L$ und deswegen $\pi_L A\gamma = A\gamma$ und in $(\dagger)$, dass orthogonale Abbildungen die $|\cdot|$-Norm erhalten.

Weiter gilt für $k \in \{1, \dots, n\}$

$$\mathbb{E}_\vartheta[\eta_k^2] = \mathbb{E}_\vartheta[\langle v_k, \xi\rangle^2] = v_k\, \mathbb{E}_\vartheta[\xi\xi^\top] v_k^\top = |v_k|^2 = 1$$

und somit

$$(n-s)\,\mathbb{E}_\vartheta[V^*] = (n-s)v.$$

□

6.3 Das lineare Gauß-Modell

Wir nehmen nun zusätzlich an, dass ξ $\mathcal{N}(0, E_n)$-verteilt ist mit $s \leqslant n$.

Satz 6.3.1: Verallgemeinerter Satz von Student [2, 12.17]

Im linearen Gauß-Modell gelten (bezüglich $\mathbb{P}_{\gamma,v}$)

1. $\hat{\gamma}$ ist $\mathcal{N}\left(\gamma, v(A^\mathsf{T} A)^{-1}\right)$-verteilt.

Wenn $s < n$ ist, dann gelten

2. $\frac{n-s}{v} V^*$ ist χ^2_{n-s}-verteilt.
3. $\frac{1}{v}|A(\hat{\gamma}-\gamma)|^2 = \frac{1}{v}\left|\pi_L X - \mathbb{E}_{(\gamma,v)}[X]\right|^2$ ist χ^2_s-verteilt und unabhängig von V^*. Insbesondere ist $\frac{|A(\hat{\gamma}-\gamma)|^2}{sV^*}$ ist $\mathcal{F}_{s,n-s}$-verteilt.
4. Ist $H \subset L$ ein Teilraum mit Dimension $r < s$ und ist $A\gamma \in H$, so ist $\frac{1}{v}|\pi_L X - \pi_H X|^2$ χ^2_{s-r}-verteilt und unabhängig von V^*. Insbesondere ist die Fisher-Statistik
$$F_{H,L} := \frac{\frac{|\pi_L X - \pi_H X|^2}{(s-r)v}}{\frac{(n-s)V^*}{(n-s)v}} = \frac{n-s}{s-r}\frac{|\pi_L X - \pi_H X|^2}{|X - \pi_L X|^2} = \frac{|A\hat{\gamma} - \pi_H X|^2}{(s-r)V^*}$$
$\mathcal{F}_{s-r,n-s}$-verteilt (wegen 2).

Merke, dass Satz 3.0.2 2 ein Spezialfall (mit $s = 1$, $A = \mathbb{1}$ und somit ist $\hat{\gamma}$ der empirische Mittelwert der X_k) von 1 und 2 ist.

Beweis. 1 Da $X = A\gamma + \sqrt{v}\xi$ gilt, ist X $\mathcal{N}(A\gamma, v\,\mathrm{id}_n)$-verteilt. Da $\hat{\gamma} = (A^\mathsf{T} A)^{-1}A^\mathsf{T} X$ eine lineare Transformation von X ist, ist auch $\hat{\gamma}$ normalverteilt mit Erwartungswert $\mathbb{E}_\vartheta[\hat{\gamma}] = \gamma$. Die Kovarianzmatrix ist
$$\begin{aligned}\mathbb{E}_\vartheta[(\hat{\gamma}-\gamma)(\hat{\gamma}-\gamma)^\mathsf{T}] &= \mathbb{E}_\vartheta[(A^\mathsf{T} A)^{-1}A^\mathsf{T}(X - A\gamma)(X - A\gamma)^\mathsf{T} A(A^\mathsf{T} A)^{-1}] \\ &= v\,\mathbb{E}_\vartheta[\xi\xi^\mathsf{T}](A^\mathsf{T} A)^{-1} = v(A^\mathsf{T} A)^{-1}.\end{aligned}$$

2 - 4 Sei $H \subset L$ ein Teilraum mit Dimension $r < s$ und $v_1, \ldots, v_n$ eine Orthonormalbasis von $\mathbb{R}^n$ mit $\mathrm{span}(v_1, \ldots, v_r) = H$ und $\mathrm{span}(v_1, \ldots, v_s) = L$. Sei Γ wieder die Matrix mit den Spalten $v_1, \ldots, v_n$ und $\eta := \Gamma^\mathsf{T}\xi$. Dann ist η $\mathcal{N}(0, \mathrm{id}_n)$-verteilt. Daraus folgt 2 mit ($\diamond$) und der Definition einer χ^2_{n-s}-Verteilung.

Wieder gilt $\pi_L = \Gamma M_s \Gamma^\mathsf{T}$ und analog $\pi_H = \Gamma M_r \Gamma^\mathsf{T}$ also folgt, dass
$$|\pi_L\xi - \pi_H\xi|^2 = |\Gamma(M_s - M_r)\Gamma^\mathsf{T}\xi|^2 = |(M_s - M_r)\eta|^2 = \sum_{k=r+1}^{s} \eta_k^2$$
χ^2_{s-r}-verteilt und wegen ($\diamond$) unabhängig von V^*.

Fall 1: $H = \{0\}$. Dann ist $r = 0$, also ist $|\pi_L\xi|^2$ χ^2_s-verteilt und unabhängig von V^*. Weiter gilt
$$A(\hat{\gamma} - \gamma) = \pi_L(X - A\gamma) = \pi_L\sqrt{v}\xi = \sqrt{v}\pi_L\xi$$
und somit $\frac{1}{v}|A(\hat{\gamma}-\gamma)|^2 = |\pi_L\xi|^2$, also folgt 3.

Fall 1: $H \neq \{0\}$**.** Sei $A\gamma \in H$. Dann gilt

$$\begin{aligned}\pi_L X - \pi_H X &= \pi_L(A\gamma + \sqrt{v}\xi) - \pi_H(A\gamma + \sqrt{v}\xi)\\ &= \cancel{A\gamma} + \sqrt{v}\pi_L\xi \cancel{- A\gamma} - \sqrt{v}\pi_H\xi = \sqrt{v}(\pi_L\xi - \pi_H\xi)\end{aligned}$$

also folgt 4. □

13.07.2022

Seien im Folgenden $s < n$.

Korollar 6.3.1 (Konfidenzbereiche im linearen Gauss-Modell)
Sei $\alpha \in (0,1)$.

1 *Ein Konfidenzellipsoid für γ zu Irrtumsniveau α ist*

$$C(\cdot) = \{\gamma \in \mathbb{R}^s : |A(\gamma - \hat{\gamma})|^2 < s f_{s,n-s;1-\alpha} V^*\},$$

wobei $f_{s,n-s;1-\alpha}$ das α-Fraktil (bzw. $(1-\alpha)$-Quantil) der $\mathcal{F}_{s,n-s}$-Verteilung ist.

Man kann $<$ durch $\leqslant$ ersetzen, da $\mathcal{F}$ eine stetige Verteilung ist.

2 *Ein Konfidenzintervall für $\tau(\gamma) = c^\mathsf{T}\gamma$, wobei $c \in \mathbb{R}^s$, zum Irrtumsniveau α ist*

$$C(\cdot) = \left[c^\mathsf{T}\hat{\gamma} - \delta\sqrt{V^*}, c^\mathsf{T}\hat{\gamma} + \delta\sqrt{V^*}\right],$$

wobei $\delta = t_{n-s;1-\frac{\alpha}{2}}\sqrt{c^\mathsf{T}(A^\mathsf{T}A)^{-1}c}$ und $t_{n-s;1-\frac{\alpha}{2}}$ das $\frac{\alpha}{2}$-Fraktil der t_{n-s}-Verteilung ist.

Hier müssen wir das abgeschlossene Intervall wählen, weil die Aussage sonst für $c = 0$ falsch ist.

3 *Sind $q_- := \chi^2_{n-s;\frac{\alpha}{2}}$ und $q_+ := \chi^2_{n-s;1-\frac{\alpha}{2}}$ die $\frac{\alpha}{2}$ bzw. $1-\frac{\alpha}{2}$-Quantile von χ^2_{n-s}, so ist*

$$C(\cdot) = \left((n-s)\frac{V^*}{q_+}, (n-s)\frac{V^*}{q_-}\right)$$

ein Konfidenzintervall für v zum Niveau α.

Man kann C durch das entsprechende abgeschlossene Intervall ersetzen, da die χ^2-Verteilung eine stetige Dichte hat.

Beweis. 1 Folgt aus Satz 6.3.1 3, denn $\mathbb{P}\left(\frac{1}{sV^*}|A(\gamma - \hat{\gamma})|^2 < f_{s,n-s,1-\alpha}\right) = 1-\alpha$ nach Definition des Fraktils.

2 Sei ohne Beschränkung der Allgemeinheit $c \neq 0$, denn für $c = 0$ ist $\tau \equiv 0$ und $C(\cdot) = \{0\}$ offenbar richtig. Nach Satz 6.3.1 1 ist $\hat{\gamma}$ $\mathcal{N}\left(\gamma, v(A^\mathsf{T}A)^{-1}\right)$-verteilt. Also ist $Z := c^\mathsf{T}\hat{\gamma}$ normalverteilt mit $\mathbb{E}_\vartheta[Z] = c^\mathsf{T}\,\mathbb{E}_\vartheta[\hat{\gamma}] = c^\mathsf{T}\gamma = \tau(\gamma)$ und

$$\mathbb{E}_\vartheta\left[(Z - \mathbb{E}_\vartheta[Z])(Z - \mathbb{E}_\vartheta[Z])^\mathsf{T}\right] = c^\mathsf{T}\,\mathbb{E}_\vartheta\left[(\hat{\gamma} - \gamma)(\hat{\gamma} - \gamma)^\mathsf{T}\right]c \overset{6.3.1\,(1)}{=} c^\mathsf{T} v(A^\mathsf{T}A)^{-1}c.$$

Somit ist Z $\mathcal{N}\left(c^\mathsf{T}\gamma, vc^\mathsf{T}(A^\mathsf{T}A)^{-1}c\right)$-verteilt. Somit ist $Z^* := \frac{Z - \mathbb{E}_\vartheta[Z]}{\sqrt{\mathbb{V}_\vartheta[Z]}} = \frac{c^\mathsf{T}(\hat{\gamma} - \gamma)}{\sqrt{v}\sqrt{c^\mathsf{T}(A^\mathsf{T}A)^{-1}c}}$ standardnormalverteilt. *Sind* Z^* und V^* unabhängig, dann sind es auch Z^* und $(n-$

$s)\frac{V^*}{v}$, welche χ^2_{n-s}-verteilt ist. Also ist $\frac{Z^*}{\sqrt{\frac{V^*}{v}}}$ t_{n-s}-verteilt. Dann folgt

$$\begin{aligned}\mathbb{P}_\vartheta\big(\tau(\gamma)\in C(\cdot)\big) &= \mathbb{P}_\vartheta\big(c^\mathsf{T}\hat\gamma-\delta\sqrt{V^*}\leqslant c^\mathsf{T}\gamma\leqslant c^\mathsf{T}\hat\gamma+\delta\sqrt{V^*}\big)\\ &= \mathbb{P}_\vartheta\big(|c^\mathsf{T}(\hat\gamma-\gamma)|\leqslant\delta\sqrt{V^*}\big)\\ &= \mathbb{P}_\vartheta\left(\frac{|Z-\tau(\gamma)|}{\sqrt{V^*}}\leqslant\delta\right)\\ &= \mathbb{P}_\vartheta\left(\frac{|Z^*|\sqrt{vc^\mathsf{T}(A^\mathsf{T}A)^{-1}c}}{\sqrt{V^*}}\leqslant\delta\right)\\ &= \mathbb{P}_\vartheta\left(\frac{|Z^*|}{\sqrt{\frac{V^*}{v}}}\leqslant\frac{\delta}{\sqrt{c^\mathsf{T}(A^\mathsf{T}A)^{-1}c}}\right)\\ &= \mathbb{P}_\vartheta\left(\left|\frac{Z^*}{\sqrt{\frac{V^*}{v}}}\right|\leqslant t_{n-s;1-\frac{\alpha}{2}}\right)=1-\alpha.\end{aligned}$$

Es verbleibt, die Unabhängigkeit von Z^* und V^* zu zeigen. Aus ($\diamond$) folgt $\frac{n-s}{v}V^*=\sum_{k=s+1}^n\eta_k^2$. Es gilt

$$\begin{aligned}A\hat\gamma=\pi_LX=\pi_L(A\gamma+\sqrt{v}\xi)=A\gamma+\sqrt{v}\pi_L\xi&=A\gamma+\sqrt{v}\Gamma M_s\underbrace{\Gamma^\mathsf{T}\xi}_{=\eta}\\ &=A\gamma+\sqrt{v}\Gamma(\eta_1,\dots,\eta_s,0,\dots,0)^\mathsf{T}.\end{aligned}$$

Es tauchen nur deterministische Größen auf, bis auf $\eta_1,\dots,\eta_s$. In V^* tauchen nur $\eta_{s+1},\dots,\eta_n$ auf, also sind V^* und $A\hat\gamma$ unabhängig. Da A bijektiv und bimessbar (messbar und die Umkehrabbildung ist messbar) ist, sind auch $\hat\gamma$ und V^* unabhängig. Somit sind Z^* und V^* unabhängig.

3. Einfach. □

Korollar 6.3.2 (Hypothesentests im linearen Gauss-Modell)
Sei $\alpha\in(0,1)$.

1. ***t-Test für $c^\mathsf{T}\gamma=m_0$.*** *Der (zweiseitige) t-Test der Hypothese $H_0: c^\mathsf{T}\gamma=m_0$ gegen $H_1: c^\mathsf{T}\gamma\neq m_0$ mit Irrtumsniveau α hat den Ablehungsbereich*

$$\left\{\left|c^\mathsf{T}\hat\gamma-m_0\right|>t_{n-s,1-\frac{\alpha}{2}}\sqrt{c^\mathsf{T}(A^\mathsf{T}A)^{-1}c}\sqrt{V^*}\right\},$$

und das gilt analog für den einseitigen Test $H_0: c^\mathsf{T}\gamma\leqslant m_0$ gegen $H_1: c^\mathsf{T}\gamma\geqslant m_0$.

2. ***F-Test von $H_0: A\gamma\in H$.*** *Sei $H\subset L$ ein r-dimensionaler Unterraum. Ist $F_{H,L}$ wie in Satz 6.3.1 definiert, so definiert der Ablehnungsbereich*

$$\{F_{H,L}>f_{s-r,n-s,1-\alpha}\}$$

einen Test zum Niveau α für die Hypothese $H_0: A\gamma\in H$ gegen die Alternative $H_1: A\gamma\notin H$.

3. ***χ^2-Test für die Varianz v.*** *Für $v_0>0$ definiert der Ablehnungsbereich*

$$\{(n-s)V^*>v_0\chi^2_{n-s;1-\alpha}\}$$

einen Test zum Niveau α für $H_0: v\leqslant v_0$ gegen $H_1: v>v_0$.

Beweis. 1 und 2 sind klar durch Bemerkung 4.1.11.

3 Sei $0 < v \leqslant v_0$. Dann gilt

$$\mathbb{P}_\vartheta\big((n-s)V^* > v_0\chi^2_{n-s;1-\alpha}\big) = \mathbb{P}_\vartheta\left((n-s)\frac{V^*}{v} > \underbrace{\frac{v_0}{v}}_{\geqslant 1}\chi^2_{n-s;1-\alpha}\right)$$
$$\leqslant \mathbb{P}_\vartheta\left(\underbrace{(n-s)\frac{V^*}{v}}_{\sim\chi^2_{n-s}} > \chi^2_{n-s;1-\alpha}\right) = \alpha.$$

□

Beispiel 6.3.3 (Motorleistung (Fortsetzung))

2 Wir testen $H_0 : \gamma_1 \geqslant 10$ zum Niveau $\alpha = 0.05$. Für $H_0 : \binom{0}{1} \cdot \gamma \geqslant 10$ ist der Ablehnungsbereich $\left\{10 - \hat{\gamma}_1 > t_{6,0.95}\sqrt{\frac{1}{n}\frac{1}{V(t)}V^*}\right\}$, wobei $t_{6,0.95} \approx 1.943$. Es ist $V(t) \approx 2.56$ sowie

$$V^* = \frac{1}{6}\sum_{k=1}^{8}(X_k - \gamma_0 - \gamma_1 t_k)^2 \approx 2.5$$

und somit $\sqrt{\frac{1}{n}\frac{1}{V(t)}V^*} \approx 0.35$. Da $10 - \hat{\gamma}_1 > 0.68$, wird H_0 abgelehnt.

3 Wir bestimmen einen Konfidenzbereich für γ_1 zum Irrtumsniveau $\alpha = 0.05$. Der Konfidenzbereich mit $c = (0,1)^\mathsf{T}$ ist $C(x) = \big(\hat{\gamma}_2 - \delta\sqrt{V^*}, \hat{\gamma}_1 + \delta\sqrt{V^*}\big)$, wobei $\delta = t_{6,0.975}\sqrt{\frac{1}{n}\frac{1}{V(t)}} \approx 2.447 \cdot 0.02$ ist. Also gilt

$$C(x) = \big(9.19 - 2.447 \cdot 0.35, 9.19 + 2.447 \cdot 0.35\big) \approx (8.33, 10.05).$$

◇

Beispiel 6.3.4 (Gauss-Produktmodell: Mittelwert schätzen)

Seien $s = 1$, $A := \mathbb{1} \in \mathbb{R}^n$ und $\gamma = m \in \mathbb{R}$, dann können wir $X = A\gamma + \sqrt{v}\xi$ schreiben, wobei ξ $\mathcal{N}(0, \mathrm{id}_n)$-verteilt sind, das heißt $X_k = m + \sqrt{v}\xi_k$ für $k \in \{1, \dots, n\}$ sind unabhängig $\mathcal{N}(m, v)$-verteilt.

Es gilt $\hat{\gamma} = \hat{m} = (A^\mathsf{T}A)^{-1}A^\mathsf{T}X$. Wegen $A^\mathsf{T}A = n$ folgt $\hat{m} = \frac{1}{n}\sum_{k=1}^n X_k$. Insbesondere ergibt sich aus Korollar 6.3.2 1 ein t-Test (mit $c = 1$) für die Hypothese $H_0 : m = m_0$ gegen die Alternative $H_1 : m \neq m_0$

◇

Beispiel 6.3.5 (Einfache lineare Regression)

Seien $s = 2$ sowie $A = (\mathbb{1}, t) \in \mathbb{R}^{n\times 2}$ für $t \in \mathbb{R}^n$, sodass $\mathbb{1}$ und t linear unabhängig sind. Ferner sei $V(t) := \frac{1}{n}\sum_{k=1}^n t_k^2 - M(t)^2$, wobei $M(t) := \frac{1}{n}\sum_{k=1}^n t_k$. Dann gilt

$$A^\mathsf{T}A = (\mathbb{1}, t)^\mathsf{T}(\mathbb{1}, t) = \begin{pmatrix} n & nM(t) \\ nM(t) & nM(t^2) \end{pmatrix}$$

und daher (cf. [2, Satz 6.1])

$$(A^\mathsf{T}A)^{-1} = \frac{1}{n^2M(t^2) - n^2M(t)^2}\begin{pmatrix} nM(t^2) & -nM(t) \\ -nM(t) & n \end{pmatrix} = \frac{1}{n}\frac{1}{V(t)}\begin{pmatrix} M(t^2) & -M(t) \\ -M(t) & 1 \end{pmatrix}$$

also

$$\begin{pmatrix} 0 & 1 \end{pmatrix}(A^\mathsf{T}A)^{-1}\begin{pmatrix} 0 \\ 1 \end{pmatrix} = \frac{1}{n}\frac{1}{V(t)}$$

und

$$\hat{\gamma} = (A^\mathsf{T} A)^{-1} A^\mathsf{T} X = \frac{1}{n} \frac{1}{V(t)} \begin{pmatrix} M(t^2) & -M(t) \\ -M(t) & 1 \end{pmatrix} \begin{pmatrix} \sum_{k=1}^n X_k \\ \sum_{k=1}^n t_k X_k \end{pmatrix}$$

◇

6.4 Varianzanalyse

19.07.2022

Siehe [2, Kap. 12.4], nicht relevant für die Juli-Prüfung.

Appendix

A Wiederholung: Gemeinsame Verteilung und Randverteilungen

25.05.2022

Sind X und Y Zufallsvariablen, so ist die Verteilung von (X, Y) die gemeinsame Verteilung und die Verteilung der einzelnen Variablen X und Y sind die Randverteilungen.

Aus der gemeinsamen Verteilung von (X, Y) kann man stets eindeutig die Randverteilungen von X und Y bestimmen. Sind die Randverteilungen gegeben, ist die gemeinsame Verteilung nicht eindeutig (bei Unabhängigkeit möglich).

Beispiel A.0.1 (Nichteindeutigkeit der gemeinsamen Verteilung)
Seien X und Y gleichverteilt auf $[0, 1]$. Es gibt viele Möglichkeiten, eine gemeinsame Verteilung zu wählen:

- ist $X(\omega) = Y(\omega)$, so gilt $\mathbb{P}[X \leqslant s, Y \leqslant t] = \mathbb{P}[X \leqslant \min(s,t)] = \min\{s,t\}$ für $s, t \in (0, 1)$.
- ist $X(\omega) = 1 - Y(\omega)$, so ist $\mathbb{P}[X \leqslant s, Y \leqslant t] = \mathbb{P}[1 - t \leqslant X \leqslant s] = \max\{0, s + t - 1\}$ für $s, t \in (0, 1)$.
- sind X und Y unabhängig, so ist $\mathbb{P}[X \leqslant s, Y \leqslant t] = \mathbb{P}[X \leqslant s]\mathbb{P}[Y \leqslant t] = st$ für $s, t \in (0, 1)$. ◇

Definition A.0.2 (Randverteilung)
Seien X und Y reelle Zufallsvariablen, deren gemeinsame Verteilung bekannt ist. Die Randverteilungen sind

$$\mathbb{P}[X \in A] = \mathbb{P}[(X, Y) \in A \times \mathbb{R}] \qquad \text{und} \qquad \mathbb{P}[Y \in A] = \mathbb{P}[(X, Y) \in \mathbb{R} \times A].$$

In diskreten Fall (wenn X und Y $\mathbb{N}$-wertig sind), gilt

$$\mathbb{P}[X = k] = \sum_{j \in \mathbb{N}} \mathbb{P}[X = k, Y = j] \qquad \text{und} \qquad \mathbb{P}[Y = j] = \sum_{k \in \mathbb{N}} \mathbb{P}[X = k, Y = j].$$

Im absolut stetigen Fall mit gemeinsamer Dichte $f_{(X,Y)}$ gilt

$$f_X(x) = \int_{\mathbb{R}} f_{(X,Y)}(x, y)\,\mathrm{d}y \qquad \text{und} \qquad f_Y(y) = \int_{\mathbb{R}} f_{(X,Y)}(x, y)\,\mathrm{d}x.$$

Sind X und Y unabhängig, so gilt

$$\mathbb{P}[(X, Y) \in A \times B] = \mathbb{P}[X \in A] \times \mathbb{P}[Y \in B].$$

Wiederholung: Satz von Bayes und bedingte Verteilung

Für $A, B \subset \Omega$ mit $\mathbb{P}[A] > 0$ und $\mathbb{P}[B] > 0$ gilt

$$\mathbb{P}[A \mid B] = \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]} = \frac{\mathbb{P}[B \mid A]\mathbb{P}[A]}{\mathbb{P}[B]}.$$

Ist $\bigcup_{j\in I} A_j = \Omega$ eine Zerlegung, gilt auch

$$\mathbb{P}[A_k \mid B] = \frac{\mathbb{P}[B \mid A_k]\mathbb{P}[A_k]}{\sum_{j\in I} \mathbb{P}[B \mid A_j]\mathbb{P}[A_j]}.$$

Für Zufallsvariablen X und Y ist

$$\mathbb{P}[Y \in A \mid X \in B] = \frac{\mathbb{P}[Y \in A, X \in B]}{\mathbb{P}[X \in B]}$$

sozusagen der Quotient aus der gemeinsamen Verteilung und der Randverteilung.

Nimmt die diskrete Zufallsvariable Y die Werte $(y_i)_{i\in I}$ an, so gilt ($\sum_{i\in I} \mathbb{P}[Y = y_i] = 1$ und)

$$\mathbb{P}[Y = y_i \mid X \in B] = \frac{\mathbb{P}[X \in B \mid Y = y_i]\mathbb{P}[Y = y_i]}{\sum_{j\in I} \mathbb{P}[X \in B \mid Y = y_j]\mathbb{P}[Y = y_j]}.$$

Sind X und Y absolut stetig bezüglich des Lebesgue-Maßes, so ist die bedingte Dichte von X gegeben Y

$$f_{X|Y}(x \mid y) := \frac{f_{(X,Y)}(x, y)}{f_Y(y)},$$

also sozusagen der Quotient der gemeinsamen Dichte und der Randdichte. Analog zum Satz von Bayes folgt

$$f_{X|Y}(x \mid y) = \frac{f_{Y|X}(y \mid x) f_X(x)}{\int_{\mathbb{R}} f_{Y|X}(y \mid z) f_X(z)\,\mathrm{d}z}.$$

B Wiederholung: Bedingte Wahrscheinlichkeit und Erwartungswert

Was ist $\mathbb{E}[X \mid Y] := \mathbb{E}[X \mid \sigma(Y)]$?

Wenn Y diskret ist, existieren $(p_i)_{i\in I}$ mit $\mathbb{P}(Y = y_i) > 0$ für alle $i \in I$ und $\sum_{i\in I} \mathbb{P}(Y = y_i) = 1$. Dann gilt $\mathbb{E}[X \mid Y](\omega) = \mathbb{E}[X \mid Y = y_i]$ wenn $Y(\omega) = y_i$.

Beispiel B.0.1 (Münzwurf)
Wir werfen eine faire Münze unabhängig. Seien Y das Ergebnis des ersten Wurfes und X die Anzahl der Versuch bis zum ersten Mal Kopf. Dann ist

$$\mathbb{E}[X \mid Y](\omega) = \begin{cases} 1, & \text{wenn } Y(\omega) = \text{Kopf}, \\ 1 + \mathbb{E}[X] = 3, & \text{wenn } Y(\omega) = \text{Zahl}. \end{cases}$$

◇

Definition B.0.2 (Bedingte Erwartung)
Seien $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum, $G \subset \mathcal{F}$ eine σ-Algebra und $X \in L^1(\Omega)$. Die Zufallsvariable X_0 sodass gilt

- X_0 ist G-messbar,
- $\mathbb{E}[X_0 \mathbb{1}_A] = \mathbb{E}[X \mathbb{1}_A]$ für alle $A \in G$,

heißt bedingte Erwartung von X gegeben G und wir schreiben $X_0 = \mathbb{E}[X \mid G]$. bedingte Erwartung

Lemma B.0.3 (Eigenschaften der bedingten Erwartung)

1. ***Linearität:*** $\mathbb{E}[aX + bY \mid G] = a\,\mathbb{E}[X \mid G] + b\,\mathbb{E}[Y \mid G]$,
2. ***Bekanntes rausziehen:*** *ist X G-messbar, so gilt* $\mathbb{E}[XY \mid G] = X \cdot \mathbb{E}[Y \mid G]$,
3. ***Unabhängigkeit:*** *sind X und Y unabhängig, so gilt* $\mathbb{E}[X \mid \sigma(Y)] = \mathbb{E}[X]$,
4. ***Konstanten:*** *ist $X = a \in \mathbb{R}$, dann ist* $\mathbb{E}[X \mid G] = a$
5. ***Monotonie:*** *ist $X \leqslant Y$, dann ist* $\mathbb{E}[X \mid Y] \leqslant \mathbb{E}[Y \mid G]$.
6. ***Jensensche Ungleichung:*** *ist $f\colon \mathbb{R} \to \mathbb{R}$ konvex und $f(X) \in L^1$, dann gilt* $f\big(\mathbb{E}[X \mid G]\big) \leqslant \mathbb{E}[f(X) \mid G]$.
7. ***Turmeigenschaft:*** *Seien $H, G \subset \mathcal{F}$ σ-Algebren mit $H \subset G$, dann gilt*

$$\mathbb{E}\big[\mathbb{E}[X \mid H] \mid G\big] = \mathbb{E}\big[\mathbb{E}[X \mid G] \mid H\big] = \mathbb{E}[X \mid H]$$

und insbesondere $\mathbb{E}\big[\mathbb{E}[X \mid G]\big] = \mathbb{E}[X]$.

Beispiel B.0.4 (Zweimaliger Würfelwurf)
Wir werfen zwei mal einen fairen Würfel. Seien X_j für $j \in \{1, 2\}$ das Ergebnis des j-ten Wurfs. Dann gilt

$$\mathbb{E}[X_1 + X_2 \mid X_1] \overset{1}{=} \mathbb{E}[X_1 \mid X_1] + \mathbb{E}[X_2 \mid X_1] \overset{2}{\underset{3}{=}} X_1 + \mathbb{E}[X_2] = X_1 + \frac{7}{2}. \tag{34}$$

Der Erwartungswert der erwarteten Augensumme bei Kenntnis des ersten Wurfs ist somit

$$\mathbb{E}\big[\,\mathbb{E}[X_1 + X_2 \mid X_1]\big] \overset{(34)}{=} \mathbb{E}\left[X_1 + \frac{7}{2}\right] = 7 = \mathbb{E}[X_1 + X_2].$$

◇

Letztlich ist $\mathbb{P}_\vartheta[A \mid \mathcal{B}]$ eine Zufallsvariable auf $\mathfrak{X}$ und es gilt $\mathbb{P}_\vartheta[X \in A \mid \sigma(T)] = \mathbb{E}_\vartheta[\mathbb{1}_A(X) \mid \sigma(T)]$ beziehungsweise

$$\mathbb{P}_\vartheta[A \mid \sigma(T)] = \mathbb{E}_\vartheta[\mathbb{1}_A \mid \sigma(T)].$$

Übersicht: Modelle, Schätzer und deren Eigenschaften

	$\mathfrak{X}$	$\rho(x,\vartheta)$	τ	$T_n(x)$	$\mathbb{E}$-treu	UMVU	M Reg.	M exp.	T reg.	$T=$ ML	Suff.	Vollst.	Konsist.
Bernoulli	$\{0,1\}$	$\vartheta^{n\hat{\mu}}(1-\vartheta)^{1-n\hat{\mu}}$	$\mathrm{id}_{[0,1]}$	$\hat{\mu}$	✓	✓	✗	✓, $\Theta=(0,1)$	✓	✓	✓	✓	✓
Binomial	$\{0,\dots,n\}$	$\binom{n}{x}\vartheta^x(1-\vartheta)^{n-x}$	$\mathrm{id}_{[0,1]}$	$\frac{1}{n}x$	✓	✓	✗	✓, $\Theta=(0,1)$	✓	✓	✓	✓	✓
Hypergeo.	$\{0,\dots,n\}$	$\frac{\binom{\vartheta}{x}\binom{N-\vartheta}{n-x}}{\binom{N}{n}}\mathbb{1}_{x\leqslant\vartheta}$	$\mathrm{id}_{\{0,\dots,N\}}$	$x\frac{N}{n}$	✓		✗	✗	/	✗			
			$\mathrm{id}_{\{0,\dots,N\}}$	$\begin{cases} N, & x=n \\ x\left\lfloor\frac{N+1}{n}\right\rfloor, & \text{sonst}\end{cases}$			✗	✗	/	✓			
Poisson	$\mathbb{N}_0$	$e^{-\vartheta}\frac{\vartheta^x}{x!}$	$\mathrm{id}_{(0,\infty)}$	$\hat{\mu}$	✓	✓	✓	✓	✓	✓	✓	✓	✓
			$e^{-3\vartheta}$	$(-2)^x$	✓		✓			/			
Geometrisch	$\mathbb{N}_{>0}$	$\vartheta(1-\vartheta)^{x-1}$	$\mathrm{id}_{[0,1]}$	$\frac{1}{x}$	✗		✗		/	✓			
	$\mathbb{N}_0$	$\vartheta(1-\vartheta)^{n\hat{\mu}}$	$\mathrm{id}_{[0,1]}$	$\hat{\mu}$							✓		
Game show	$[0,\infty)$	$\mathcal{U}_{[0,\vartheta]}$	$\mathrm{id}_{(0,\infty)}$	$x_{(n)}$	✗	✗	✗	✗	/	✓			✓
				$\frac{n+1}{n}x_{(n)}$	✓	✓							✓
				$2\hat{\mu}$	✓		✗	✗	/	✗			✓
				$e\left(\prod_{k=1}^n x_k\right)^{\frac{1}{n}}$	✓								✓
Exponentiell	$(0,\infty)$	$\vartheta e^{-\vartheta x}$	$\frac{1}{\mathrm{id}_{(0,\infty)}}$	$\hat{\mu}$	✓	✓	✓	✓	✓	/			✓
			id	$\frac{1}{\hat{\mu}}$	✗					✓			
Verschob. Exp.	$\mathbb{R}$	$e^{\vartheta-x}\mathbb{1}_{[\vartheta,\infty)}(x)$	$\mathrm{id}_{\mathbb{R}}$	$x_{(1)}$	✗	✗	✗	✗		✓			✓
Gauss I	$\mathbb{R}$	$\mathcal{N}_{m,\vartheta}$	$\sqrt{\mathrm{id}_{(0,\infty)}}$	$\sqrt{\frac{\pi}{2}}\frac{1}{n}\sum_{k=1}^n \lvert x_k-m\rvert$	✓				✓	/			
			$\mathrm{id}_{(0,\infty)}$	$\frac{1}{n}\sum_{k=1}^n(x_k-m)^2$	✓	✓	✓	✓	✓	✓	✓	✓	
Gauss II	$\mathbb{R}$	$\mathcal{N}_{\vartheta}$	$\mathrm{id}_{\mathbb{R}\times(0,\infty)}$	$\left(\hat{\mu},\frac{1}{n}\sum_{k=1}^n(x_k-\hat{\mu})^2\right)$	✗	✓	✓	(✓)		✓			
		$n>1$	$\mathrm{id}_{\mathbb{R}\times(0,\infty)}$	$\left(\hat{\mu},\frac{1}{n-1}\sum_{k=1}^n(x_k-\hat{\mu})^2\right)$	✓	✓	✓	(✓)		✗			
Gauss III	$\mathbb{R}$	$\mathcal{N}_{\vartheta,\sigma^2}$	$\mathrm{id}_{\mathbb{R}}$	$\hat{\mu}$	✓	✓	✓	✓	✓	✓	✓	✓	✓
Gleichverteilung	$\mathbb{R}$	$\mathcal{U}_{[\vartheta-\frac{1}{2},\vartheta+\frac{1}{2}]}$	$\mathrm{id}_{\mathbb{R}}$	$\frac{1}{2}\left(x_{(1)}+x_{(n)}\right)$	✓	$\nexists$	✗	✗	/	✓			
			$\mathrm{id}_{\mathbb{R}}$	$\hat{\mu}$	✓	$\nexists$	✗	✗	/	✓			✓
Laplace	$\mathbb{R}$	$\frac{1}{2}e^{\vartheta-x}$	$\mathrm{id}_{\mathbb{R}}$	$\mathrm{Median}(x_1,\dots,x_{n=2m})$	✓	✓	✓	✓	✓	✓	✓	✓	
Pareto	$[k,\infty)$	$z\mapsto\vartheta k^{\vartheta}z^{-\vartheta-1}$	$\mathrm{id}_{(0,\infty)}$	$\left(\frac{1}{n}\sum_{j=1}^n\ln(x_j)-\ln(k)\right)^{-1}$	✗	✗	✓	✗	✗?	✓			

Tabelle 4: Überblick über die statistischen Standardmodelle $M=\left(\mathfrak{X},\mathcal{F},(\mathbb{P}_\vartheta)_{\vartheta\in\Theta}\right)$ mit Kenngröße τ und Schätzer T für τ. Wir betrachten meist das n-fache Produktmodell von M für $n\geqslant 1$. Für diskrete Modelle ist $\mathcal{F}=\mathcal{P}(\mathfrak{X})$ und für stetige $\mathcal{F}=\mathcal{B}(\mathfrak{X})$. Letztlich ist $\hat{\mu}:=\frac{1}{n}\sum_{k=1}^n x_k$ der empirische Mittelwert. Ist M exponentiell bezüglich T, so sind M und T regulär und T ist UMVU, suffizient und vollständig.

Index

T

U

V

W

Literatur

[1] Breiman, Leo: Statistics: with a View Toward Applications. Houghton Mifflin, 1973, ISBN 9780395042328. https://books.google.de/books?id=xJsZAQAAIAAJ.

[2] Georgii, Hans Otto: Stochastik. de Gruyter, 3. Auflage, 2007. https://doi.org/10.1515/9783110215274.

[3] Rüschendorf, Ludger: Mathematische Statistik, Band 62 der Reihe Masterclass. Springer Spektrum Berlin, Heidelberg, 1. Auflage, 2014. https://doi.org/10.1007/978-3-642-41997-3.

[4] Scheutzow, Michael: Wahrscheinlichkeitstheorie I Skript.