Technische Universität Dresden

Skript:

Mathematische Statistik

Verfasser Franziska Kühn

Daten Prof. Dr. Dietmar Ferger
Wintersemester 2010/11
Hauptstudium

Inhaltsverzeichnis

1

Stochastische Modelle

1.1 Beispiel: Schätzung von Populationsgrößen

- Betrachte Teich mit unbekannter Anzahl N von Fischen. Problem: Schätze N.
- Vorgehen: M Fische fangen, markieren, wieder aussetzen. Später n Fische fangen, davon seien m markiert. Erwarte, dass folgende Bedingung gilt:

  $$\frac{m}{n} \approx \frac{M}{N} \tag{*}$$

  d.h. Anteil der markierten Fische im zweiten Fang ≈ Anteil der markierten Fische im Teich. Auflösen der „Gleichung" liefert plausiblen (naiven) Schätzwert

  $$\hat{N} := \left[\frac{M \cdot n}{m}\right]$$

  für N.

1.2 Beispiel: Qualitätskontrolle

- Produktion von bestimmten Bauteilen. Seien $p \in (0,1)$ unbekannter Ausschußanteil, $p_0 \in (0,1)$ vorgegebener zulässiger Ausschußanteil. Bei $p > p_0$: Produktionsprozess verbessern, bei $p \leq p_0$ ist Produktionsprozess okay.
- Naheliegendes Vorgehen: Prüfe n Bauteile, erhalte Liste $\underline{x}_n := (x_1, \ldots, x_n)$ mit $x_i = 1$ ($x_i = 0$), falls i-tes Bauteil defekt (intakt). Erwarte für den relativen Anteil

  $$r_n = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i$$

  der defekten Bauteile in der Strichprobe $x_n = (x_1, \ldots, x_n)$: $r_n \approx p$.
- Somit eine mögliche Entscheidungsvorschrift: Falls $r_n > p_0$, so entscheide, dass $H_0 : p > p_0$ richtig ist, ansonsten Entscheidung für $H_1 : p \leq p_0$.

In beiden Beispielen: Beobachtetes Ergebnis (m bzw. r_n) ist zufällig, also Schätzung bzw. Entscheidung kann falsch sein. Daher Ziel: Mathematische Beschreibung von Zufallsexperimenten wie in 1.1 und 1.2. Dies ermöglicht eine objektive Beurteilung des Risikos falscher Entscheidungen.

Fortsetzung Beispiel 1.1 Auftretende Größen:

N	M,n	m
nicht zufällig unbekannt interessierende Größe (Parameter)	nicht zufällig bekannt frei wählbare Parameter	zufällig bekannt zufällige Beobachtung

Modellierung: Wir fassen m als Realisierung einer Zufallsvariablen über einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$, d.h. $m = X(w)$ für $w \in \Omega$. Damit beschreibt X die zufällige Anzahl der markierten Fische im zweiten Fang. Sinnvolle Verteilung von X gemäß Laplace-Ansatz:

$$\mathbb{P}[X = m] = \frac{\binom{M}{m} \cdot \binom{N-M}{n-m}}{\binom{N}{n}}$$

Beachte:

(i). Verteilung $\mathbb{P}_X := \mathbb{P} \circ X^{-1}$ von X hängt von N ab.

(ii). Jedes $N \geq M$ ist möglich.

(iii). Ein $N \geq M$ ist das wahre N.

(iv). Es liegt eine Realisierung $X(w) = m$ von X vor (auf Grund des in der Realität ausgeführten Zufallsexperiments).

(v). Auf der Basis von m ist N zu schätzen.

1.3 Definition

Seien $N, M, n \in \mathbb{N}$ mit $M, n \leq N$. Die diskrete Wahrscheinlichkeitsverteilung

$$\text{Hyp}(N, M, n)(\{m\}) := \frac{\binom{M}{m} \cdot \binom{N-M}{n-m}}{\binom{N}{n}}$$

wobei

$$\max(0, n - N + M) \leq m \leq \min(m, M)$$

heißt hypergeometrische Verteilung zu (N, M, n).

Fortsetzung Beispiel 1.2 Auftretende Größen:

p	n, p_0	$\underline{x}_n = (x_1, \ldots, x_n) \in \{0,1\}^n$
nicht zufällig unbekannt interessierender Parameter	nicht zufällig bekannt (frei wählbare) Parameter	zufällig bekannt beobachtetes Datum

Modellierung: Fassen $\underline{x}_n$ auf als Realisierung der Zufallsvariablen $\underline{X}_n = (X_1, \ldots, X_n)$ mit $X_1, \ldots, X_n$ iid (independant and identically distributed), $X_i \sim \beta_p :=$ Bin(1,p). Unabhängigkeitsannahme nur sinnvoll bei sehr großer Produktion.

$$\mathbb{P}_{\underline{X}_n} = \bigotimes_{i=1}^{n} \mathbb{P}_{X_i} = \bigotimes_{i=1}^{n} \beta_p := (\beta_p)^n$$

Beachte:

(i). Verteilung von $\underline{X}_n$ hängt von p ab.

(ii). Jedes $p \in (0,1)$ ist möglich.

(iii). Ein $p \in (0,1)$ ist das wahre p.

(iv). Es liegt eine Realisierung $\underline{X}_n(w) = \underline{x}_n$ von $\underline{X}_n$ vor.

(v). Auf Basis von $\underline{x}_n$ ist eine Aussage über das unbekannte p zu treffen.

Gemeinsame Struktur in Beispielen 1.1 und 1.2: Das stochastische Modell ist gegeben durch

- Zufallsvariable X mit Werten in Maßraum $(\mathcal{X}, \mathcal{F})$ von unbekannter Verteilung $\mathbb{P}_X$, die aber in einer bekannten Klasse $\mathcal{P}$ von Verteilungen auf $(\mathcal{X}, \mathcal{F})$ liegt.
- Realisierung $X(w) =: x \in \mathcal{X}$ modelliert das Ergebnis einer Durchführung des Zufallsexperiments in der Realität.

Ziel: Aussage über die unbekannte Verteilung $\mathbb{P}_X \in \mathcal{P}$ auf der Basis von x (Datum). Statistische Fragestellung:

(i). In Beispiel 1.1: $\mathcal{X} = \mathbb{N}_0, \mathcal{F} = \mathcal{P}(\mathbb{N}_0)$, $\mathcal{P} = \{\text{Hyp}(N, M, n) : N \geq M\} = \mathcal{P}_{M,n}$

(ii). In Beispiel 1.2: $\mathcal{X} = \{0,1\}^n$, $\mathcal{F} = \mathcal{P}(\{0,1\}^n)$, $\mathcal{P} = \{(\beta_p)^n : p \in (0,1)\} = \mathcal{P}_n$

1.4 Definition

Ein stochastisches Modell $(X, \mathcal{X}, \mathcal{F}, \mathcal{P})$ besteht aus

(i). Messraum $(\mathcal{X}, \mathcal{F})$ (Stichprobenraum)

(ii). einer Zufallsvariablen $X : (\Omega, \mathcal{A}, \mathbb{P}) \to (\mathcal{X}, \mathcal{F})$ (Datum)

(iii). Familie $\mathcal{P}$ mit
$$\mathcal{P} \subseteq \mathcal{P}(\mathcal{F}) := \{\mathbb{P} : \mathcal{F} \to [0, \infty]; \mathbb{P}\,\text{Maß auf}\,\mathcal{F}\}$$
mit $\mathbb{P}_X = \mathbb{P} \circ X^{-1} \in \mathcal{P}$ (Verteilungsannahme).

Interpretation:

- X beschreibt Zufallsereignis mit $\mathbb{P}_X \in \mathcal{P}$.
- Es existiert genau ein $\mu_0 \in \mathcal{P}$ mit $\mathbb{P}_X = \mu_0$, wobei μ_0 unbekannt, aber $\mathcal{P}$ bekannt ist.
- Die Realisierungen $X(w) = x \in \mathcal{X}$ modellieren die in der Realität beobachteten Ergebnisse.

1.5 Beispiel: Lebensdauer eines elektronischen Bauteils

- Unabhängige Beobachtungen von n Exemplaren ergibt Datenvektor / Datum $\underline{x}_n = (x_i)_{i=1,\ldots,n} \in \mathbb{R}^n$, wobei x_i die beobachtete Lebensdauer des i-ten Bauteils ist.
- Modell: $\underline{X}_n = (X_i)_{i=1,\ldots,n}$ mit $(X_i)_{i=1,\ldots,n}$ iid und $X_i \sim \text{Exp}(\lambda)$, $\lambda > 0$. Sinnvoll nur bei nicht alternden Bauteilen wegen der sogenannten Gedächtnislosigkeit („Nichtalterungseigenschaft“) von $\text{Exp}(\lambda)$: Gilt $X \sim \text{Exp}(\lambda)$, dann folgt für $s > 0$:
$$\begin{aligned}\mathbb{P}[X \leq s + t | X > t] &= \frac{\mathbb{P}[t \leq X \leq t + s]}{\mathbb{P}[X > t]} = \frac{1 - e^{-\lambda \cdot (s+t)} - (1 - e^{-\lambda \cdot t})}{1 - (1 - e^{-\lambda \cdot t})} \\ &= 1 - e^{-\lambda \cdot s} = \mathbb{P}[X \leq s]\end{aligned}$$
Wähle $\mathcal{X} = \mathbb{R}^n$, $\mathcal{F} = \mathcal{B}(\mathbb{R}^n)$, $\mathcal{P} = \{(\text{Exp}(\lambda))^n; \lambda > 0\}$ ($n \in \mathbb{N}$ bekannt)
- Gemeinsamkeit zu Beispiel 1.2: Ein Zufallsexperiment wird n-mal unabhängig wiederholt.

1.6 Bemerkung

Die einmalige Durchführung eines Zufallsexperiments wird modelliert durch $(X, \mathcal{X}, \mathcal{F}, \mathcal{P})$. Dann ist die n-malige unabhängige Wiederholung des Zufallsexperiments (Gesamtexperiment) beschreibbar durch $(\underline{X}_n, \mathcal{X}^n, \mathcal{F}^n, \mathcal{P}^n)$ (n-faches Produktmodell zu $(X, \mathcal{X}, \mathcal{F}, \mathcal{P})$), wobei $\underline{X}_n := (X_i)_{i \leq n}$ und X_i beschreibt Ergebnis in der i-ten Wiederholung, $\mathcal{F}^n = \bigotimes_{i=1}^n \mathcal{F}, \mathcal{P}^n := \{\mu^n, \mu \in \mathcal{P}\}$.

Beachte:
$$\mathbb{P}_{\underline{X}_n} = \mu^n, \mu \in \mathcal{P} \Leftrightarrow (X_i)_{i \leq n}\,\text{i.i.d mit}\,X_i \sim \mu \in \mathcal{P} \tag{1.1}$$
Daher erfolgt die Beschreibung des n-fachen Produktmodells häufig durch die rechte Seite von (1.1) unter der genauen Angabe von $\mathcal{P}$.

1.7 Beispiel: Bernoulli-Experiment

Das Modell

$$\left(\underline{X}_n = (X_i)_{i\leq n}, \{0,1\}^n, \mathcal{P}(\{0,1\}^n), \{(\beta_p)^n \colon p \in (0,1)\}\right)$$

ist n-faches Produktmodell zum Bernoulli-Experiment $(X, \{0,1\}, \mathcal{P}(\{0,1\}), \{\beta_p; p \in (0,1)\})$ und heißt n-faches Bernoulli-Experiment. Alternativ gemäß (1.1): $(X_i)_{i\leq n}$ iid mit $X_i \sim \beta_p$, $p \in (0,1)$.

Bisherige Verteilungsannahmen:

(i). $\mathcal{P} = \{\mathrm{Hyp}(N,M,n); N \geq M\}$ mit M, n bekannt

(ii). $\mathcal{P} = \{(\beta_p)^n; p \in (0,1)\}$ mit $n \in \mathbb{N}$ bekannt.

(iii). $\mathcal{P} = \{(\mathrm{Exp}(\lambda))^n; \lambda > 0\}$ mit $n \in \mathbb{N}$ bekannt.

Gemeinsame Struktur:

$$\mathcal{P} = \{\mu_\theta; \theta \in \Theta\} \qquad \Theta \subseteq \mathbb{R}^d, d \geq 1 \tag{1.2}$$

(i). $\Theta = \{k \in \mathbb{N}; k \geq M\} = \{M, M+1, \ldots\}$, $\theta = N$, $\mu_\theta = \mathrm{Hyp}(N,M,n)$

(ii). $\Theta = (0,1)$, $\theta = p$, $\mu_\theta = (\beta_p)^n$

(iii). $\Theta = (0,\infty)$, $\theta = \lambda$, $\mu_\theta = (\mathrm{Exp}(\lambda))^n$

Falls in (1.2) $\psi : \Theta \to \mathcal{P}, \theta \mapsto \mu_\theta$ (Parametrisierung) bijektiv ist, so heißt ψ identifizierend. Damit ist insbesondere die wahre Verteilung von X identifizierbar mit wahrem Parameter $\theta_0 \in \Theta$ mit $\mu_{\theta_0} = \mathbb{P} \circ X^{-1}$.
In diesem Fall heißt $\mathcal{P}$ parametrische Verteilung und Θ Parameterraum.

$$(\Omega, \mathcal{A}, \mathbb{P} = \mathbb{P}_\theta) \xrightarrow{X} (\mathcal{X}, \mathcal{F}, \mu_\theta = \mathbb{P}_\theta \circ X^{-1}) \qquad (\theta \in \Theta)$$

2 Schätzer und grundlegende Eigenschaften

Ziel: Das unbekannte θ_0 oder allgemein eine Funktion $\gamma(\theta_0)$ ist aus dem Datum $x \in \mathcal{X}$ zu „schätzen".

2.1 Definition

Sei $(X, \mathcal{X}, \mathcal{F}, \mathcal{P})$ ein statistisches Modell, $\gamma : \Theta \to \mathbb{R}^d$. Dann heißt $\hat{\Theta} \in \mathcal{M}(\mathcal{F})$ ein Schätzer für $\gamma(\theta)$ und $\hat{\Theta}(x)$ heißt Schätzwert zur Beobachtung $x \in \mathcal{X}$. Damit ist $\hat{\Theta} \circ X \in \mathcal{M}(\mathcal{A})$ und somit beispielsweise die Untersuchung

$$\mathbb{P}_\theta\big[\|\hat{\Theta} \circ X - \gamma(\theta)\| \leq \varepsilon\big]$$

für $\varepsilon > 0$ möglich.

2.2 Beispiel

$X_1, \ldots, X_n$ i.i.d, $X_i \sim N(\mu, \sigma^2)$; $\mu \in \mathbb{R}$ und $\sigma^2 > 0$ unbekannt. Dann

$$\Theta = \mathbb{R} \times (0, \infty) \subseteq \mathbb{R}^2 \qquad\qquad \theta = (\mu, \sigma^2)$$

Mögliche Parameteranteile:

(i). $\gamma(\theta) = \gamma(\mu, \sigma^2) = \mu$

(ii). $\gamma(\theta) = \gamma(\mu, \sigma^2) = \sigma^2$

(iii). $\gamma(\theta) = \gamma(\mu, \theta) = \mu^2 + \sigma^2 = \mathbb{E}_\theta(X_1^2)$

(iv). $\gamma(\theta) = (\mu, \sigma^2)$

Mögliche Schätzer z.B. für (i):

$$\hat{\Theta}(\underline{x}_n) := \frac{1}{n} \cdot \sum_{i=1}^n x_i$$
$$\tilde{\Theta}(\underline{x}_n) := \prod_{i=1}^n x_i$$
$$\Theta^*(\underline{x}_n) := x_1$$

für $\underline{x}_n \in \mathbb{R}^n$. Offenbar notwendig:

(i). mathematische Kriterien für die Qualität von Schätzern

(ii). Konstruktionsprinzipien für Schätzer

Zunächst (i) für $\gamma : \Theta \to \mathbb{R}$. Im Folgenden liegt stets ein parametrisches statistisches Modell $(X, \mathcal{X}, \mathcal{F}, \mathcal{P})$ mit $\mathcal{P}$ aus (1.2) zugrunde, d.h.

$$\mathcal{P} = \{P_\theta; \theta \in \Theta\}$$

2.3 Definition

$\Theta \subseteq \mathbb{R}^d$, $\gamma : \Theta \to \mathbb{R}$. Der Schätzer $\hat{\Theta} : \mathcal{X} \to \mathbb{R}$ heißt erwartungstreu (für $\gamma(\theta)$) $:\Leftrightarrow$

$$\forall \theta \in \Theta : \mathbb{E}_\theta(\hat{\Theta}(X)) = \gamma(\theta) \tag{2.1}$$

wobei $\mathbb{E}_\theta$ = Erwartungswertbildung bzgl. $\mathbb{P}_\theta$, d.h.

$$\begin{aligned}\mathbb{E}_\theta(\hat{\Theta}(X)) &:= \int_\Omega \hat{\Theta}(X(\omega))\, d\mathbb{P}_\theta(dw) \\ &= \int_{\mathcal{X}} \Theta(\hat{x})\, \underbrace{\mathbb{P}_\theta \circ X^{-1}(dx)}_{P_\theta} = \int \hat{\Theta}\, dP_\theta\end{aligned}$$

unter Verwendung des Transformationssatzes. Nach dem starken Gesetz der großen Zahlen liefert „$\hat{\Theta}$ im Mittel den Parameteranteil $\gamma(\Theta)$“.
Interpretation: $X_1, \dots, X_m, \dots$ i.i.d mit $X_i \sim P_\theta = \mathbb{P} \circ X_1^{-1}$. Starkes Gesetz der großen Zahlen:

$$\forall \theta \in \Theta : \hat{\gamma}_m := \frac{1}{m} \cdot \sum_{i=1}^m \hat{\Theta}(X_i) \to \underbrace{\mathbb{E}_\theta(\hat{\Theta}(X_1))}_{\overset{(2.1)}{=}\gamma(\theta) = \gamma(\theta_0)} \qquad (m \to \infty) \qquad \mathbb{P}_\theta \text{fast sicher}$$

für den wahren aber unbekannten Parameter $\theta = \theta_0$.

Beispiel 2.2 (Fortsetzung) $\gamma(\theta) = \gamma(\mu, \sigma^2) = \mu$. Dann:

$$\forall \theta \in \Theta : \mathbb{E}_{(\mu,\sigma^2)}(\hat{\Theta}(\underline{X}_n)) = \mathbb{E}_{(\mu,\sigma^2)}\left(\frac{1}{n}\sum_{i=1}^n X_i\right) = \mu$$

für alle $\theta \in \Theta$, d.h. $\hat{\Theta}$ ist erwartungstreu.

$$\begin{aligned}\forall \theta \in \Theta : \mathbb{E}_{(\mu,\sigma^2)}\left(\tilde{\Theta}(\underline{X}_n)\right) &= \mathbb{E}_{(\mu,\sigma^2)}\left(\prod_{i=1}^n X_i\right) \overset{\text{unabh}}{=} \prod_{i=1}^n \mathbb{E}_{(\mu,\sigma^2)} X_i \\ &= \mu^n\end{aligned}$$

also $\tilde{\Theta}$ ist für alle $n \geq 2$ nicht erwartungstreu.

$$\forall \theta \in \Theta : \mathbb{E}_{(\mu,\sigma^2)}(\Theta^*(\underline{X}_n)) = \mathbb{E}_{(\mu,\sigma^2)}(X_1) = \mu$$

d.h. Θ^* ist erwartungstreu. Beispiel 2.2 zeigt, dass neben der Erwartungstreue ein weiteres Qualitätskriterium notwendig ist. Sinnvoll: Verwende $\hat{\Theta}$ mit möglichst kleinem mittlerem quadratischen Schätzfehler (MQS):

$$\mathbb{E}_\theta\big((\hat{\Theta}(X) - \gamma(\theta))^2\big) \qquad (\theta \in \Theta)$$

Falls $\hat{\Theta}$ erwartungstreu, so gilt:

$$\mathbb{E}_\theta\big((\hat{\Theta}(X) - \gamma(\theta))^2\big) = \mathbb{V}_\theta(\hat{\Theta}(X))$$

2.4 Definition

$\gamma : \Theta \to \mathbb{R}$, $\hat{\Theta}_1$ und $\hat{\Theta}_2$ erwartungstreue Schätzer für $\gamma(\theta)$. Dann:

(i). $\hat{\Theta}_1$ gleichmäßig besser als $\hat{\Theta}_2$ $:\Leftrightarrow \forall \theta \in \Theta : \mathbb{V}_\theta(\hat{\Theta}_1) \leq \mathbb{V}_\theta(\hat{\Theta}_2)$

(ii). $\hat{\Theta}_1$ heißt gleichmäßig bester erwartungstreuer Schätzer für $\gamma(\theta)$ $:\Leftrightarrow \forall \theta \in \Theta$ und $\forall \hat{\Theta_2}$ erwartungstreue Schätzer für $\gamma(\theta)$: $\mathbb{V}_\theta(\hat{\Theta}_1(X)) \leq \mathbb{V}_\theta(\hat{\Theta}_2(X))$

Beispiel 2.2 (Fortsetzung)

$$\begin{aligned}
\mathbb{V}_\theta(\hat{\Theta}(\underline{X}_n)) &= \mathbb{V}_\theta\left(\frac{1}{n}\cdot\sum_{i=1}^n X_i\right) \overset{\text{unabh}}{=} \frac{1}{n^2}\cdot\sum_{i=1}^n \mathbb{V}_\theta(X_i) \\
&= \frac{1}{n^2}\cdot(n\cdot\sigma^2) = \frac{\sigma^2}{n} \\
\mathbb{V}_\theta(\Theta^*(\underline{X}_n)) &= \mathbb{V}_\theta(X_1) = \sigma^2 \geq \frac{1}{n}\cdot\sigma^2 = \mathbb{V}_\theta(\hat{\Theta}(\underline{X}_n))
\end{aligned}$$

mit $\theta \in \Theta$ beliebig, d.h. $\hat{\Theta}$ ist gleichmäßig besser als Θ^*. Unter bestimmten Voraussetzungen an $\mathcal{P}$ können gleichmäßig beste erwartungstreue Schätzer angegeben werden (s. Kapitel 5).

3 Das Maximum-Likelihood-Prinzip

Betrachte zunächst nur diskrete Modelle, d.h. der Stichprobenraum $\mathcal{X}$ ist abzählbar. Idee: Wähle als Schätzwert für θ zur Beobachtung $x \in \mathcal{X}$ denjenigen Parameter, der dem beobachteten Ereignis $[X = x]$ die größte Wahrscheinlichkeit zuordnet (Plausibilitätsprinzip). Deshalb:

3.1 Definition

Sei $\mathcal{X}$ abzählbar. Setze

$$L(\theta, x) := \mathbb{P}_\theta[X = x] = P_\theta(\{x\}) \qquad (\theta \in \Theta, x \in \mathcal{X})$$

Dann heißt $L(\cdot, x) : \Theta \to [0,1]$ Likelihood-Funktion zur Beobachtung $x \in \mathcal{X}$. Jede Maximalstelle von $L(\cdot, x)$ heißt Maximum-Likelihood-Schätzwert für θ zur Beobachtung x. Also θ^* ist Maximum-Likelihood-Schätzwert zu $X \Leftrightarrow L(\theta^*, x) \geq L(\theta, x)$ für alle $\theta \in \Theta$. Äquivalente Formulierungen:

$$\begin{aligned} L(\theta^*, x) &= \max_{\theta \in \Theta} L(\theta, x) \\ \theta^* &= \operatorname*{argmax}_{\theta \in \Theta} L(\theta, x) \end{aligned}$$

Beachte: $\theta^* = \theta^*(x)$.

3.2 Beispiel

$X \sim \mathrm{Hyp}(N, M, n)$, $\theta = N \in \Theta = \{M, M+1, \ldots\}$, $\mathcal{X} = \mathbb{N}_0$ mit M, n feste Parameter (bekannt). Dann:

$$L(N, m) = \mathbb{P}_N[X = m] \overset{1.3}{=} \frac{\binom{M}{m} \cdot \binom{N-M}{n-m}}{\binom{N}{n}}$$

Eine elementare Rechnung liefert:

$$\begin{aligned} L(N-1, m) \overset{<}{\underset{>}{=}} L(N, m) &\Leftrightarrow N \cdot m \overset{<}{\underset{>}{=}} n \cdot M \\ &\overset{m \geq 1}{\Leftrightarrow} N \overset{<}{\underset{>}{=}} \frac{n \cdot M}{m} \end{aligned}$$

Damit: $\hat{N} = \left[\frac{n \cdot M}{m}\right]$ ist Maximum-Likelihood-Schätzwert für $m \geq 1$. Stimmt mit dem „naiven" Schätzwert aus Beispiel 1.1 überein.
Falls $m = 0$, so ist wegen $(*)$ $L(\cdot, m) = L(\cdot, 0)$ streng monoton wachsend, d.h. es existiert kein ML-Schätzwert zu $m = 0$. Ferner: Gilt $\frac{n \cdot M}{m} \notin \mathbb{N}_0$, so ist $\hat{N}$ einzige Maximalstelle. Ist dagegen $\frac{n \cdot M}{m} \in \mathbb{N}_0$, so sind $\frac{n \cdot M}{m}$ und $\frac{n \cdot M}{m} - 1$ Maximalstellen.

Folgerung aus 3.2: ML-Schätzwerte müssen nicht existieren und sind gegebenenfalls mitunter nicht eindeutig.

3.3 Beispiel

Seien $X_1, \dots, X_n$ i.i.d, $X_i \sim \beta_p$ $(i = 1, \dots, n)$ für $p \in [0,1]$, $\underline{x}_n = (x_1, \dots, x_n) \in \{0,1\}^n$,

$$\begin{aligned} L(p, \underline{x}_n) &= \mathbb{P}_p[\underline{X}_n = \underline{x}_n] = \mathbb{P}_p\left(\bigcap_{i=1}^n [X_i = x_i]\right) \\ &\overset{\text{unabh}}{=} \prod_{i=1}^n \mathbb{P}_p[X_i = x_i] = \prod_{i=1}^n p^{x_i} \cdot (1-p)^{1-x_i} \\ &= p^{\sum_{i=1}^n x_i} \cdot (1-p)^{n - \sum_{i=1}^n x_i} \end{aligned} \tag{*}$$

Maximierung von $L(\cdot, \underline{x}_n)$:

(i). $\underline{x}_n = 0$. Dann:

$$L(p, \underline{0}) = 1 \cdot (1-p)^n \Rightarrow \hat{p} = 0$$

$\hat{p}$ ist eindeutige Maximalstelle.

(ii). $\underline{x}_n = 1$: Dann:

$$L(p, \underline{1}) = p^n \Rightarrow \hat{p} = 1$$

$\hat{p}$ ist eindeutige Maximalstelle.

(iii). $\underline{x}_n \notin \{\underline{0}, \underline{1}\}$, d.h. $\exists i, j : x_i = 0, x_j = 1$. Aus $(*)$ folgt:

$$L(0, \underline{x}_n) = 0 = L(1, \underline{x}_n)$$

und $L(\cdot, \underline{x}_n) > 0$ auf $(0,1)$, also muss Maximalstelle in $(0,1)$ liegen. Da $x \mapsto \log(x)$ streng monoton wachsend auf $(0, \infty)$, folgt

$$\hat{p} := \underset{p \in (0,1)}{\operatorname{argmax}}\, L(p, \underline{x}_n) \Leftrightarrow \hat{p} = \underset{p \in (0,1)}{\operatorname{argmax}} \log L(p, \underline{x}_n)$$

Betrachte deshalb:

$$\begin{aligned} \log L(p, \underline{x}_n) &= \sum_{i=1}^n x_i \cdot \log p + \left(n - \sum_{i=1}^n x_i\right) \cdot \log(1-p) \\ \Rightarrow \frac{d}{dp} L(p, \underline{x}_n) &= \frac{1}{p} \cdot \sum_{i=1}^n x_i - \frac{1}{1-p} \cdot \left(n - \sum_{i=1}^n x_i\right) \overset{!}{=} 0 \\ \Rightarrow \hat{p} &= \frac{1}{n} \cdot \sum_{i=1}^n x_i \end{aligned}$$

$\hat{p}$ ist eindeutige Maximalstelle von $L(\cdot, \underline{x}_n)$.

Fazit:

$$\hat{p} = \overline{x}_n := \frac{1}{n} \cdot \sum_{i=1}^n x_i$$

ist ML-Schätzwert für p für jede Beobachtung $\underline{x}_n$.

3.4 Bemerkung

(i). Übergang zu $\ell(\theta, x) := \log L(\theta, x)$ vereinfacht häufig das Maximierungsproblem. $\ell(\cdot, x)$ heißt log-Likelihood-Funktion (zur Beobachtung x).

(ii). $\Theta \subset \mathbb{R}^d$ und $\ell(\cdot, x)$ sei auf $\mathring{\Theta} := \operatorname{int}(\Theta)$ partiell differenzierbar, so sind die ML-Schätzwerte $\theta = (\theta_1, \dots, \theta_d)$ in $\mathring{\Theta}$ Lösungen von

$$\forall j = 1, \dots, d : \frac{\partial}{\partial \theta_j} \ell(\theta, x) = 0 \tag{3.1}$$

(sogenannte log-Likelihood-Gleichungen). Beachte:

- (3.1) nicht immer explizit lösbar; dann numerische Verfahren verwenden.
- Lösungen von (3.1) sind mitunter keine ML-Schätzwerte.

Ausdehnung des ML-Prinzips auf absolut-stetige Modelle $(X, \mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \mathcal{P})$ mit

$$\mathcal{P} = \{P_\theta = f_\theta\, \lambda^n; \theta \in \Theta\} \qquad (\Theta \subseteq \mathbb{R}^d)$$

d.h.

$$\mathbb{P}_\theta[X \in B] = P_\theta(B) := \int f_\theta \cdot 1_B \, d\lambda^n$$

für $B \in \mathcal{B}(\mathbb{R}^n)$, also f_θ Lebesgue-Dichte von X bzw. P_θ. Dann:

$$\begin{aligned}\mathbb{P}_\theta[X = x] = \mathbb{P}_\theta[X \in \{x\}] &= \int_{\mathbb{R}^n} f_\theta(y) \cdot 1_{\{x\}}(y)\, \lambda^n(dy) \\ &= \int_{\mathbb{R}^n} 1_{\{x\}}(y) \cdot f_\theta(x)\, \lambda^n(dy) = f_\theta(x) \cdot \lambda^n(\{x\}) = 0\end{aligned}$$

Eine Festlegung $L(\theta, x) := \mathbb{P}_\theta[X = x]$ lieferte $L(\theta, x) = 0$ für alle $\theta \in \Theta, x \in \mathbb{R}^n$, d.h. jedes θ wäre ML-Schätzwert zu jeder Beobachtung x. Ist offensichtlich unsinnig. Deshalb kleine Modifikation:

$$\begin{aligned}\mathbb{P}_\theta[X = x] &\overset{\varepsilon \text{ klein}}{\approx} \mathbb{P}_\theta[X \in (x - \varepsilon, x + \varepsilon)] \\ &= \int_{(x-\varepsilon, x+\varepsilon)} f_\theta(y)\, \lambda^n(dy) \approx \int_{(x-\varepsilon, x+\varepsilon)} f_\theta(x)\, \lambda^n(dy) \\ &= f_\theta(x) \cdot \lambda^n((x - \varepsilon, x + \varepsilon)) = f_\theta(x) \cdot (2\varepsilon)^n \\ \Rightarrow \mathbb{P}_\theta[X = x] &\overset{\varepsilon \text{ klein}}{\approx} \mathbb{P}_\theta[X \in (x - \varepsilon, x + \varepsilon)] \overset{\varepsilon \text{ klein}}{\approx} (2\varepsilon)^n \cdot f_\theta(x)\end{aligned}$$

Hoffnung:

$$\begin{aligned}\underset{\theta \in \Theta}{\operatorname{argmax}}\, \mathbb{P}_\theta[X \in (x - \varepsilon, x + \varepsilon)] &\approx \underset{\theta \in \Theta}{\operatorname{argmax}} (2\varepsilon)^n \cdot f_\theta(x) \\ &= \underset{\theta \in \Theta}{\operatorname{argmax}}\, f_\theta(x)\end{aligned}$$

3.5 Definition

Im absolut-stetigen Modell setze $L(\theta, x) := f_\theta(x)$ für $\theta \in \Theta, x \in \mathbb{R}^n$. $L(\cdot, x) : \Theta \to [0, \infty)$ heißt Likelihood-Funktion zu $x \in \mathbb{R}^n$. Jede Maximalstelle von $L(\cdot, x)$ heißt Maximum-Likelihood-Schätzwert für θ (zur Beobachtung $x \in \mathbb{R}^n$).

3.6 Beispiel

$X_1, \ldots, X_n$ i.i.d, $X_i \sim \text{Exp}(\lambda)$ $(i = 1, \ldots, n)$, $\lambda \in (0, \infty)$ unbekannt. Dann $X := \underline{X}_n = (X_1, \ldots, X_n) \sim (\text{Exp}(\lambda))^n$, d.h. besitzt λ^n-Dichte

$$\begin{aligned}f_\lambda(\underline{x}_n) &= \prod_{i=1}^n 1_{(0,\infty)}(x_i) \cdot \lambda \cdot e^{-\lambda \cdot x_i} \\ \Rightarrow \forall \underline{x}_n \in (0, \infty)^n : L(\lambda, \underline{x}_n) &= \prod_{i=1}^n \lambda \cdot e^{-\lambda \cdot x_i} = \lambda^n \cdot e^{-\lambda \cdot \sum_{i=1}^n x_i}\end{aligned}$$

Beachte: $\mathbb{P}_\lambda[\underline{X}_n \notin (0, \infty)^n] = 0$ für alle $\lambda > 0$. Maximierung von $L(\cdot, \underline{x}_n)$: $L(\cdot, \underline{x}_n) > 0$ und $L(\cdot, \underline{x}_n)$ stetig auf $(0, \infty)$,

$$\lim_{\lambda \to 0} L(\lambda, \underline{x}_n) = 0$$

$$\lim_{\lambda \to \infty} L(\lambda, \underline{x}_n) = 0$$

$\Rightarrow$ Es existiert eine (globale) Maximalstelle (Satz vom Maximum) in $(0,\infty)$.

$$\ell(\lambda,\underline{x}_n) = \log L(\lambda,\underline{x}_n) = n\cdot\log(\lambda) - \lambda\cdot\sum_{i=1}^n x_i$$
$$\frac{\partial}{\partial\lambda}\ell(\lambda,\underline{x}_n) = \frac{n}{\lambda} - \sum_{i=1}^n x_i \overset{!}{=} 0$$
$$\Rightarrow \hat\lambda = \frac{n}{\sum_{i=1}^n x_i} = \frac{1}{\overline{x}_n}$$

$\hat\lambda$ ist eindeutiger ML-Schätzwert für λ zu $\underline{x}_n \in (0,\infty)^n$.

3.7 Beispiel

$X_1,\dots,X_n$ i.i.d, $X_i \sim N(\mu,\sigma^2)$ $(i=1,\dots,n)$ mit $(\mu,\sigma^2)\in\mathbb{R}\times(0,\infty) =: \Theta$ (Vgl. 2.2).

$$\begin{aligned} L(\mu,\sigma^2,\underline{x}_n) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\cdot\sigma^2}}\cdot\exp\left(-\frac{(x_i-\mu)^2}{\sigma^2}\right) \\ &= (2\pi\cdot\sigma^2)^{-\frac{n}{2}}\cdot\exp\left(-\frac{1}{2\sigma^2}\cdot\sum_{i=1}^n (x_i-\mu)^2\right) \end{aligned}$$

Fallunterscheidung:

(i). $\underline{x}_n := (x_1,\dots,x_n)\in\mathbb{R}^n$ mit $x_1 = x_2 = \ldots = x_n$. Dann:

$$\begin{aligned} L(\mu,\sigma^2,\underline{x}_n) &\overset{\mu:=x_1}{=} L(x_1,\sigma^2,\underline{x}_n) = \frac{1}{(2\pi\cdot\sigma^2)^{\frac{n}{2}}}\cdot 1 \\ &\to\infty \qquad (\sigma^2\to 0) \end{aligned}$$

also hat $L(\cdot,\cdot,\underline{x}_n)$ in Θ keine Maximalstelle.

(ii). $\underline{x}_n\in\mathbb{R}^n$ hat min. zwei verschiedene Komponenten. $L(\mu,\sigma^2,\underline{x}_n)\to 0$, falls $\mu\to\pm\infty$ oder $\sigma^2\to 0$ bzw. $\sigma^2\to\infty$. Deswegen existiert eine globale Maximalstelle in Θ, für die notwendig gilt:

$$0 \overset{!}{=} \frac{\partial}{\partial\mu}\ell(\mu,\sigma^2,\underline{x}_n) = \frac{1}{\sigma^2}\sum_{i=1}^n (x_i-\mu)$$
$$0 \overset{!}{=} \frac{\partial}{\partial\sigma^2}\ell(\mu,\sigma^2,\underline{x}_n) = -\frac{n}{\sigma^2} + \frac{1}{2\cdot(\sigma^2)^2}\cdot\sum_{i=1}^n (x_i-\mu)^2$$

Eindeutige Lösung ist

$$\hat\theta := (\hat\mu,\hat\sigma^2) = (\overline{x}_n,\sigma_n^2)$$

wobei

$$\overline{x}_n := \frac{1}{n}\cdot\sum_{i=1}^n x_i \qquad \sigma_n^2 := \frac{1}{n}\cdot\sum_{i=1}^n (x_i-\overline{x}_n)^2$$

$\hat\theta$ ist eindeutiger ML-Schätzwert für (μ,σ^2) zu $\underline{x}_n$.

Bezeichnung: Für $\underline{x}_n = (x_1,\dots,x_n)\in\mathbb{R}^n$ heißt

$$\overline{x}_n := \frac{1}{n}\cdot\sum_{i=1}^n x_i$$

empirisches Mittel (Strichprobenmittel) und

$$s_n^2 := \frac{1}{n}\cdot\sum_{i=1}^n (x_i-\overline{x}_n)^2$$

empirische Varianz (Stichprobenvarianz). Bislang nur ML-Schätzwerte betrachtet, aber nicht ML-Schätzer im Sinne von Definition 2.1. Dazu:

3.8 Definition

Sei $(X, \mathcal{X}, \mathcal{F}, \mathcal{P})$ parametrisches statistisches Modell mit diskreten oder absolut-stetigen Verteilungen P_θ, $\theta \in \Theta \subseteq \mathbb{R}^d$, d.h.

$$L(\theta, x) = \begin{cases} P_\theta(\{x\}) & \mathcal{X} \text{ abzählbar} \\ f_\theta(x) = \dfrac{dP_\theta}{d\lambda^n}(x) & \mathcal{X} = \mathbb{R}^n \end{cases}$$

Eine $\mathcal{F} - \mathcal{B}(\mathbb{R}^d)$-messbare Abbildung $\hat{\Theta} : \mathcal{X} \to \mathbb{R}^d$ heißt Maximum-Likelihood-Schätzer (MLS) für θ, wenn $\hat{\Theta}(x)$ für jedes $x \in \mathcal{X} \backslash A$ Maximum-Likelihood-Schätzwert für θ ist, wobei $P_\theta(A) = 0$ für alle $\theta \in \Theta$.

Bisherige Beispiele:

(i). In Beispiel 3.2 ist $\mathcal{X} = \mathbb{N}_0$, $A = \{0\}$. Beachte:

$$\forall N \in \mathbb{N} : P_\theta(A) = \text{Hyp}(N, M, n)(\{0\}) > 0$$

d.h. es gibt keinen ML-Schätzer.

(ii). In Beispiel 3.3: $\mathcal{X} = \{0,1\}^n, A = \varnothing$, $\hat{\theta}(\underline{x}_n) = \overline{x}_n$ ist MLS.

(iii). In Beispiel 3.6: $\mathcal{X} = \mathbb{R}^n$, $A = \mathbb{R}^n \backslash (0, \infty)^n$,

$$\hat{\lambda}(\underline{x}_n) := 1_{(0,\infty)^n}(\underline{x}_n) \cdot \frac{1}{\overline{x}_n}$$

Da für alle $\lambda > 0$

$$\begin{aligned} P_\lambda(A) &= \mathbb{P}_\lambda[\underline{X}_n \in A] = 1 - \mathbb{P}_\lambda\left(\bigcap_{i=1}^n [X_i > 0]\right) \\ &= 1 - \prod_{i=1}^n \mathbb{P}_\lambda[X_i > 0] = 0 \end{aligned}$$

ist $\hat{\lambda}$ MLS für λ.

(iv). In Beispiel 3.7: $\mathcal{X} = \mathbb{R}^n, A = \{\underline{x}_n \in \mathbb{R}^n; x_1 = x_2 = \ldots = x_n\}$, $\hat{\Theta}(\underline{x}_n) = (\overline{x}_n, \sigma_n^2)$,

$$\begin{aligned} P_{(\mu,\sigma^2)}(A) &= \mathbb{P}_{(\mu,\sigma^2)}[\underline{X}_n \in A] = \mathbb{P}_{(\mu,\sigma^2)}[X_1 = \ldots = X_n] \\ &\leq \mathbb{P}_{(\mu,\sigma^2)}[X_1 = X_2] \overset{\text{Fubini}}{=} 0 \end{aligned}$$

4 Dominierte Verteilungsannahmen und Exponentialfamilien

Ziel: Simultane Behandlung von diskreten und absolut-stetigen Modellen.

4.1 Definition

Sei $(X, \mathcal{X}, \mathcal{F}, \mathcal{P})$ ein parametrisches statistisches Modell, $\mu : \mathcal{F} \to [0, \infty]$ sei ein σ-finites Maß. $\mathcal{P}$ ist durch μ dominiert ($\mathcal{P} << \mu$) $:\Leftrightarrow \forall \theta \in \Theta \, \exists \mu$–Dichte$f_\theta : \mathcal{X} \to [0, \infty)$ von P_θ. Schreibweise:

$$f_\theta = \frac{dP_\theta}{d\mu}$$

4.2 Beispiel

$\mathcal{P}$ in absolut-stetigen Modellen ist durch $\mu = \lambda^n$ dominiert.

4.3 Beispiel

Sei X Zufallsvariable in $\mathcal{X} = \{x_i; i \in \mathbb{N}\}$ mit $X \sim P_\theta$, $\theta \in \Theta$. Sei

$$\mu := \sum_{i \in \mathbb{N}} \delta_{x_i}$$

das Zählmaß auf $\mathcal{X}$. Dann $\mathcal{P} = \{P_\theta; \theta \in \Theta\} << \mu$ wobei $f_\theta(x) = P_\theta(\{x\})$ für $x \in \mathcal{X}$ (sogenannte Zähldichte). Dazu:

$$\mu(F) = \sum_{i \in \mathbb{N}} \delta_{x_i}(F) = \sharp\{x_i; x_i \in F\}$$

Dann

$$\begin{aligned}
\int_F f_\theta(x) \, \mu(dx) &= \int 1_F(x) \cdot P_\theta(\{x\}) \left(\sum_{i \in \mathbb{N}} \delta_{x_i}(dx) \right) \\
&\overset{\text{Ü2}}{=} \sum_{i \in \mathbb{N}} \int 1_F(x) \cdot P_\theta(\{x\}) \, \delta_{x_i}(dx) = \sum_{i \in \mathbb{N}} 1_F(x_i) \cdot P_\theta(\{x_i\}) \\
&= \sum_{i \in \mathbb{N}, x_i \in F} P_\theta(\{x_i\}) \overset{\sigma\text{–additiv}}{=} P_\theta \left(\dot{\bigcup_{i \in \mathbb{N}, x_i \in F}} \{x_i\} \right) \\
&= P_\theta(F)
\end{aligned}$$

also $P_\theta(\{x\}) = \frac{dP_\theta}{d\mu}$.

4.4 Definition

Sei $(X, \mathcal{X}, \mathcal{F}, \mathcal{P})$ mit $\mathcal{P} << \mu$. Für $\theta \in \Theta \subseteq \mathbb{R}^d$ sei $f_\theta : \mathcal{X} \to [0, \infty)$ die μ-Dichte von P_θ. Setze

$$L(\theta, x) := f_\theta(x) \qquad (\theta \in \Theta, x \in \mathcal{X})$$

Dann heißt $L(\cdot, x) : \Theta \to [0, \infty)$ Likelihood-Funktion zu $x \in \mathcal{X}$. Eine $\mathcal{F}-\mathcal{B}(\mathbb{R}^d)$-messbare Abbildung $\hat{\Theta} : \mathcal{X} \to \mathbb{R}^d$ heißt Maximum-Likelihood-Schätzer für θ, wenn

$$\forall x \in \mathcal{X} \backslash A : \hat{\Theta}(x) = \underset{\theta \in \Theta}{\operatorname{argmax}} \, L(\theta, x)$$

wobei $P_\theta(A) = 0$ für alle $\theta \in \Theta$. $\hat{\Theta}(x)$ heißt Maximum-Likelihood-Schätzwert für θ zu $x \in \mathcal{X}$.

Wegen 4.2 und 4.3 ist Definition 4.4 eine Verallgemeinerung von Definition 3.1, 3.5 und 3.8.

Erinnere an die Bedeutung von Produktmodellen.

4.5 Satz

Sei $(\underline{X}_n, \mathcal{X}^n, \mathcal{F}^n, \mathcal{P}^n)$ das n-fache Produktmodell zu $(X, \mathcal{X}, \mathcal{F}, \mathcal{P})$ aus 4.4. Dann gilt: $\mathcal{P}^n << \mu^n := \bigotimes_{i=1}^n \mu$, denn

$$f_{\theta,n}(\underline{x}_n) := \prod_{i=1}^n f_\theta(x_i) \qquad (\underline{x}_n \in \mathcal{X}^n)$$

ist μ^n-Dichte von P_θ^n.

Beweis: Übung/MAST

Viele parametrische Verteilungsannahmen $\mathcal{P}$ haben eine gemeinsame Struktur. Dazu:

4.6 Definition

Sei $(X, \mathcal{X}, \mathcal{F}, \mathcal{P})$ ein statistisches Modell, μ ein σ-finites Maß auf $\mathcal{F}$ und c, $q_1, \ldots, q_k$: $\Theta \to \mathbb{R}$ beliebige Funktionen sowie $h, T_1, \ldots, T_k : \mathcal{X} \to \mathbb{R}$ messbar und $h \geq 0$. Dann heißt $\mathcal{P}$ k-parametrische Exponentialfamilie in $q := (q_1, \ldots, q_k)$ und $T := (T_1, \ldots, T_k)$ bzgl. $\mu :\Leftrightarrow \forall \theta \in \Theta$ ist

$$f_\theta(x) = c(\theta) \cdot h(x) \cdot \exp\left(\sum_{j=1}^k q_j(\theta) \cdot T_j(x)\right) \qquad (x \in \mathcal{X})$$

eine μ-Dichte von P_θ.

4.7 Bemerkung

(i). Exponentialfamilien sind dominiert.

(ii). $\forall \theta \in \Theta$:

$$1 = \int f_\theta \, d\mu = c(\theta) \cdot \int h(x) \cdot \exp\left(\sum_{j=1}^k q_j(\theta) \cdot T_j(x)\right) d\mu$$

$$\Rightarrow 0 < c(\theta) = \frac{1}{\int h(x) \cdot \exp\left(\sum_{j=1}^k q_j(\theta) \cdot T_j(x)\right) d\mu}$$

d.h. $c(\theta)$ ist bereits durch h, q, T eindeutig bestimmt (nur Normierungsfaktor).

(iii). $\forall \theta \in \Theta, \forall F \in \mathcal{F}$:

$$P_\theta(F) = \int_F f_\theta \, d\mu = c(\theta) \cdot \int_F \exp\left(\sum_{j=1}^k q_j(\theta) \cdot T_j(x)\right) \cdot \underbrace{h(x) \, \mu(dx)}_{=:\nu(dx)}$$

wobei $\nu := h \cdot \mu$. Also kann h in das dominierende Maß „gesteckt“ werden, d.h. q und T sind die relevanten Größen.

(iv). $\{x \in \mathcal{X}; f_\theta(x) > 0\} = \{x \in \mathcal{X}; h(x) > 0\}$, d.h. Träger von f_θ hängt nicht von θ ab.

4.8 Satz

Sei $(\underline{X}_n, \mathcal{X}^n, \mathcal{F}^n, \mathcal{P}^n)$ das Produktmodell zu $(X, \mathcal{X}, \mathcal{F}, \mathcal{P})$ mit k-parametrischer Exponentialfamilie $\mathcal{P}$ in q, T bzgl. μ. $\mathcal{P}^n$ ist k-parametrische Exponentialfamilie in $q, T_n = (T_{n1}, \dots, T_{nk})$ bzgl. μ^n, wobei

$$T_{nj}(\underline{x}_n) = \sum_{i=1}^n T_j(x_i) \qquad (j = 1, \dots, k)$$

Beweis:

- Gemäß 4.5 hat P_θ^n die μ^n-Dichte

$$\begin{aligned} f_{\theta,n}(\underline{x}_n) &\overset{4.5}{=} \prod_{i=1}^n f_\theta(x_i) \\ &= \prod_{i=1}^n c(\theta) \cdot h(x_i) \cdot \exp\left(\sum_{j=1}^k q_j(\theta) \cdot T_j(x_i)\right) \\ &= c(\theta)^n \cdot \left(\prod_{i=1}^n h(x_i)\right) \cdot \exp\left(\sum_{j=1}^k q_j(\theta) \cdot \left(\sum_{i=1}^n T_j(x_i)\right)\right) \end{aligned}$$

4.9 Beispiel: Einfaches Bernoulli-Experiment

$\mathcal{P} = \{\beta_p; p \in (0,1)\}$, $\mathcal{X} = \{0,1\}$, also hat β_p die Zähldichte

$$\begin{aligned} f_p(x) &= \beta_p(\{x\}) = p^x \cdot (1-p)^{1-x} \\ &= \exp(\log(p^x \cdot (1-p)^{1-x})) = \exp\left(x \cdot \log p + (1-x) \cdot \log(1-p)\right) \\ &= \exp\left(x \cdot \log\left(\frac{p}{1-p}\right) + \log(1-p)\right) \\ &= \underbrace{(1-p)}_{c(p)} \cdot \exp\left(\underbrace{\log\left(\frac{p}{1-p}\right)}_{q(p)} \cdot \underbrace{x}_{T(x)}\right) \end{aligned}$$

Somit ist $\mathcal{P}$ eine 1-parametrige Exponentialfamilie. Mit 4.8 folgt: $\mathcal{P}^n$ im n-fachen Bernoulli-Experiment ist 1-parametrige Exponentialfamilie in $q(p) = \log\left(\frac{p}{1-p}\right)$ und

$$T_n(\underline{x}_n) = \sum_{i=1}^n x_i$$

4.10 Beispiel: Normalverteilungs-Modell

$X \sim N(\mu, \sigma^2)$ mit $\mu \in \mathbb{R}, \sigma^2 > 0$. Dann hat X die Lebesgue-Dichte

$$\begin{aligned} f_{(\mu,\sigma^2)}(x) &= \frac{1}{\sqrt{2\pi \cdot \sigma^2}} \cdot \exp\left(-\frac{1}{2\sigma^2} \cdot (x-\mu)^2\right) \\ &= \underbrace{\frac{1}{\sqrt{2\pi \cdot \sigma^2}} \cdot \exp\left(-\frac{\mu^2}{2\sigma^2}\right)}_{c(\mu,\sigma^2)} \cdot \exp\left(\underbrace{-\frac{1}{2\sigma^2}}_{=:q_1(\mu,\sigma^2)} \cdot \underbrace{x^2}_{=:T_1(x)} + \underbrace{\frac{\mu}{\sigma^2}}_{=:q_2(\mu,\sigma^2)} \cdot \underbrace{x}_{=:T_2(x)}\right) \end{aligned}$$

$\mathcal{P} = \{N(\mu, \sigma^2); (\mu, \sigma^2) \in \mathbb{R} \times (0, \infty)\}$ ist 2-parametrige Exponentialfamilie $q = (q_1, q_2), T = (T_1, T_2)$. Nach 4.8: Im zugehörigen n-fachen Produktmodell (d.h. $X_1, \dots, X_n$ i.i.d mit $X_i \sim N(\mu, \sigma^2)$) ist $\mathcal{P}^n$ eine 2-parametrige Exponentialfamilie in

$$\begin{aligned} q(\mu, \sigma^2) &= \left(-\tfrac{1}{2\sigma^2} \quad \tfrac{\mu}{\sigma^2}\right) \\ T_n(\underline{x}_n) &= \left(\textstyle\sum_{i=1}^n x_i^2 \quad \sum_{i=1}^n x_i\right) \end{aligned}$$

4.11 Beispiel: Exponentialmodell

$X \sim \mathrm{Exp}(\lambda)$ mit $\lambda > 0$. X hat Lebesgue-Dichte

$$f_\lambda(x) = \underbrace{\lambda}_{c(\lambda)} \cdot \underbrace{1_{(0,\infty)}(x)}_{h(x)} \cdot \exp\left(\underbrace{-\lambda}_{q(\lambda)} \cdot \underbrace{x}_{T(x)} \right)$$

Also: $\mathcal{P} = \{f_\lambda \, \lambda(dx); \lambda > 0\}$ ist 1-parametrige Exponentialfamilie in $q(\lambda) = \lambda$ und $T(x) = x$. Entsprechend wieder Übertragung auf $\mathcal{P}^n$ möglich.

4.12 Beispiel

Sei $X = d + \varepsilon$ mit $d \in \mathbb{R}, \varepsilon \sim \mathrm{Exp}(\lambda)$, $\lambda > 0$. X hat Dichte

$$f_{\lambda,d}(x) = f_\lambda(x-d) = \lambda \cdot 1_{(d,\infty)}(x) \cdot \exp\left(-\lambda \cdot (x-d)\right)$$

$\mathcal{P}$ ist für kein k eine k-parametrische Exponentialfamilie, denn der Träger

$$\{f_{\lambda,d} > 0\} = (d, \infty)$$

hängt von d ab, vgl. 4.7(iv).

Im Folgenden Beschränkung auf 1-parametrige Exponentialfamilien, $\Theta \subseteq \mathbb{R}$.

4.13 Satz

Sei $(X, \mathcal{X}, \mathcal{F}, \mathcal{P})$ statistisches Modell mit $\mathcal{P} = \{P_\theta; \theta \in \Theta \subseteq \mathbb{R}\}$ eine 1-parametrige Exponentialfamilie in q, T bzgl. μ, d.h.

$$L(\theta, x) = c(\theta) \cdot h(x) \cdot \exp(q(\theta) \cdot T(x)) \qquad (x \in \mathcal{X}, \theta \in \Theta)$$

Ist $q : \Theta \to \mathbb{R}$ differenzierbar, so gilt für jedes messbare $g : \mathcal{X} \to \mathbb{R}$ mit $\mathbb{E}_\theta |g(X)| < \infty$ für alle $\theta \in \Theta$: Die Funktion $f : \Theta \to \mathbb{R}$

$$\begin{aligned} f(\theta) := \mathbb{E}_\theta(g(X)) &\overset{\text{TFS}}{=} \int_{\mathcal{X}} g(x) P_\theta(dx) \\ &= \int_{\mathcal{X}} g(x) \cdot L(\theta, x) \, \mu(dx) \end{aligned}$$

ist differenzierbar auf Θ und es gilt:

$$f'(\theta) = \frac{d}{d\theta} \int g(x) \cdot L(\theta, x) \, \mu(dx) = \int g(x) \cdot \frac{d}{d\theta} L(\theta, x) \, \mu(dx)$$

für alle $\theta \in \Theta$. (D.h. es gilt die Vertauschung von Differentiation und Integration.)

Als Hilfsmittel für den Beweis von 4.13 verwende Lemma 4.14:

4.14 Lemma

Sei $(\mathcal{X}, \mathcal{F}, \nu)$ ein Maßraum, $T : \mathcal{X} \to \mathbb{R}$ messbar, $\Theta \subseteq \mathbb{R}$ offen, $q : \Theta \to \mathbb{R}$ differenzierbar mit

$$f(\theta) := \int \exp(q(\theta) \cdot T(x)) \, \nu(dx) < \infty$$

für alle $\theta \in \Theta$. Dann $f : \Theta \to \mathbb{R}$ differenzierbar mit

$$f'(\theta) = \int T(x) \cdot q'(\theta) \cdot \exp(q(\theta) \cdot T(x)) \, \nu(dx)$$

für $\theta \in \Theta$.

Beweis:

- Zu $\theta \in \Theta$ existiert ein $h_0 > 0$ mit $B(\theta, h_0) \subseteq \Theta$.
- 1.Schritt: $q(\theta) = \theta$, dann

$$\begin{aligned}\frac{f(\theta+h)-f(\theta)}{h} &= \int \frac{\exp\left((\theta+h)\cdot T(x)\right)-\exp\left(\theta\cdot T(x)\right)}{h}\,\nu(dx)\\ &= \int e^{\theta\cdot T}\cdot \underbrace{\frac{\exp(h\cdot T)-\exp(0\cdot T)}{h}}_{\to \frac{d}{dy}e^{y\cdot T}\big|_{y=0}=T\ (h\to 0)}\,d\nu \qquad (*)\end{aligned}$$

Ziel: Anwendung des Satzes von der dominierten Konvergenz. Dazu erforderlich ist eine integrierbare Majorante.

$$\begin{aligned}\left|\frac{e^{h\cdot T}-1}{h}\right| &\le \frac{1}{|h|}\cdot\left|\sum_{i=1}^{\infty}\frac{(h\cdot T)^i}{i!}\right|\\ &\le \sum_{i=1}^{\infty}\frac{|h|^{i-1}\cdot|T|^i}{i!} \le \frac{1}{h_0}\cdot\sum_{i=0}^{\infty}\frac{h_0^i\cdot|T|^i}{i!}\\ &= \frac{1}{h_0}\cdot e^{h_0\cdot|T|} \le \frac{1}{h_0}\cdot\left(e^{h_0\cdot T}+e^{-h_0\cdot T}\right)\\ \Rightarrow \left|e^{\theta\cdot T}\cdot\frac{e^{h\cdot T}-1}{h}\right| &\le \frac{1}{h_0}\cdot\left(e^{(\theta+h_0)\cdot T}+e^{(\theta-h_0)\cdot T}\right)\end{aligned}$$

... integrierbar, da nach Voraussetzung $\theta \pm h_0 \in \Theta$. Damit Behauptung für $q(\theta) = \theta$.

- q differenzierbar, also $f(\theta) = \overline{f}(q(\theta))$ mit

$$\overline{f}(\theta) := \int e^{\theta\cdot T}\,d\nu$$

Aus Kettenregel folgt: f ist differenzierbar mit

$$\begin{aligned}f'(\theta) &= \overline{f}'(q(\theta))\cdot q'(\theta)\\ &= \int T\cdot e^{q(\theta)\cdot T}\cdot q'(\theta)\,d\nu\end{aligned}$$

Beweis 4.13:

$$\begin{aligned}f(\theta) &= \int g(x)\cdot L(\theta,x)\,\mu(dx)\\ &= c(\theta)\cdot\underbrace{\int e^{q(\theta)\cdot T}\cdot g\cdot h\,d\mu}_{=:\tilde{f}}\end{aligned}$$

(i). Ist $g \ge 0$, so ist $\nu := g\cdot h\,\mu$ ein Maß und

$$\tilde{f}(\theta) = \int e^{q(\theta)\cdot T}\,d\nu$$

Gemäß 4.14 folgt:

$$\begin{aligned}\tilde{f}'(\theta) &= \int T\cdot q'(\theta)\cdot e^{q(\theta)\cdot T}\,d\nu\\ &= \int T\cdot g\cdot h\cdot q'(\theta)e^{q(\theta)\cdot T}\,d\mu \qquad (*)\end{aligned}$$

Gemäß Bemerkung 4.7(ii):

$$\begin{aligned}c(\theta) &= \left(\int h\cdot e^{q(\theta)\cdot T}\,d\mu\right)^{-1}\\ &= \left(\int e^{q(\theta)\cdot T}\,d\sigma\right)^{-1}\end{aligned}$$

mit $\sigma := h\,\mu$. Also $c(\theta)$ differenzierbar (Integralausdruck differenzierbar nach 4.14). Aus Produktregel folgt:

$$f(\theta) = c(\theta)\cdot\tilde{f}(\theta)$$

ist differenzierbar und

$$\begin{aligned}
f'(\theta) &= c'(\theta)\cdot\tilde{f}(\theta) + c(\theta)\cdot\tilde{f}'(\theta)\\
&\overset{(*)}{=} c'(\theta)\cdot\int e^{q(\theta)\cdot T}\cdot g\cdot h\,d\mu + c(\theta)\cdot\int T\cdot g\cdot h\cdot q'(\theta)\cdot e^{q(\theta)\cdot T}\,d\mu\\
&= \int \underbrace{g\cdot h\cdot(c'(\theta)+c(\theta)\cdot q'(\theta)\cdot T)\cdot e^{q(\theta)\cdot T}}_{=\frac{d}{d\theta}L(\theta,x)}\,d\mu
\end{aligned}$$

denn

$$L(\theta,x) = c(\theta)\cdot h(x)\cdot e^{q(\theta)\cdot T(x)}$$

(ii). g beliebig, dann $g = g^+ - g^-$.

$$\begin{aligned}
f(\theta) &= \int (g^+(x) - g^-(x))\cdot L(\theta,x)\,\mu(dx)\\
&= \int g^+(x)\cdot L(\theta,x)\,\mu(dx) - \int g^-(x)\cdot L(\theta,x)\,\mu(dx)
\end{aligned}$$

Aus (i) folgt, dass f differenzierbar mit

$$\begin{aligned}
f'(\theta) &= \int g^+(x)\cdot\frac{d}{d\theta}L(\theta,x)\,\mu(dx) - \int g^-(x)\cdot\frac{d}{d\theta}L(\theta,x)\,\mu(dx)\\
&= \int (g^+(x) - g^-(x))\cdot\frac{d}{d\theta}L(\theta,x)\,\mu(dx)\\
&= \int g(x)\cdot\frac{d}{d\theta}L(\theta,x)\,\mu(dx)
\end{aligned}$$

5

Informationsungleichung

Ziel: Herleitung einer unteren Schranke für die Varianz erwartungstreuer Schätzer in sogenannten regulären (glatten) Modellen.

5.1 Voraussetzungen

(5.1) $\Omega \subseteq \mathbb{R}$ offen

(5.2) $\{(\theta, x) \in \Theta \times \mathcal{X}; L(\theta, x) > 0\} = \Theta \times A$ für ein $A \in \mathcal{F}$, d.h. alle μ-Dichten f_θ von P_θ haben denselben Träger

$$\{x \in \mathcal{X}; f_\theta(x) > 0\} =: A$$

(5.3) $L'(\theta, x) := \frac{\partial}{\partial\theta} L(\theta, x)$ existiert für alle $\theta \in \Theta$, $x \in A$

(5.4) $\ell(\theta, x) := \log L(\theta, x)$, $\theta \in \Theta$, $x \in A$. Nach (5.3): $\ell(\theta, x)$ differenzierbar nach θ und $\ell'(\theta, x) := \frac{\partial}{\partial\theta} \ell(\theta, x)$. Die Funktion $\ell'(\theta, x)$ heißt Score-Funktion. Forderung:

$$\forall \theta \in \Theta : \mathbb{E}_\theta(\ell'(\theta, X)) = 0$$

(5.5) Sei

$$\begin{aligned} I(\theta) := \mathbb{V}_\theta(\ell'(\theta, X)) &\overset{(5.4)}{=} \mathbb{E}((\ell'(\theta, X))^2) \\ &\in (0, \infty) \qquad (\theta \in \Theta) \end{aligned}$$

I heißt Fischer-Information.

Bemerkung: Voraussetzungen wie in (5.1)-(5.5) nennt man Regularitätsvoraussetzungen.

5.2 Satz

Es gelten Voraussetzungen (5.1)-(5.5), ferner sei $g : \mathcal{X} \to \mathbb{R}$ messbar (später $g = \hat{\Theta}$) mit $\mathbb{E}_\theta(g(X)^2) < \infty$ für alle $\theta \in \Theta$. Außerdem sei $f : \Theta \to \mathbb{R}$ mit

$$f(\theta) := \mathbb{E}_\theta(g(X)) \overset{\text{TFS}}{=} \int g(x) \cdot L(\theta, x)\, \mu(dx)$$

differenzierbar mit

$$f'(\theta) = \int g(x) \cdot L'(\theta, x)\, \mu(dx) \tag{5.7}$$

für alle $\theta \in \Theta$. Dann gilt:

$$\forall \theta \in \Theta : \mathbb{V}_\theta(g(X)) \geq \frac{f'(\theta)^2}{I(\theta)}$$

Beweis:

$$
\begin{aligned}
f'(\theta) &\overset{(5.7)}{=} \int g(x)\cdot L'(\theta,x)\,\mu(dx) \overset{(5.2)}{=} \int_A g(x)\cdot L'(\theta,x)\,\mu(dx)\\
&= \int_A g(x)\cdot \underbrace{\frac{L'(\theta,x)}{L(\theta,x)}}_{=\ell'(\theta,x)} \cdot \underbrace{L(\theta,x)\,\mu(dx)}_{P_\theta(dx)}\\
&\overset{\text{TFS}}{=} \mathbb{E}_\theta(g(X)\cdot \ell'(\theta,X)) \overset{(5.4)}{=} \mathbb{E}_\theta((g(X)-f(\theta))\cdot \ell'(\theta,X))\\
\Rightarrow (f'(\theta))^2 &= \mathbb{E}_\theta\Bigg(\underbrace{(g(X)-f(\theta))}_{=:u}\cdot\underbrace{\ell'(\theta,X)}_{=:v}\Bigg)^2\\
&= (|\mathbb{E}(u\cdot v)|)^2 \overset{\text{CSU}}{\le} \mathbb{E}(u^2)\cdot\mathbb{E}(v^2)\\
&= \mathbb{E}((g(X)-f(\theta))^2)\cdot\mathbb{E}(\ell'(\theta,X)^2)\\
&= \mathbb{V}_\theta(g(X))\cdot I(\theta)
\end{aligned}
$$

5.3 Korollar: Informationsungleichung

Es gelten (5.1)-(5.5), $\gamma:\Theta\to\mathbb{R}$ zu schätzende Funktion, $\hat\Theta:\mathcal{X}\to\mathbb{R}$ erwartungstreuer Schätzer für $\gamma(\theta)$ mit $\mathbb{E}_\theta(\hat\Theta(X)^2)<\infty$. Ist γ differenzierbar mit

$$\gamma'(\theta) = \int \hat\Theta(x)\cdot L'(\theta,x)\,\mu(dx) \tag{5.8}$$

so gilt

$$\mathbb{V}_\theta(\hat\Theta(X)) \ge \frac{(\gamma'(\theta))^2}{I(\theta)} \tag{5.9}$$

Beweis:

- Setze in Satz 5.2 $g := \hat\Theta$, dann

$$f(\theta) = \mathbb{E}_\theta(g(X)) = \mathbb{E}_\theta(\hat\Theta(X)) = \gamma(\theta)$$

5.4 Bemerkung

Sei $\mathcal{P}$ mit (5.1)-(5.5) derart, dass alle erwartungstreuen Schätzer für $\gamma(\theta)$ die Vertauschbarkeitsbedingung von (5.8) erfüllen. Kann man dann für einen erwartungstreuen Schätzer $\hat\Theta$ zeigen, dass

$$\forall\theta\in\Theta : \mathbb{V}_\theta(\hat\Theta(X)) = \frac{(\gamma'(\theta))^2}{I(\theta)}$$

so zeigt 5.3, dass $\hat\Theta$ gleichmäßig bester erwartungstreuer Schätzer für $\gamma(\theta)$ ist.
Zeige im Folgenden, dass 1-parametrige Exponentialfamilien obige Eigenschaften besitzen.

5.5 Satz

Sei $\mathcal{P} = \{P_\theta;\theta\in\Theta\}$ mit $\Theta\subseteq\mathbb{R}$ offen eine 1-parametrige Exponentialfamilie bzgl. μ in q und T. Ist q differenzierbar, so sind (5.1)-(5.4) erfüllt. Ferner ist für jede messbare Funktion $g:\mathcal{X}\to\mathbb{R}$ mit $\mathbb{E}_\theta|g(X)|<\infty$ für alle $\theta\in\Theta$ die Funktion

$$f(\theta) := \mathbb{E}_\theta(g(X))$$

differenzierbar und es gilt (5.7).

Beweis:

- (5.1) gilt nach Voraussetzung. Da

$$L(\theta, x) = f_\theta(x) = \underbrace{c(\theta)}_{>0} \cdot h(x) \cdot \underbrace{\exp(q(\theta) \cdot T(x))}_{>0}$$

ist (5.2) erfüllt, denn

$$\begin{aligned} \{(\theta, x) \in \Theta \times \mathcal{X}; L(\theta, x) > 0\} &= \Theta \times \underbrace{\{h > 0\}}_{=:A} \\ \Rightarrow A = h^{-1}((0, \infty)) &\in \mathcal{F} \end{aligned}$$

Beweis von 4.13 zeigt, dass $c(\theta)$ differenzierbar, also $L(\cdot, x)$ differenzierbar für alle $x \in \mathcal{X}$ und somit (5.3) erfüllt. Bleibt (5.4):

$$\begin{aligned} \mathbb{E}_\theta(\ell'(\theta, X)) &= \int_\Omega \ell'(\theta, X(w))\, \mathbb{P}_\theta(dw) \overset{\text{TFS}}{=} \int_\mathcal{X} \ell'(\theta, x)\, P_\theta(dx) \\ &= \int_A \ell'(\theta, x) \cdot L(\theta, x)\, \mu(dx) = \int_A \frac{L'(\theta, x)}{L(\theta, x)} \cdot L(\theta, x)\, \mu(dx) \\ &= \int_A \frac{\partial}{\partial\theta} L(\theta, x)\, \mu(dx) \overset{4.13}{=} \frac{\partial}{\partial\theta} \underbrace{\int L(\theta, x)\, \mu(dx)}_{1} = 0 \end{aligned}$$

Schließlich (5.7) gilt gemäß 4.13.

5.6 Satz

Sei $\mathcal{P}$ wie in Satz 5.5 mit $I(\theta) > 0$ für alle $\theta \in \Theta$. Ist $T : \mathcal{X} \to \mathbb{R}$ erwartungstreu für $\gamma(\theta)$, wobei γ differenzierbar auf Θ, so gilt

$$\forall \theta \in \Theta : \mathbb{V}_\theta(T(X)) = \frac{(\gamma'(\theta))^2}{I(\theta)}$$

Beweis:

- Nach Definition gilt

$$\begin{aligned} \ell(\theta, x) &= \log c(\theta) + \log h(x) + q(\theta) \cdot T(x) \\ \Rightarrow \ell'(\theta, x) &= \frac{c'(\theta)}{c(\theta)} + q'(\theta) \cdot T(x) \end{aligned}$$

Aus (5.4) folgt wegen Satz 5.5 außerdem

$$\begin{aligned} 0 \overset{(5.4)}{=} \mathbb{E}_\theta(\ell'(\theta, X)) &= \frac{c'(\theta)}{c(\theta)} + q'(\theta) \cdot \mathbb{E}_\theta(T(X)) \\ \Leftrightarrow \frac{c'(\theta)}{c(\theta)} &= -q'(\theta) \cdot \mathbb{E}_\theta(T(X)) \qquad (*) \end{aligned}$$

- Außerdem gilt

$$
\begin{aligned}
\gamma'(\theta) &= \frac{d}{d\theta}\mathbb{E}_\theta(T(X)) = \frac{d}{d\theta}\int_A T(x)\cdot L(\theta,x)\,\mu(dx) \\
&\overset{4.13}{=} \int_A T(x)\cdot L'(\theta,x)\,\mu(dx) \\
&= \int_A T(x)\cdot \frac{L'(\theta,x)}{L(\theta,x)}\cdot L(\theta,x)\,\mu(dx) \\
&= \int_A T(x)\cdot \ell'(\theta,x)\,P_\theta(dx) \overset{\text{TFS}}{=} \mathbb{E}_\theta(T(X)\cdot \ell'(\theta,X)) \\
&= \mathbb{E}_\theta\left(T(X)\cdot\left(\frac{c'(\theta)}{c(\theta)} + q'(\theta)\cdot T(X)\right)\right) \\
&= \frac{c'(\theta)}{c(\theta)}\cdot \mathbb{E}_\theta(T(X)) + q'(\theta)\cdot \mathbb{E}_\theta(T(X)^2) \\
&\overset{*}{=} q'(\theta)\cdot\left(\mathbb{E}_\theta(T(X)^2) - \mathbb{E}(T(X))^2\right) \\
&= q'(\theta)\cdot \mathbb{V}_\theta(T(X))
\end{aligned}
$$

Wegen

$$I(\theta) = \mathbb{V}_\theta(\ell'(\theta,X)) = (q'(\theta))^2\cdot \mathbb{V}_\theta(T(X))$$

folgt somit

$$\frac{(\gamma'(\theta))^2}{I(\theta)} = \mathbb{V}_\theta(T(X))$$

5.7 Bemerkung

Es gilt:

$$\forall\theta\in\Theta : I(\theta) > 0 \Leftrightarrow q'(\theta) \neq 0 \wedge \mathbb{V}_\theta(T(X)) > 0$$

Dann folgt:

$$
\begin{aligned}
\gamma'(\theta) &= q'(\theta)\cdot \mathbb{V}_\theta(T(X)) = \frac{I(\theta)}{q'(\theta)} \\
\Rightarrow I(\theta) &= \gamma'(\theta)\cdot q'(\theta)
\end{aligned}
$$

Spezialfall: $I(\theta) = q'(\theta)$ für $\gamma = \mathrm{id}_\Theta$. Aus 5.3-5.6 folgt sofort:

5.8 Satz

Sei $(X,\mathcal{X},\mathcal{F},\mathcal{P})$ ein statistisches Modell, $\Omega\subseteq\mathbb{R}$ offen, $\gamma:\Theta\to\mathbb{R}$ differenzierbar, $\mathcal{P}$ eine 1-parametrige Exponentialfamilie in q und T mit q differenzierbar und T erwartungstreu für $\gamma(\theta)$. Ist $I(\theta) > 0$ für alle $\theta\in\Theta$, dann gilt für alle erwartungstreuen Schätzer $\hat{\Theta}$ für $\gamma(\theta)$ mit $\mathbb{E}_\theta(\hat{\Theta}(X)^2) < \infty$ für alle $\theta\in\Theta$, dass

$$\mathbb{V}_\theta(\hat{\Theta}(X)) \geq \frac{(\gamma'(\theta))^2}{I(\theta)} = \mathbb{V}_\theta(T(X))$$

für alle $\theta\in\Theta$. Damit ist T gleichmäßig bester erwartungstreuer Schätzer für $\gamma(\theta)$, für den die untere Schranke in der Informationsungleichung angenommen wird.

5.9 Beispiel

Sei $X\sim\beta_p$, dann folgt aus 4.9, dass

$$q(p) = \log p - \log(1-p)$$

differenzierbar ist und $T(x) = x$.

$$\begin{aligned}
\mathbb{E}_p(T(X)) &= p = \gamma(p) \\
\mathbb{V}_p(T(X)) &= p \cdot (1-p) > 0 \\
q'(p) &= \frac{1}{p} + \frac{1}{1-p} \neq 0 \\
\overset{5.7}{\Rightarrow} I(p) &= q'(p) = \frac{1}{p \cdot (1-p)}
\end{aligned}$$

Somit folgt aus 5.8, dass T gleichmäßig bester erwartungstreuer Schätzer für p ist.

5.10 Beispiel

Sei $X \sim \operatorname{Exp}(\lambda)$ mit $\lambda > 0$. Aus 4.11 bekannt: $q(\lambda) = \lambda \neq 0, T(x) = x$,

$$\begin{aligned}
\mathbb{E}_\lambda(T(X)) &= \frac{1}{\lambda} =: \gamma(\lambda) \\
\mathbb{V}_\lambda(T(X)) &= \frac{1}{\lambda^2} > 0 \\
&\overset{5.7}{\Rightarrow} 0 < I(\lambda) = \gamma'(\lambda) \cdot q'(\lambda) = \frac{1}{\lambda^2}
\end{aligned}$$

Aus 5.8 folgt, dass T optimaler Schätzer für $\gamma(\lambda) = \frac{1}{\lambda}$.

5.11 Beispiel

Sei $X \sim N(\mu, \sigma_0^2)$ mit $\mu \in \mathbb{R}$ ($\sigma_0^2 > 0$ bekannt),

$$f_\mu(x) = \underbrace{\frac{1}{\sqrt{2\pi \cdot \sigma_0^2}} \cdot \exp\left(-\frac{\mu^2}{2\sigma_0^2}\right)}_{c(\mu)} \cdot \underbrace{\exp\left(-\frac{x^2}{2\sigma_0^2}\right)}_{h(x)} \cdot \exp\left(\underbrace{\frac{\mu}{\sigma_0^2}}_{q(\mu)} \cdot \underbrace{x}_{T(x)}\right)$$

also $\mathcal{P} = \{N(\mu, \sigma_0^2); \mu \in \mathbb{R}\}$ ist 1-parametrige Exponentialfamilie in q und T.

$$\begin{aligned}
\mathbb{E}_\mu(T(X)) &= \mathbb{E}_\mu(X) = \mu = \gamma(\mu) \\
\mathbb{V}_\mu(T(X)) &= \sigma_0^2 > 0 \\
q'(\mu) &= \frac{1}{\sigma_0^2} \neq 0
\end{aligned}$$

Aus 5.7 folgt, dass 5.8 anwendbar.

5.12 Beispiel

Sei $X \sim N(\mu_0, \sigma^2)$ mit $\sigma^2 \in \mathbb{R}^+$ (μ_0 bekannt).

$$f_{\sigma^2}(x) = \underbrace{\frac{1}{\sqrt{2\pi\sigma^2}}}_{c(\sigma^2)} \cdot \exp\left(\underbrace{-\frac{1}{2\sigma^2}}_{q(\sigma^2)} \cdot \underbrace{(x-\mu_0)^2}_{T(x)}\right)$$

d.h. $\mathcal{P} = \{N(\mu_0, \sigma^2); \sigma^2 > 0\}$ ist 1-parametrige Exponentialfamilie in q und T.

$$\begin{aligned}
\mathbb{E}_{\sigma^2}(T(X)) &= \mathbb{E}_{\sigma^2}((X-\mu_0)^2) = \mathbb{V}_{\sigma^2}(X) = \sigma^2 = \gamma(\sigma^2) \\
\mathbb{V}_{\sigma^2}(T(X)) &= \mathbb{V}_{\sigma^2}((X-\mu_0)^2) = \mathbb{E}_{\sigma^2}((X-\mu_0)^4) - (\mathbb{E}_{\sigma^2}((X-\mu_0)^2))^2 \\
&= 3 \cdot \sigma^4 - \sigma^4 = 2\sigma^4 > 0 \\
q'(\sigma^2) &= \frac{1}{2 \cdot (\sigma^2)^2} \neq 0
\end{aligned}$$

Mit 5.7 folgt, dass Satz 5.8 anwendbar ist.

Wie sieht das in Produktmodellen aus?

5.13 Satz

$(X, \mathcal{X}, \mathcal{F}, \mathcal{P})$ erfülle (5.1)-(5.5). Dann erfüllt das zugehörige n-fache Produktmodell $(\underline{X}_n, \mathcal{X}^n, \mathcal{F}^n, \mathcal{P}^n)$ $(n \in \mathbb{N})$ ebenfalls die entsprechenden Voraussetzungen (5.1)-(5.5). Insbesondere gilt:

$$I_n(\theta) = n \cdot I(\theta)$$

wobei $I(\theta)$ die Fischer-Information in $(X, \mathcal{X}, \mathcal{F}, \mathcal{P})$ und $I_n(\theta)$ die Fischer-Information in $(\underline{X}_n, \mathcal{X}^n, \mathcal{F}^n, \mathcal{P}^n)$ bezeichne.

Beweis:

(5.1) Ist klar.

(5.2) Sei $L(\theta, x) = f_\theta(x)$ die Likelihood-Funktion in $(X, \mathcal{X}, \mathcal{F}, \mathcal{P})$. $L_n(\theta, \underline{x}_n)$ ist Likelihood-Funktion in $(\underline{X}_n, \mathcal{X}^n, \mathcal{F}^n, \mathcal{P}^n)$. Aus 4.5:

$$L_n(\theta, \underline{x}_n) = \prod_{i=1}^n L(\theta, x_i) \qquad (\theta \in \Theta, \underline{x}_n \in \mathcal{X}^n) \tag{1}$$

Nach Voraussetzung (5.2) für das einfache Modell gilt:

$$\{(\theta, x) \in \Theta \times \mathcal{X}; L(\theta, x) > 0\} = \Theta \times A \qquad (A \in \mathcal{F})$$

Somit:

$$\begin{aligned}
&\{(\theta, \underline{x}_n) \in \Theta \times \mathcal{X}^n; L(\theta, \underline{x}_n) = 0\} \\
&= \left\{(\theta, \underline{x}_n) \in \Theta \times \mathcal{X}^n; \prod_{i=1}^n L(\theta, x_i) = 0\right\} \\
&= \Theta \times A^n
\end{aligned}$$

mit $A^n \in \mathcal{F}^n$.

(5.3) Folgt aus (1).

(5.4) Es gilt:

$$\begin{aligned}
\ell_n(\theta, \underline{x}_n) &= \log L_n(\theta, \underline{x}_n) = \sum_{i=1}^n \ell(\theta, x_i) \\
\Rightarrow \ell_n'(\theta, \underline{x}_n) &= \sum_{i=1}^n \ell'(\theta, x_i) \qquad (2) \\
\Rightarrow \mathbb{E}_\theta(\ell_n'(\theta, \underline{X}_n)) &= \mathbb{E}\left(\sum_{i=1}^n \ell'(\theta, X_i)\right) = \sum_{i=1}^n \mathbb{E}(\ell'(\theta, X_i)) \\
&\overset{(5.4)}{=} 0
\end{aligned}$$

(5.5) Gemäß Definition:

$$\begin{aligned}
I_n(\theta) &= \mathbb{V}_\theta(\ell_n'(\theta, \underline{X}_n)) \overset{(2)}{=} \mathbb{V}_\theta\left(\sum_{i=1}^n \ell'(\theta, X_i)\right) \\
&\overset{\text{iid}}{=} \sum_{i=1}^n \underbrace{\mathbb{V}_\theta(\ell'(\theta, X_i))}_{=I(\theta)>0} = n \cdot I(\theta) > 0
\end{aligned}$$

Jetzt Formulierung von 5.8 für n-faches Produktmodell:

5.14 Satz

Es gelten Voraussetzungen und Bezeichnungen von 5.8. Dann gilt für das zugehörige Produktmodell $(\underline{X}_n, \mathcal{X}^n, \mathcal{F}^n, \mathcal{P}^n)$:

(1) $T_n(\underline{x}_n) \coloneqq \frac{1}{n} \cdot \sum_{i=1}^n T(x_i)$ ist erwartungstreuer Schätzer für $\gamma(\theta)$.

(2) Für alle erwartungstreuen Schätzer $\hat{\Theta}_n : \mathcal{X}^n \to \mathbb{R}$ für $\gamma(\theta)$ gilt:

$$\mathbb{V}_\theta(\hat{\Theta}_n(\underline{X}_n)) \geq \frac{(\gamma'(\theta))^2}{n \cdot I(\theta)} = \mathbb{V}_\theta(T_n(\underline{X}_n))$$

für alle $\theta \in \Theta$.

Also ist $T_n : \mathcal{X}^n \to \mathbb{R}$ gleichmäßig bester erwartungstreuer Schätzer für $\gamma(\theta)$.

Beweis:

(1) Es gilt:

$$\mathbb{E}_\theta(T_n(\underline{X}_n)) = \frac{1}{n} \cdot \sum_{i=1}^n \underbrace{\mathbb{E}_\theta(T(X_i))}_{=\gamma(\theta)} = \gamma(\theta)$$

(2) Gemäß 4.8 ist $\mathcal{P}^n$ eine 1-parametrige Exponentialfamilie in q und $\sum_{i=1}^n T(x_i)$ und damit auch in $n \cdot q$ und $\frac{1}{n} \cdot \sum_{i=1}^n T(x_i) = T_n(\underline{x}_n)$. Aus $I(\theta) > 0$ folgt nach 5.13 auch

$$I_n(\theta) = n \cdot I(\theta) > 0$$

Aus 5.8 und 5.13:

$$\begin{aligned}\mathbb{V}_\theta(\hat{\Theta}_n(\underline{X}_n)) &\overset{5.8}{\geq} \frac{(\gamma'(\theta))^2}{I_n(\theta)} \overset{5.13}{=} \frac{(\gamma'(\theta))^2}{n \cdot I(\theta)} \\ &\overset{5.8}{=} \mathbb{V}_\theta(T_n(\underline{X}_n))\end{aligned}$$

5.15 Beispiel

Seien $X_1, \ldots, X_n$ iid mit $X_i \sim \beta_p$ $(i = 1, \ldots, n)$ und $p \in (0,1)$. Aus 5.9 und 5.14:

$$\overline{X}_n \coloneqq \frac{1}{n} \cdot \sum_{i=1}^n X_i$$

ist gleichmäßig bester erwartungstreuer Schätzwert für p.

5.16 Beispiel

Seien $X_1, \ldots, X_n$ iid mit $X_i \sim \text{Exp}(\lambda)$ für $i = 1, \ldots, n$ $(\lambda > 0)$. Aus 5.10 und 5.14: $\overline{X}_n$ ist gleichmäßig bester erwartungstreuer Schätzer für $\frac{1}{\lambda}$. Beachte $\frac{1}{\overline{X}_n}$ (Maximum-Likelihood-Schätzer für λ) ist nicht erwartungstreu für λ, denn:

$$\mathbb{E}\left(\frac{1}{\overline{X}_n}\right) \overset{(!)}{=} \underbrace{\frac{n}{n-1}}_{\to 1\ (n \to \infty)} \cdot \lambda \neq \lambda$$

aber $\overline{X}_n^{-1}$ ist asymptotisch erwartungstreu für λ, d.h.

$$\forall \lambda > 0 : \mathbb{E}_\lambda\left(\frac{1}{\overline{X}_n}\right) \to \lambda \qquad (n \to \infty)$$

5.17 Beispiel

Seien $X_1, \dots, X_n$ iid mit $X_i \sim N(\mu, \sigma_0^2)$ für $i = 1, \dots, n$ mit $\sigma_0^2 > 0$ bekannt. Aus 5.11 und 5.14: $\overline{X}_n$ ist optimal für μ.

5.18 Beispiel

Seien $X_1, \dots, X_n$ iid mit $X_i \sim N(\mu_0, \sigma^2)$ für $i = 1, \dots, n$ mit $\mu_0 \in \mathbb{R}$ bekannt. Aus 5.12 und 5.14:

$$\frac{1}{n} \cdot \sum_{i=1}^{n} (X_i - \mu_0)^2$$

optimal für σ^2.

5.19 Bemerkung

Die Schätzer in 5.15, 5.17 und 5.18 sind ML-Schätzer.

6 Die Prüfverteilungen

Ziel: Zusammenstellung wichtiger Verteilungen in der Statistik.

Konvention: Sei Q eine Verteilung. Bezeichne mit Q auch eine nach Q verteilte Zufallsvariable, also: Q ist Zufallsvariable mit Verteilung Q. Beispiel:

$$\mathbb{E}(N(0,1)^{2k}) = 1 \cdot 3 \cdots (2k-1)$$

6.1 Definition

Seien $X_1, \ldots, X_m$ iid mit $X_i \sim N(0,1)$ für $i = 1, \ldots, m$. Die Verteilung von $\sum_{i=1}^m X_i^2$ heißt χ^2-Verteilung mit m Freiheitsgraden. Notation: χ_m^2.

6.2 Bemerkung

(1) Charakteristik:

$$\begin{aligned}
\mathbb{E}(\chi_m^2) &= \mathbb{E}\left(\sum_{i=1}^m X_i^2\right) = m \cdot \underbrace{\mathbb{E}(N(0,1)^2)}_{=1} = m \\
\mathbb{V}(\chi_m^2) &= \mathbb{V}\left(\sum_{i=1}^m X_i^2\right) \stackrel{\text{iid}}{=} \sum_{i=1}^m \mathbb{V}(X_i^2) = m \cdot \mathbb{V}(N(0,1)^2) \\
&= m \cdot (\mathbb{E}(N(0,1)^4) - (\mathbb{E}(N(0,1)))^2) = m \cdot (3-1) = 2m
\end{aligned}$$

(2) Es gilt:

$$\begin{aligned}
\frac{\chi_m^2 - m}{\sqrt{2m}} = \frac{\sum_{i=1}^m (X_i^2 - \mathbb{E}(X_i^2))}{\sqrt{m \cdot \mathbb{V}(X_1^2)}} = \frac{1}{\sqrt{m}} \cdot \sum_{i=1}^m \underbrace{\frac{X_i^2 - \mathbb{E}(X_i^2)}{\sqrt{\mathbb{V}(X_i^2)}}}_{=:\xi_i} \\
\stackrel{\text{i.V.}}{\to} N(0,1) \qquad (m \to \infty)
\end{aligned}$$

nach zentralem Grenzwertsatz.

6.3 Satz

χ_m^2 hat Lebesgue-Dichte

$$f_{\chi_m^2}(x) = 1_{(0,\infty)}(x) \cdot \frac{1}{2^{\frac{m}{2}} \cdot \Gamma\left(\frac{m}{2}\right)} \cdot x^{\frac{m}{2}-1} \cdot e^{-\frac{x}{2}}$$

Dabei ist

$$\Gamma(y) := \int_0^\infty t^{y-1} \cdot e^{-t}\, dt \qquad (y > 0)$$

Beweis: MAST.

6.4 Bemerkung

Setze $m = 2$. Dann

$$f_{\chi_2^2}(x) = 1_{(0,\infty)}(x) \cdot \frac{1}{2} \cdot e^{-\frac{x}{2}}$$

d.h. $f_{\chi_2^2}$ ist Dichte von $\operatorname{Exp}\left(\frac{1}{2}\right)$. Damit: $X^2 + Y^2 \sim \operatorname{Exp}\left(\frac{1}{2}\right)$, falls $X \sim N(0,1)$ und $Y \sim N(0,1)$ unabhängig.

6.5 Definition

Seien χ_m^2 und χ_n^2 unabhängig. Dann heißt die Verteilung von

$$\frac{\frac{1}{m} \cdot \chi_m^2}{\frac{1}{n}\chi_n^2}$$

F-Verteilung mit (m,n) Freiheitsgraden. Notation: $F_{m,n}$

6.6 Satz

$F_{m,n}$ hat Lebesgue-Dichte

$$f_{F_{m,n}}(x) = 1_{(0,\infty)}(x) \cdot \frac{\Gamma\left(\frac{n+m}{2}\right)}{\Gamma\left(\frac{m}{2}\right) \cdot \Gamma\left(\frac{n}{2}\right)} \cdot m^{\frac{m}{2}} \cdot n^{\frac{n}{2}} \cdot \frac{x^{\frac{m}{2}-1}}{(n + m \cdot x)^{\frac{n+m}{2}}}$$

Beweis:

- Es ist

$$\frac{\frac{1}{m} \cdot \chi_m^2}{\frac{1}{n} \cdot \chi_n^2} = \underbrace{\frac{n}{m}}_{=:\alpha} \cdot \underbrace{\frac{\chi_m^2}{\chi_n^2}}_{=:Z}$$

  Da $Z > 0$ fast sicher sei o.E. $x > 0$. Es bezeichne f_X die Dichte einer Zufallsvariablen X. Dann folgt:

$$\begin{aligned}
f_Z(x) &\overset{\text{MAST}}{=} \int_{\mathbb{R}} f_{(\chi_m^2, \chi_n^2)}(x \cdot y, y) \cdot |y| \, dy \\
&\overset{\text{unabh}}{=} \int_{(0,\infty)} f_{\chi_m^2}(x \cdot y) \cdot f_{\chi_n^2}(y) \cdot y \, dy \\
&\overset{6.3}{=} \underbrace{\frac{1}{2^{\frac{m}{2}} \cdot \Gamma\left(\frac{m}{2}\right) \cdot 2^{\frac{n}{2}} \cdot \Gamma\left(\frac{n}{2}\right)}}_{=:\frac{1}{C_{m,n}}} \cdot \int_0^\infty (x \cdot y)^{\frac{m}{2}-1} \cdot e^{-\frac{x \cdot y}{2}} \cdot y^{\frac{n}{2}-1} \cdot e^{-\frac{y}{2}} \cdot y \, dy \\
&= \frac{1}{C_{m,n}} \cdot x^{\frac{m}{2}-1} \cdot \int_0^\infty y^{\frac{n+m}{2}-1} \cdot e^{-\frac{y}{2}\cdot(x+1)} \, dy \\
&\overset{t := \frac{y}{2}\cdot(x+1)}{=} \frac{1}{C_{m,n}} \cdot x^{\frac{m}{2}-1} \cdot \int_0^\infty 2^{\frac{m+n}{2}-1} \cdot (1+x)^{-\frac{m+n}{2}+1} \cdot t^{\frac{m+n}{2}-1} \cdot e^{-t} \cdot \frac{2}{1+x} \, dt \\
&= \frac{1}{\Gamma\left(\frac{m}{2}\right) \cdot \Gamma\left(\frac{n}{2}\right)} \cdot x^{\frac{m}{2}-1} \cdot (1+x)^{-\frac{m+n}{2}} \cdot \underbrace{\int_0^\infty t^{\frac{m+n}{2}-1} \cdot e^{-t} \, dt}_{=\Gamma\left(\frac{m+n}{2}\right)} \\
&= \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right) \cdot \Gamma\left(\frac{n}{2}\right)} \cdot \frac{x^{\frac{m}{2}-1}}{(1+x)^{\frac{m+n}{2}}}
\end{aligned}$$

  Wegen $f_{\alpha \cdot Z}(x) = \frac{1}{\alpha} \cdot f_Z\left(\frac{x}{\alpha}\right)$ (Übung 4) folgt die Behauptung.

6.7 Definition

$N(0,1)$ und χ_n^2 seien unabhängig. Die Verteilung von

$$\frac{N(0,1)}{\sqrt{\frac{1}{n}\cdot\chi_n^2}}$$

heißt T-Verteilung mit n Freiheitsgraden. Notation: t_n

Ziel: Herleitung der Dichte von t_n

6.8 Definition/Satz

(i). Eine reelle Zufallsvariable $X:(\Omega,\mathcal{A},\mathbb{P})\to(\mathbb{R},\mathcal{B}(\mathbb{R}))$ mit $\mathbb{P}\circ X^{-1}=\mathbb{P}\circ(-X^{-1})$ heißt symmetrisch um Null (verteilt). Notation: $\mathcal{L}(X)=\mathcal{L}(-X)$

(ii). Ist X Zufallsvariable mit stetiger Verteilungsfunktion F. Dann:

  (a) X symmetrisch um Null $\Leftrightarrow 1-F(x)=F(-x)$ für alle $x\in\mathbb{R}$.

  (b) X symmetrisch um Null $\Leftrightarrow 1-F(x)=F(-x)$ für alle $x>0$.

Beweis:

(ii). $\mathbb{P}_X=\mathbb{P}_{-X}\Leftrightarrow\forall x\in\mathbb{R}:\mathbb{P}[X\le x]=\mathbb{P}[-X\le x]$ nach Maßeindeutigkeitssatz. Nach Definition folgt:

$$\begin{aligned}&\mathbb{P}_X=\mathbb{P}_{-X}\\ \Leftrightarrow\ &F(x)=\mathbb{P}[X\ge -x]=1-\mathbb{P}[X<-x]=1-F((-x)-0)\overset{\text{F stetig}}{=}1-F(-x)\\ \Leftrightarrow\ &(a)\end{aligned}$$

Aus (a) folgt offensichtlich (b). Noch zu zeigen: $(b)\Rightarrow(a)$. Sei $x<0$, dann $-x>0$, also nach (b):

$$1-F(-x)=F(-(-x))=F(x)$$

Für $x=0$ folgt die Behauptung aus der Stetigkeit von F mit Grenzübergang $x\downarrow 0$.

6.9 Korollar

Sei X Zufallsvariable mit Lebesgue-Dichte $f:=f_X$ und F differenzierbar. Dann sind äquivalent:

(i). X symmetrisch um Null

(ii). $\forall x\in\mathbb{R}:f(x)=f(-x)$

(iii). $\forall x>0:f(x)=f(-x)$

Beweis:

- $(i)\Rightarrow(ii)$:
  Nach 6.8 gilt $1-F(x)=F(-x)$ für alle $x\in\mathbb{R}$. Differentiation ergibt mit $F'=f$:

  $$\forall x\in\mathbb{R}:-f(x)=-f(-x)\Leftrightarrow f(x)=f(-x)$$

- $(ii)\Rightarrow(iii)$: Klar.

- $(iii) \Rightarrow (i)$:
  Sei $x > 0$. Beachte:

$$\begin{aligned}
1 &= \int_{-\infty}^{\infty} f(x)\,dx = \int_{-\infty}^{0} f(x)\,dx + \int_{0}^{\infty} f(x)\,dx \\
&= -\int_{\infty}^{0} f(-t)\,dt + \int_{0}^{\infty} f(x)\,dx \\
&= \int_{0}^{\infty} f(-t)\,dt + \int_{0}^{\infty} f(x)\,dx \\
&\overset{(iii)}{=} \int_{0}^{\infty} f(t)\,dt + \int_{0}^{\infty} f(x)\,dx \\
&= 2\cdot\int_{0}^{\infty} f(t)\,dt = 2\cdot\int_{-\infty}^{0} f(t)\,dt \\
\Rightarrow F(0) &= \int_{-\infty}^{0} f(t)\,dt = \frac{1}{2}
\end{aligned}$$

  Damit folgt für alle $x > 0$:

$$\begin{aligned}
F(x) &= \int_{-\infty}^{x} f(t)\,dt = \frac{1}{2} + \int_{0}^{x} f(t)\,dt \\
&\overset{(iii)}{=} \frac{1}{2} + \int_{0}^{x} f(-t)\,dt \overset{u=-t}{=} \frac{1}{2} + \int_{-x}^{0} f(u)\,du \\
&= \frac{1}{2} + \int_{-\infty}^{0} f(u)\,du - \int_{-\infty}^{-x} f(u)\,du \\
&= \frac{1}{2} + \frac{1}{2} - F(-x) = 1 - F(-x)
\end{aligned}$$

  Mit Satz 6.8 folgt die Behauptung.

6.10 Beispiel

$N(0,\sigma^2)$ symmetrisch um 0 für alle $\sigma^2 > 0$.

Berechnung der Dichte einer symmetrisch um Null verteilten Zufallsvariablen:

6.11 Satz

Sei X symmetrisch um Null mit $\mathbb{P}[X = 0] = 0$. Ferner existiere $g \geq 0$ mit

$$\mathbb{P}[0 < X \leq x] = \int_{0}^{x} g(u)\,du$$

Dann ist

$$f(x) := \begin{cases} g(x) & x \geq 0 \\ g(-x) & x < 0 \end{cases}$$

Lebesgue-Dichte von X.

Beweis:

- Es gilt:

$$\begin{aligned}
1 &= \underbrace{\mathbb{P}[X<0]}_{=\mathbb{P}[-X<0]} + \underbrace{\mathbb{P}[X=0]}_{=0} + \mathbb{P}[X>0] \\
&\overset{\text{symm}}{=} \mathbb{P}[X>0] + \mathbb{P}[X>0] = 2\cdot\mathbb{P}[X>0] = 2\cdot\mathbb{P}[X<0] \\
\Rightarrow \frac{1}{2} &= \mathbb{P}[X<0] = \mathbb{P}[X>0]
\end{aligned}$$

Also:

$$\begin{aligned}\frac{1}{2} &= \lim_{x\to\infty} \mathbb{P}[0 < X \le x] = \lim_{x\to\infty} \int_0^x g(u)\,du \\ &= \lim_{x\to\infty} \int_{\mathbb{R}} 1_{(0,x]}(u)\cdot g(u)\,du \overset{\text{mK}}{=} \int_0^\infty g(u)\,du \\ &= \int_0^\infty f(u)\,du \overset{t=-u}{=} \int_{-\infty}^0 f(t)\,dt\end{aligned}$$

denn $f(u) = f(-u)$ für alle $u \in \mathbb{R}$. Somit folgt für alle $x > 0$:

$$\begin{aligned}\mathbb{P}[X \le x] &= \mathbb{P}[X < 0] + \mathbb{P}[0 < X \le x] = \frac{1}{2} + \int_0^x f(u)\,du \\ &= \int_{-\infty}^0 f(u)\,du + \int_0^x f(u)\,du = \int_{-\infty}^x f(u)\,du\end{aligned}$$

Für $x \le 0$ geht man analog vor.

6.12 Satz

t_n hat die Dichte

$$f_{t_n}(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\cdot\sqrt{\pi\cdot n}} \cdot \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}$$

Beweis:

- Nach Definition ist $t_n = \frac{X}{Y}$ mit $X \sim N(0,1)$, $Y \sim \sqrt{\frac{1}{n}\cdot\chi_n^2}$ und X, Y unabhängig. Gemäß 6.10: $\mathbb{P}\circ X^{-1} = \mathbb{P}\circ(-X)^{-1}$. Damit:

$$\begin{aligned}\mathbb{P}\circ(X,Y)^{-1} &\overset{X \perp\!\!\!\perp Y}{=} \mathbb{P}\circ X^{-1} \otimes \mathbb{P}\circ Y^{-1} = \mathbb{P}\circ(-X)^{-1} \otimes \mathbb{P}\circ Y^{-1} \\ &\overset{-X \perp\!\!\!\perp Y}{=} \mathbb{P}\circ(-X,Y)^{-1}\end{aligned}$$

  d.h. $(X,Y) \overset{\mathcal{L}}{=} (-X,Y)$. Damit

$$t_n = \frac{X}{Y} \overset{\mathcal{L}}{=} \frac{-X}{Y} = -t_n$$

  also t_n symmetrisch um Null nach Definition. Weiterhin gilt

$$\mathbb{P}[t_n = 0] = \mathbb{P}\left[\frac{X}{Y} = 0\right] = \mathbb{P}[X = 0] = 0$$

  Somit sind Voraussetzungen von Satz 6.11 erfüllt.

$$\begin{aligned}\mathbb{P}[0 < t_n \le x] &\overset{\text{symm}}{=} \frac{1}{2}\cdot\mathbb{P}[-x \le t_n \le x] = \frac{1}{2}\cdot\mathbb{P}[t_n^2 \le x^2] \\ &= \frac{1}{2}\cdot\mathbb{P}\left[\frac{(N(0,1))^2}{\frac{1}{n}\cdot\chi_n^2} \le x^2\right] \\ &= \frac{1}{2}\cdot\mathbb{P}[F_{1,n} \le x^2] = \frac{1}{2}\cdot\int_0^{x^2} f_{F_{1,n}}(u)\,du \\ &\overset{y=\sqrt{u}}{=} \int_0^x \underbrace{f_{F_{1,n}}(y^2)\cdot y}_{=:g(y)}\,dy\end{aligned}$$

  Mit 6.11 folgt die Behauptung.

7
Konfidenzintervalle

Sei $(X, \mathcal{X}, \mathcal{F}, \mathcal{P})$ ein statistisches Modell mit

$$\mathcal{P} = \{P_\theta : \theta \in \Theta\} \qquad (\Theta \subseteq \mathbb{R}^d)$$

Sei $\gamma : \Theta \to \mathbb{R}$ die zu schätzende Funktion und $\hat{\Theta} : \mathcal{X} \to \mathbb{R}$ ein Schätzer für $\gamma(\theta)$. Bisher: Auswahl von $\hat{\Theta}$ nach bestimmten Gütekriterien (Erwartungstreue, Minimierung des mittleren quadratischen Schätzfehlers). Dabei geleitet von der Vorstellung: Für gute Schätzer sollte mit großer Wahrscheinlichkeit für den zufälligen Schätzwert $\hat{\Theta}(X) \approx \gamma(\theta_0)$ gelten, wobei θ_0 wahrer Parameter. Gemäß Häufigkeitsinterpretation: Die meisten konkreten Schätzwerte $\hat{\theta}(x)$ liegen nahe bei $\gamma(\theta_0)$. Ziel: Präzisierung unserer Forderungen.

7.1 Beispiel

Seien $X_1, \ldots, X_n$ iid mit $X_i \sim N(\mu, \sigma^2)$, wobei $\mu \in \mathbb{R}$ unbekannt und $\sigma^2 > 0$ bekannt. Gemäß 5.19 ist

$$\overline{X}_n = \frac{1}{n} \cdot \sum_{i=1}^{n} X_i$$

MLS für μ und gemäß 5.17 überdies optimal. Trotzdem gilt:

$$\mathbb{P}_\mu[\overline{X}_n \neq \mu] = 1$$

für alle $\mu \in \mathbb{R}$. Mit Wahrscheinlichkeit 1 liefert $\overline{X}_n$ den falschen Wert. Kompromiss zwischen der sogenannten Punktschätzung: Finde ein zufälliges Intervall $I \subseteq \mathbb{R}$ (möglichst klein) mit

$$\mathbb{P}_\mu[\mu \in I] \geq 1 - \alpha \tag{7.1}$$

für alle $\mu \in \mathbb{R}$. Dabei ist $1 - \alpha$ eine vorgegebene Schranke, die als „groß" erachtet wird. Mit einer Wahrscheinlichkeit von mindestens $1 - \alpha$ enthält das zufällige I den wahren Parameter μ. Typische Konstruktion von I:

$$\overline{X}_n - \mu = \sum_{i=1}^{n} \underbrace{\frac{X_i - \mu}{n}}_{\sim N\left(0, \frac{\sigma^2}{n^2}\right)} \overset{\perp\!\!\!\perp}{\sim} N\left(0, \frac{\sigma^2}{n}\right)$$

Damit folgt für alle $\mu \in \mathbb{R}$, $a < b \in \mathbb{R}$:

$$\begin{aligned}\mathbb{P}_\mu[a \leq \overline{X}_n - \mu \leq b] &= \mathbb{P}\left[a \leq \frac{\sigma}{\sqrt{n}} \cdot N(0,1) \leq b\right] = \mathbb{P}\left[a \cdot \frac{\sqrt{n}}{\sigma} \leq N(0,1) \leq b \cdot \frac{\sqrt{n}}{\sigma}\right] \\ &= \Phi\left(b \cdot \frac{\sqrt{n}}{\sigma}\right) - \Phi\left(\frac{\sqrt{n}}{\sigma} \cdot a\right)\end{aligned}$$

mit Φ Verteilungsfunktion von $N(0,1)$. Wähle $a^* < b^*$ so, dass

$$\Phi\left(b^* \cdot \frac{\sqrt{n}}{\sigma}\right) - \Phi\left(\frac{\sqrt{n}}{\sigma} \cdot a^*\right) = 1 - \alpha \tag{7.2}$$

Dann

$$\begin{aligned}
&\forall \mu \in \mathbb{R} : \mathbb{P}_\mu[a^* \le \overline{X}_n - \mu \le b^*] = 1-\alpha \\
\Leftrightarrow\ &\forall \mu \in \mathbb{R} : \mathbb{P}_\mu\left[\overline{X}_n - b^* \le \mu \le \overline{X}_n - a^*\right] = 1-\alpha \\
\Leftrightarrow\ &\forall \mu \in \mathbb{R} : \mathbb{P}_\mu[\mu \in \underbrace{[\overline{X}_n - b^*, \overline{X}_n - a^*]}_{=:I_n(\underline{X}_n)}] = 1-\alpha
\end{aligned}$$

d.h. $\mathbb{P}[\mu \in I_n(\underline{X}_n)] = 1-\alpha$ für alle $\mu \in \mathbb{R}$.

Klar: Für alle $\alpha \in (0,1)$ existieren (sogar unendlich viele) $a^* < b^*$ mit (7.2). Wünschenswert: $I = I(a^*, b^*)$ mit minimaler Länge $b^* - a^*$. Lösung später. Ebenfalls wünschenswert: I symmetrisch um $\overline{X}_n$, dann $a^* = -b^*$ notwendig. Dann

$$\begin{aligned}
(7.2) &\Leftrightarrow \Phi\left(\frac{\sqrt{n}}{\sigma}\cdot b^*\right) - \Phi\left(-\frac{\sqrt{n}}{\sigma}\cdot b^*\right) = 1-\alpha \\
&\overset{6.8,6.10}{\Leftrightarrow} 2\cdot\Phi\left(\frac{\sqrt{n}}{\sigma}\cdot b^*\right) - 1 = 1-\alpha \\
&\Leftrightarrow \Phi\left(\frac{\sqrt{n}}{\sigma}\cdot b^*\right) = 1-\frac{\alpha}{2} \\
&\Leftrightarrow b^* = \frac{\sigma}{\sqrt{n}}\cdot\Phi^{-1}\left(1-\frac{\alpha}{2}\right)
\end{aligned}$$

Man erhält das um $\overline{X}_n$ symmetrische Intervall

$$I_n = \left[\overline{X}_n - \frac{\sigma}{\sqrt{n}}\cdot u_{1-\frac{\alpha}{2}}, \overline{X}_n + \frac{\sigma}{\sqrt{n}}\cdot u_{1-\frac{\alpha}{2}}\right] \tag{7.3}$$

mit $u_\beta := \beta$-Quantil von $N(0,1)$, d.h. $\Phi(u_\beta) = \beta$, d.h. $u_\beta = \Phi^{-1}(\beta)$.

7.2 Definition

Sei $(X, \mathcal{X}, \mathcal{F}, \mathcal{P})$ mit $\mathcal{P} = \{P_\theta; \theta \in \Theta\}$, $\Theta \subseteq \mathbb{R}^d$, $\gamma : \Theta \to \mathbb{R}$, $\alpha \in (0,1)$. $\mathcal{C} : \mathcal{X} \to \{I; I \subseteq \mathbb{R} \text{ Intervall }\}$ heißt Konfidenzintervall für $\gamma(\theta)$ zum Konfidenzniveau $1-\alpha$:$\Leftrightarrow$

(i). $F_\theta := \{x \in \mathcal{X}; \gamma(\theta) \in \mathcal{C}(x)\} \in \mathcal{F}$ für alle $\theta \in \Theta$

(ii). $\mathbb{P}_\theta[\gamma(\theta) \in \mathcal{C}(X)] \ge 1-\alpha$ für alle $\theta \in \Theta$

Für $x \in \mathcal{X}$ heißt $\mathcal{C}(x)$ konkretes Konfidenzintervall zur Beobachtung $x \in \mathcal{X}$.

7.3 Bemerkung

(i). $[\gamma(\theta) \in \mathcal{C}(X)] = X^{-1}(F_\theta) \in \mathcal{A}$, da $X : (\Omega, \mathcal{A}) \to (\mathcal{X}, \mathcal{F})$-messbar und $F_\theta \in \mathcal{F}$ gemäß 7.2(i). Damit sind die Wahrscheinlichkeiten in 7.2(ii) wohldefiniert.

(ii). „$\ge$“ in 7.2(ii), weil nicht alle Verteilungen $\mathbb{P} \circ X^{-1}$ stetig sind.

(iii). Klar: $\mathcal{C}(x) = \mathbb{R}$ erfüllt 7.2(i) und (ii), enthält aber keine Information über die Lage von $\gamma(\theta)$. In der Tat sind im Gegensatz möglichst kurze Konfidenzintervalle gesucht.

(iv). α in der Praxis „klein“, konventionelle Werte: $\alpha \in \{0.1, 0.01, 0.05, 0.001\}$

(v). Seien $u, o : \mathcal{X} \to \mathbb{R}$ $\mathcal{F}$-messbar mit $u \le 0$, dann erfüllt

$$\mathcal{C}(x) := [u(x), o(x)]$$

die Bedingung 7.2(i), denn

$$\begin{aligned}
\{x \in \mathcal{X}; \gamma(\theta) \in \mathcal{C}(x)\} &= \{x \in \mathcal{X}; u(x) \le \gamma(\theta) \le o(x)\} \\
&= u^{-1}((-\infty, \gamma(\theta)]) \cap o^{-1}([\gamma(\theta), \infty)) \in \mathcal{F}
\end{aligned}$$

(vi). Wie in Beispiel 7.1 sind Konfidenzintervalle häufig vom Typ $[\hat{\Theta}(x) - b, \hat{\Theta}(x) + a]$, wobei $\hat{\Theta}$ ein („guter“) Schätzer ist.

(vii). Häufigkeitsinterpretation von 7.2(ii): Wird das Konfidenzintervall $\mathcal{C}$ zum Niveau $1-\alpha$ in der Praxis sehr oft zum Berechnen konkreter Konfidenzintervalle $\mathcal{C}(x)$ angewendet, so liegt in mindestens $100\cdot(1-\alpha)$ % der Fälle $\gamma(\theta) \in \mathcal{C}(x)$. Beachte: Im Einzelfall, d.h. für ein konkretes Konfidenzintervall gilt, entweder $\gamma(\theta) \in \mathcal{C}(x)$ oder nicht. Insbesondere ist die Sprechweise „Das konkrete Konfidenzintervall enthält mit Wahrscheinlichkeit von $1-\alpha$ das unbekannte $\gamma(\theta)$.“ ist falsch!

Häufigkeitsinterpretation im Detail: Sei $X : (\Omega, \mathcal{A}, \mathbb{P}_\theta) \to (\mathcal{X}, \mathcal{F})$ und seien $X_1, \dots, X_N, \dots$ iid mit $X_i \overset{\mathcal{L}}{=} X$. Sei $\mathcal{C} : \mathcal{X} \to \{I; I \subseteq \mathbb{R} \text{ Intervall }\}$ vorgegeben. Dies liefert $\mathcal{C}(X_1), \dots, \mathcal{C}(X_N), \dots$.

$$\frac{1}{N}\sum_{i=1}^{N} 1_{[\gamma(\theta)\in\mathcal{C}(X_i)]} \to \mathbb{P}[\gamma(\theta) \in \mathcal{C}(X)] \geq 1-\alpha \qquad (N \to \infty)$$

7.4 Beispiel

Minimiere die Länge des Konfidenzintervalls $I = [\overline{X}_n - b, \overline{X}_n - a]$ aus Beispiel 7.1. Gemäß Konstruktion erfüllen $a < b$ die Gleichung (7.2). Somit gesucht

$$\begin{aligned}(a_0, b_0) &:= \operatorname{argmin}\left\{b-a; \Phi\left(\frac{\sqrt{n}}{\sigma}\cdot b\right) - \Phi\left(\frac{\sqrt{n}}{\sigma}\cdot a\right) = 1-\alpha\right\}\\ &\overset{\text{Ü6,2a}}{=} \frac{\sigma}{\sqrt{n}}\cdot(y_0, x_0)\end{aligned}$$

wobei

$$(y_0, x_0) := \operatorname{argmin}\{x-y; \Phi(x) - \Phi(y) = 1-\alpha\}$$

Setze $y(x) := \Phi^{-1}(\Phi(x) - 1 + \alpha)$ mit $x > \Phi^{-1}(1-\alpha) = u_{1-\alpha}$. Aus Übung 6 folgt $(y_0, x_0) = (y(\tilde{x}_0), \tilde{x}_0)$, wobei

$$\tilde{x}_0 := \operatorname{argmin}\{x - y(x); x > \Phi^{-1}(1-\alpha)\}$$

Sei $\ell(x) := x - y(x)$, dann ist ℓ stetig auf $(u_{1-\alpha}, \infty)$ mit $\ell(x) \to \infty$ für $x \to u_{1-\alpha}$ und für $x \to \infty$. Also existiert eine Minimalstelle, für die notwendigerweise gilt $\ell'(x) = 0$.

$$\begin{aligned}0 = \ell'(x) &= 1 - y'(x)\\ &= 1 - \frac{1}{\Phi'(\Phi^{-1}(\Phi(x)-1+\alpha))}\cdot\underbrace{\Phi'(x)}_{\varphi(x)}\\ &= 1 - \frac{1}{\varphi(\Phi^{-1}(\Phi(x)-1+\alpha))}\cdot\underbrace{\varphi(x)}_{>0}\\ \Leftrightarrow \varphi(x) &= \varphi(\Phi^{-1}(\Phi(x)-1+\alpha))\\ \Leftrightarrow x &= \pm\Phi^{-1}(\Phi(x)-1+\alpha)\\ \Rightarrow x &= -\Phi^{-1}(\Phi(x)-1+\alpha)\\ \Leftrightarrow -x &= \Phi^{-1}(\Phi(x)-1+\alpha)\\ \Leftrightarrow \underbrace{\Phi(-x)}_{\overset{6.9}{=}1-\Phi(x)} &= \Phi(x)-1+\alpha\\ \Leftrightarrow x &= \Phi^{-1}\left(1-\frac{\alpha}{2}\right) = u_{1-\frac{\alpha}{2}}\end{aligned}$$

(Lösung mit „+“ entfällt, weil dann $\Phi(x) = \Phi(x) - 1 + \alpha$ folgen würde. Widerspruch zu $\alpha \in (0,1)$)
Mit Einsetzen folgt: Das symmetrische Intervall aus (7.3) hat auch minimale Länge und ist überdies sogar eindeutig bestimmt.

7.5 Beispiel: Normalverteilungsmodell bei unbekannter Varianz

Seien $X_1, \ldots, X_n$ iid mit $X_i \sim N(\mu, \sigma^2)$ mit $\mu \in \mathbb{R}, \sigma^2 > 0$ beide unbekannt. Gesucht ist Konfidenzintervall zum Niveau $1 - \alpha$ für μ. Das Konfidenzintervall aus 7.1 ist von der Gestalt

$$\left[\overline{X}_n - \frac{\sigma}{\sqrt{n}} \cdot x, \overline{X}_n - \frac{\sigma}{\sqrt{n}} \cdot y\right]$$

und damit nicht berechenbar, da σ unbekannt.

Idee: Ersetze σ^2 durch einen Schätzer, zum Beispiel durch den (erwartungstreuen) Schätzer

$$S_n^2 := \frac{1}{n-1} \cdot \sum_{i=1}^{n} (X_i - \overline{X}_n)^2$$

Man erhält (durch Plug-In) das zufällige Intervall

$$\left[\overline{X}_n - \frac{S_n}{\sqrt{n}} \cdot x, \overline{X}_n - \frac{S_n}{\sqrt{n}} \cdot y\right]$$

Zur Wahl von $y < x$ beachte, dass

$$\mathbb{P}_{\mu,\sigma^2}\left[\mu \in \left[\overline{X}_n - \frac{S_n}{\sqrt{n}} \cdot x, \overline{X}_n - \frac{S_n}{\sqrt{n}} \cdot y\right]\right] = \mathbb{P}_{\mu,\sigma^2}\left[y \le \underbrace{\frac{\sqrt{n} \cdot (\overline{X}_n - \mu)}{S_n}}_{=:T_n} \le x\right] \tag{7.4}$$

Ist die Verteilung $Q_n := \mathcal{L}(T_n | \mathbb{P}_{\mu,\sigma^2}) := \mathbb{P}_{\mu,\sigma^2} \circ T_n^{-1}$ unabhängig von (μ, σ^2), so wähle $y < x$ derart, dass $Q_n([y, x]) = 1 - \alpha$. Aus (7.4) folgt dann, dass dieses Intervall ein Konfidenzintervall von μ zum Konfidenzniveau $1 - \alpha$ ist. Deshalb bestimme Q_n.
Beachte zunächst:

$$\begin{aligned}
T_n &= \sqrt{n} \cdot \frac{\frac{1}{n} \cdot \sum_{i=1}^n (X_i - \mu)}{\sqrt{\frac{1}{n-1} \cdot \sum_{i=1}^n \left((X_i - \mu) - \frac{1}{n} \cdot \sum_{j=1}^n (X_j - \mu)\right)^2}} \\
&= \sqrt{n} \cdot \frac{\frac{1}{n} \cdot \sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma}\right)}{\sqrt{\frac{1}{n-1} \cdot \sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma} - \frac{1}{n} \cdot \sum_{j=1}^n \frac{X_j - \mu}{\sigma}\right)^2}} \\
&= g_n(Y_1, \ldots, Y_n)
\end{aligned}$$

mit $Y_i := \frac{X_i - \mu}{\sigma}$ $(1 \le i \le n)$, dann $Y_1, \ldots, Y_n$ iid mit $Y_j \sim N(0,1)$ unter $\mathbb{P}_{\mu,\sigma^2}$. Damit folgt:

$$\begin{aligned}
\mathbb{P}_{\mu,\sigma^2} \circ T_n^{-1} &= \mathbb{P}_{\mu,\sigma^2}(g_n \circ (Y_1, \ldots, Y_n))^{-1} \\
&= (\mathbb{P}_{\mu,\sigma^2} \circ (Y_1, \ldots, Y_n)^{-1}) \circ g_n^{-1} = N(0,1)^n \circ g_n^{-1} = Q_n
\end{aligned}$$

T_n ist sogenannte verteilungsfreie Statistik. Die Verteilung Q_n kann ist durch Monte-Carlo-Simulation approximierbar.
Einschub zur Monte-Carlo-Simulation: Sei $\underline{X}_n^{(N)} = (X_1^{(N)}, \ldots, X_n^{(N)})$. Erzeugte $X_1^{(j)}, \ldots, X_n^{(j)}$ iid mit $X_i^{(j)} \sim N(0,1)$ für $j = 1, \ldots, N$. Dann $\underline{X}_n^{(1)}, \ldots, \underline{X}_n^{(N)} \sim \underline{X} = (X_1, \ldots, X_n)$. Durch Einsetzen erhält man $T_n^{(1)}, \ldots, T_n^{(N)}$ iid mit $T_n^{(i)} \sim Q_n$. Starkes Gesetz der großen Zahlen:

$$\frac{1}{n} \cdot \sum_{i=1}^{N} 1_{y \le T_n^{(i)} \le x} \overset{N \to \infty}{\to} \mathbb{P}[y \le T_n \le x] = Q_n([y, x]) \qquad \text{fast sicher}$$

7.6 Satz

Seien $X_1, \ldots, X_n$ iid mit $X_i \sim N(\mu, \sigma^2)$. Dann gilt:

(i). $\overline{X}_n, S_n^2$ unabhängig

(ii). $\frac{(n-1)\cdot S_n^2}{\sigma^2}$ ist χ^2-verteilt mit $n-1$ Freiheitsgraden.

(iii). Es gilt: $T_n = \sqrt{n} \cdot \frac{\overline{X}_n - \mu}{S_n} \sim t_{n-1}$

Beweis: lineare Modelle; (iii) folgt leicht aus (i),(ii) mit Definition der t_n-Verteilung)

Bemerkung:

(i). Die Verteilungsfunktion F_{t_n} der t_n-Verteilung ist stetig und streng monoton wachsend, da Dichte $f_{t_n} > 0$.

(ii). Definiere $t_{n,\alpha} := F_{t_n}^{-1}(\alpha)$ als α-Quantil von t_n. Quantile sind bekannt (Tafel, Programm-Pakete). Beachte: Wegen t_n symmetrisch um Null gilt $t_{n,1-\alpha} = -t_{n,\alpha}$ für alle $\alpha \in (0,1)$, $n \in \mathbb{N}$.

7.7 Beispiel

Wenn man in (7.4) zum Beispiel $x = t_{n-1,1-\frac{\alpha}{2}}, y = t_{n-1,\frac{\alpha}{2}}$ wählt, folgt

$$\begin{aligned}
&\mathbb{P}_{\mu,\sigma^2}\left[\mu \in \left[\overline{X}_n - \frac{S_n}{\sqrt{n}} t_{n-1,1-\frac{\alpha}{2}}, \overline{X}_n + \frac{S_n}{\sqrt{n}} \cdot t_{n-1,1-\frac{\alpha}{2}}\right]\right] \\
&= \mathbb{P}_{\mu,\sigma^2}\left[t_{n-1,\frac{\alpha}{2}} \le \sqrt{n} \cdot \frac{\overline{X}_n - \mu}{S_n} \le t_{n-1,1-\frac{\alpha}{2}}\right] \\
&= \underbrace{F_{t_{n-1}}\left(t_{n-1,1-\frac{\alpha}{2}}\right)}_{1-\frac{\alpha}{2}} - \underbrace{F_{t_{n-1}}\left(t_{n-1,\frac{\alpha}{2}}\right)}_{\frac{\alpha}{2}} = 1-\alpha
\end{aligned}$$

d.h. $\left[\overline{X}_n - \frac{S_n}{\sqrt{n}} t_{n-1,1-\frac{\alpha}{2}}, \overline{X}_n + \frac{S_n}{\sqrt{n}} \cdot t_{n-1,1-\frac{\alpha}{2}}\right]$ ist Konfidenzintervall für μ zum Niveau $1-\alpha$.

Bemerkung:

(i). Man kann zeigen, dass

$$F_{t_n}(x) \uparrow \Phi(x) \qquad (n \to \infty) \tag{7.5}$$

für alle $x \ge 0$, woraus folgt, dass

$$t_{n,\alpha} \downarrow u_\alpha = \Phi^{-1}(\alpha) \qquad (n \to \infty) \tag{7.6}$$

für alle $\alpha \in \left[\frac{1}{2}, 1\right)$.

7.8 Lemma

Seien $\chi^2_{n_1}, \ldots, \chi^2_{n_m}$ unabhängig. Dann $\sum_{j=1}^m \chi^2_{n_j} \sim \chi^2_n$ mit $n := \sum_{i=1}^m n_j$.

Beweis:

- Seien $N_i^{(j)}$, $1 \le i \le n_j, 1 \le j \le m$ iid $\sim N(0,1)$. Gemäß Definition gilt

$$\chi^2_{n_j} = \sum_{i=1}^{n_j} (N_i^{(j)})^2 \qquad (1 \le j \le m)$$

Daraus folgt:

$$\sum_{j=1}^m \chi^2_{n_j} = \sum_{j=1}^m \sum_{i=1}^{n_j} (N_i^{(j)})^2 \sim \chi^2_n$$

mit $n := \sum_{j=1}^m n_j$.

7.9 Beispiel

Seien $X_1,\dots,X_n$ iid mit $X_i \sim \mathrm{Exp}(\lambda)$, $\lambda > 0$. Gesucht ist Konfidenzintervall für λ zum Niveau $1-\alpha$. Der MLS für λ ist $\hat\lambda_n := \frac{1}{\overline{X}_n}$ (vgl. 3.6, nicht erwartungstreu). Es ist

$$\begin{aligned}\mathbb{P}_\lambda[2\lambda\cdot X_i \le x] &\overset{x\ge 0}{=} \mathbb{P}_\lambda\left[X_i \le \frac{x}{2\lambda}\right] = 1-\exp\left(-\lambda\cdot\frac{x}{2\lambda}\right)\\ &\Rightarrow 2\lambda\cdot X_i \sim \mathrm{Exp}\left(\frac{1}{2}\right) \overset{6.4}{=} \chi_2^2\\ &\Rightarrow \frac{2n\cdot\lambda}{\hat\lambda_n} = 2n\cdot\overline{X}_n\cdot\lambda = \sum_{i=1}^n \underbrace{2\lambda\cdot X_i}_{\sim\chi_2^2} \overset{7.8}{\sim} \chi_{2n}^2\end{aligned}$$

Sei $\chi^2_{m,\beta} := F^{-1}_{\chi^2_m}(\beta)$ das β-Quantil der χ^2_m-Verteilung. Für alle $\lambda > 0$ gilt dann:

$$1-\alpha = \mathbb{P}_\lambda\left[\chi^2_{2n,\frac{\alpha}{2}} \le \frac{2n\cdot\lambda}{\hat\lambda_n} \le \chi^2_{2n,1-\frac{\alpha}{2}}\right] = \mathbb{P}_\lambda\left[\lambda\in\left[\hat\lambda_n\cdot\frac{\chi^2_{2n,\frac{\alpha}{2}}}{2n}, \hat\lambda_n\cdot\frac{\chi^2_{2n,1-\frac{\alpha}{2}}}{2n}\right]\right]$$

Also ist $\left[\hat\lambda_n\cdot\frac{\chi^2_{2n,\frac{\alpha}{2}}}{2n}, \hat\lambda_n\cdot\frac{\chi^2_{2n,1-\frac{\alpha}{2}}}{2n}\right]$ ein Konfidenzintervall für λ zum Niveau $1-\alpha$.

Allgemeines Prinzip in den Beispielen 7.1, 7.5, 7.7, 7.9 erkennbar: Transformiere Schätzer (zum Beispiel „gute“ Schätzer) unter Einbeziehung des Parameteranteils $\gamma(\theta)$ derart, dass die Verteilung der Transformierten nicht von θ abhängt. Formal:
Sei $\hat\Theta : \mathcal{X} \to \mathbb{R}$ ein Schätzer für $\gamma(\theta)\in\mathbb{R}$, $\theta\in\Theta\subseteq\mathbb{R}^d$. Finde Transformation (messbare Abbildung) $H : \mathbb{R}\times\mathbb{R}\times\Omega\to\mathbb{R}$ so, dass

$$\mathcal{L}(H(\gamma(\theta),\hat\Theta(X),\cdot)|\mathbb{P}_\theta) = \mathbb{P}_\theta\circ H(\gamma(\theta),\hat\Theta(X),\cdot)^{-1}$$

nicht von θ abhängt (und idealerweise bekannt ist, sonst beliebig genaue Approximation durch Monte-Carlo-Simulation).
Bisherige Beispiele:

(i). In Beispiel 7.1: $\theta = \mu\in\Theta = \mathbb{R}$, $\hat\Theta_n(\underline{x}_n) = \overline{x}_n$, $\mathcal{X} = \mathbb{R}^n$, $\gamma(\theta) = \theta$,

$$H(\gamma(\theta),\hat\Theta_n(\underline{X}_n)) = H_\sigma(\mu,\overline{X}_n) = \frac{\overline{X}_n-\mu}{\sigma} \sim N(0,1)$$

(ii). In Beispiel 7.5/7.7: $\theta = (\mu,\sigma^2)\in\Theta = \mathbb{R}\times(0,\infty)$, $\gamma((\mu,\sigma^2)) = \mu$, $\hat\Theta_n(\underline{x}_n) = \overline{x}_n, \mathcal{X} = \mathbb{R}^n$,

$$H(\gamma(\theta),\hat\Theta_n(\underline{X}_n),\cdot) = H(\mu,\overline{X}_n,\cdot) = \frac{\overline{X}_n-\mu}{S_n} = H_{S_n}(\mu,\overline{X}_n) \sim t_{n-1}$$

(iii). In Beispiel 7.9: $\theta = \lambda\in\Theta = (0,\infty)$, $\gamma = \mathrm{id}$, $\hat\Theta_n(\underline{x}_n) = \frac{1}{\overline{x}_n} =: \hat\lambda_n$,

$$H(\theta,\hat\Theta_n(\underline{X}_n)) = H(\lambda,\hat\lambda_n) = \frac{2n\cdot\lambda}{\hat\lambda_n} \sim \chi^2_{2n}$$

Es sei im allgemeinen Fall F die Verteilungsfunktion von $H(\gamma(\theta),\hat\Theta(X),\cdot)$, d.h.

$$F(x) = \mathbb{P}_\theta[H(\gamma(\theta),\hat\Theta(X),\cdot)\le x] \qquad (x\in\mathbb{R})$$

und bezeichne F^{-1} die zugehörige Quantilfunktion. Damit:

$$\begin{aligned}&\mathbb{P}_\theta\left[F^{-1}\left(\frac{\alpha}{2}\right) \le H(\gamma(\theta),\hat\Theta(X),\cdot) \le F^{-1}\left(1-\frac{\alpha}{2}\right)\right]\\ &= \underbrace{F\left(F^{-1}\left(1-\frac{\alpha}{2}\right)\right)}_{\ge 1-\frac{\alpha}{2}} - \underbrace{F\left(F^{-1}\left(\frac{\alpha}{2}-\right)\right)}_{\le\frac{\alpha}{2}} \ge 1-\alpha\end{aligned}$$

für alle $\theta \in \Theta$. Falls F stetig gilt sogar „=“. Kann man die Ungleichungen

$$F^{-1}\left(\frac{\alpha}{2}\right) \le H(\gamma(\theta), \hat{\Theta}(X), \cdot) \le F^{-1}\left(1 - \frac{\alpha}{2}\right) \tag{*}$$

nach $\gamma(\theta)$ auflösen, so erhält man im Allgemeinen einen Konfidenzbereich für $\gamma(\theta)$ zum Niveau $1-\alpha$. Für die praktische Umsetzung notwendig:

- Die Quantile $F^{-1}\left(\frac{\alpha}{2}\right), F^{-1}\left(1-\frac{\alpha}{2}\right)$ müssen bekannt sein. Kein Problem, falls F bekannt, sonst Approximation durch Monte-Carlo-Simulation.
- $(*)$ nach $\gamma(\theta)$ auflösbar. Sonst ist der Konfidenzbereich für $\gamma(\theta)$ nur implizit durch $(*)$ gegeben. Lässt aber bei konkreter Beobachtung $X(w) = x \in \mathcal{X}$ mitunter numerische oder graphische Verfahren zu:

  Fixiere w und betrachte $u \mapsto H(u, \hat{\Theta}(X(w)), w)$, $u \in \mathbb{R}$. Gesucht sind $u \in \mathbb{R}$ mit

  $$q_{\frac{\alpha}{2}} \le H(u, \hat{\Theta}(X(w)), w) \le q_{1-\frac{\alpha}{2}}$$

Die Transformation $H : \mathbb{R} \times \mathbb{R} \times \Omega \to \mathbb{R}$ heißt Pivot (Drehpunkt) und obiges Prinzip heißt Pivotisieren der Statistik $\hat{\Theta}$.
Typischerweise betrachten wir Produktmodelle, d.h. $X = \underline{X}_n = (X_1, \ldots, X_n)$ mit $X_i \sim P_\theta$ iid (vgl. 1.6). Hierbei ist das exakte Pivotisieren bei festem n oft problematisch und -bis auf klassische Modelle- eher die Ausnahme. Ausweg für großen Stichprobenumfang n: Approximatives Pivotisieren.

Approximatives Pivotisieren Sei $\hat{\Theta}_n : \mathcal{X}^n \to \mathbb{R}$ ein Schätzer für $\gamma(\theta)$. Die Abbildung $H : \mathbb{R} \times \mathbb{R} \times \Omega \to \mathbb{R}$ heißt approximatives Pivot, falls

$$\mathbb{P}_\theta \circ H_n(\gamma(\theta), \hat{\Theta}_n(\underline{X}_n), \cdot)^{-1} \xrightarrow{w} Q \qquad (n \to \infty) \tag{7.7}$$

$$\Leftrightarrow Z_n := H_n(\gamma(\theta), \hat{\Theta}_n(\underline{X}_n), \cdot) \xrightarrow{\text{i.V}} Z \qquad \text{unter } \mathbb{P}_\theta$$

$$\Leftrightarrow \mathbb{P}_\theta[Z_n \le x] \to \mathbb{P}_\theta[Z \le x] =: F(x) \tag{7.8}$$

wobei $\mathbb{P}_\theta \circ Z^{-1}$ nicht von θ abhängt und x ein Punkt, in dem F stetig ist.
Falls F stetig ist, so gilt:

$$\begin{aligned}&\mathbb{P}_\theta\left[F^{-1}\left(\frac{\alpha}{2}\right) < H_n(\gamma(\theta), \hat{\Theta}_n(\underline{X}_n), \cdot) \le F^{-1}\left(1-\frac{\alpha}{2}\right)\right]\\ &= \mathbb{P}_\theta[Z_n \le q_{1-\frac{\alpha}{2}}] - \mathbb{P}_\theta[Z_n \le q_{\frac{\alpha}{2}}] \to F(q_{1-\frac{\alpha}{2}}) - F(q_{\frac{\alpha}{2}}) = 1-\alpha \qquad (n \to \infty)\end{aligned}$$

In diesem Sinne ist die Menge

$$B_n := \left\{u \in \mathbb{R}; F^{-1}\left(\frac{\alpha}{2}\right) \le H_n(u, \hat{\Theta}(\underline{X}_n), \cdot) \le F^{-1}\left(1-\frac{\alpha}{2}\right)\right\}$$

ein asymptotischer Konfidenzbereich für $\gamma(\theta)$ zum Niveau $1-\alpha$, d.h.

$$\lim_{n\to\infty} \mathbb{P}_\theta\left(\gamma(\theta) \in B_n\right) = 1-\alpha$$

Rest wie beim exakten Pivotisieren.

7.10 Beispiel: n-faches Bernoulli-Experiment

Seien $X_1, \ldots, X_n$ iid mit $X_i \sim \beta_p$, $p \in (0,1)$. $\overline{X}_n$ ist optimaler Schätzer für p. Zentraler Grenzwertsatz liefert

$$\sqrt{n} \cdot \frac{(\overline{X}_n - p)}{\sqrt{p \cdot (1-p)}} \xrightarrow{\text{i.V.}} N(0,1) \qquad (n \to \infty)$$

unter $\mathbb{P}_p$ für alle $p \in (0,1)$. Damit:

$$\mathbb{P}_p\left[u_{\frac{\alpha}{2}} < \sqrt{n}\cdot\frac{(\overline{X}_n - p)}{\sqrt{p\cdot(1-p)}} \le u_{1-\frac{\alpha}{2}}\right] \to \Phi(u_{1-\frac{\alpha}{2}}) - \Phi(u_{\frac{\alpha}{2}}) = 1-\alpha$$

für alle $p \in (0,1)$. Auflösen nach p mittels Lemma 7.11. Mit $s := \frac{u_{1-\frac{\alpha}{2}}}{\sqrt{n}}, x = \overline{X}_n$ folgt

$$\left[T_-\left(\overline{X}_n, \frac{u^2_{1-\frac{\alpha}{2}}}{n}\right), T_+\left(\overline{X}_n, \frac{u^2_{1-\frac{\alpha}{2}}}{n}\right)\right]$$

ist ein asymptotisches Konfidenzintervall für p zum Niveau $1-\alpha$.

7.11 Lemma

Für $p \in (0,1), s > 0$, $0 \le x \le 1$ sind äquivalent:

(i). $-\sqrt{s} \le \frac{x-p}{\sqrt{p\cdot(1-p)}} \le \sqrt{s}$

(ii). $T_-(x,s) \le p \le T_+(x,s)$ wobei

$$T_\pm(x,s) = \frac{x + \frac{s}{2} \pm \sqrt{\frac{s^2}{4} + s\cdot x\cdot(1-x)}}{1+s}$$

Beweis: Nachrechnen.

Die Vorgehensweise aus Beispiel 7.10 ist übertragbar auf sogenannte nicht-parametrische Verteilungsannahmen, zum Beispiel

$$\mathcal{P} = \left\{P; P \text{ W-Verteilung}, \int_{\mathbb{R}} x^2\, P(dx) < \infty\right\}$$

Ziel: Schätzung des Parameteranteils $\gamma(P) := \int x\, P(dx) = \mathbb{E}X$ für $X \sim P$.

7.12 Beispiel

Seien $X_1, \dots, X_n$ iid mit $X_i \sim P \in \mathcal{P}$, d.h. $\mathbb{E}(X_i^2) < \infty$, $\gamma(P) = \mathbb{E}X_i$. Annahme:

$$\sigma^2 := \sigma^2(P) = \mathbb{V}(X_i) = \int x^2\, P(dx) - \left(\int x\, P(dx)\right)^2$$

Gemäß zentralem Grenzwertsatz gilt:

$$\sqrt{n}\cdot\frac{\overline{X}_n - \gamma(P)}{\sqrt{\sigma^2}} \overset{\text{i.V}}{\to} N(0,1)$$

$$\Rightarrow \mathbb{P}_P\left[\gamma(P) \in \left[\overline{X}_n - \frac{u_{1-\frac{\alpha}{2}}}{\sqrt{n}}\cdot\sigma, \overline{X}_n + \frac{u_{1-\frac{\alpha}{2}}}{\sqrt{n}}\cdot\sigma\right]\right] \to (1-\alpha) \qquad (n\to\infty)$$

also ist dies ein asymptotisches Konfidenzintervall bei bekanntem σ^2.
Falls $\sigma = \sigma^2(P)$ unbekannt, ersetze es durch einen Schätzer. $\sigma^2 : \mathcal{P} \to \mathbb{R}$ ist ein statistisches Funktional. Setze

$$P_n := \frac{1}{n}\cdot\sum_{i=1}^n \delta_{X_i}$$

(empirische Verteilung). Dann

$$\begin{aligned} P_n(F) &= \frac{1}{n}\cdot\sum_{i=1}^n \delta_{X_i}(F) = \frac{1}{n}\sum_{i=1}^n 1_{[X_i\in F]} \\ &\to \mathbb{P}_P[X_1 \in F] = P(F) \qquad (F \in \mathcal{B}(\mathbb{R})) \end{aligned}$$

$P_n \to P$ fast sicher für alle $F \in \mathcal{B}(\mathbb{R})$. Plug-In-Methode: $\sigma_n^2 := \sigma^2(P_n)$. Dann:

$$\begin{aligned}\sigma_n^2 &= \int x^2\,P_n(dx) - \left(\int x\,P_n(dx)\right)^2\\ &= \frac{1}{n}\cdot\sum_{i=1}^n X_i^2 - \left(\frac{1}{n}\cdot\sum_{i=1}^n X_i\right)^2\\ &= \frac{1}{n}\cdot\sum_{i=1}^n X_i^2 - \overline{X}_n^2 = \frac{1}{n}\cdot\sum_{i=1}^n (X_i - \overline{X}_n)^2\end{aligned}$$

σ_n^2 ist jedoch kein erwartungstreuer Schätzer. Aber

$$S_n^2 := \frac{1}{n-1}\cdot\sum_{i=1}^n (X_i - \overline{X}_n)^2 = \frac{n}{n-1}\cdot\sigma_n^2$$

ist erwartungstreuer Schätzer. Einsetz-Methode liefert:

$$\sqrt{n}\cdot\frac{\overline{X}_n - \gamma(P)}{S_n} = \underbrace{\frac{\sigma}{S_n}}_{=:Z_n}\cdot\underbrace{\frac{\sqrt{n}\cdot(\overline{X}_n - \gamma(p))}{\sigma}}_{=:V_n \overset{\text{i.V.}}{\to} N(0,1)}$$

Außerdem:

$$\begin{aligned}\sigma_n^2 &= \underbrace{\frac{1}{n}\cdot\sum_{i=1}^n X_i^2}_{\to\mathbb{E}(X_1^2)\text{ f.s}} - \Bigg(\underbrace{\frac{1}{n}\cdot\sum_{i=1}^n X_i}_{\to\mathbb{E}X_1\text{ f.s}}\Bigg)^2 \to \mathbb{E}X_1^2 - (\mathbb{E}X_1^2) = \mathbb{V}(X_1) = \sigma^2 \qquad \text{f.s}\\ \Rightarrow S_n^2 &= \frac{n}{n-1}\cdot\sigma_n^2 \to \sigma^2 \qquad \text{f.s}\end{aligned}$$

Daraus folgt insbesondere $S_n^2 \overset{P}{\to} \sigma^2$. Aus Stetigkeit erhält man $Z_n \overset{P}{\to} 1$ für $n \to \infty$ und somit $Z_n \cdot V_n \overset{\text{i.V}}{\to} N(0,1)$. Auflösen liefert, dass

$$\left[\overline{X}_n - \frac{S_n}{\sqrt{n}}\cdot u_{1-\frac{\alpha}{2}}, \overline{X}_n + \frac{S_n}{\sqrt{n}}\cdot u_{1-\frac{\alpha}{2}}\right]$$

ein asymptotisches Konfidenzintervall für $\gamma(P)$ zum Niveau $1-\alpha$ bei unbekanntem $\sigma^2 > 0$ ist (Vergleich mit Konfidenzintervall aus Beispiel 7.7: $t_{n-1,1-\frac{\alpha}{2}}$ anstelle von $u_{1-\frac{\alpha}{2}}$ hier).

7.13 Definition

Seien $(\mathcal{X},\mathcal{F})$ ein Stichprobenraum, $(X_i)_{i\in\mathbb{N}}$ iid Zufallsvariablen mit Werten in $\mathcal{X}$ (über einem Wahrscheinlichkeitsraum $(\Omega,\mathcal{A},\mathbb{P})$) mit $\mathbb{P}\circ X_i^{-1} = P \in \mathcal{P}$ und $\gamma : \mathcal{P} \to \mathbb{R}$ eine beliebige Funktion (statistisches Funktional). Sei $C_n : \mathcal{X}^n \to \{I; I \subseteq \mathbb{R}\}$ mit

$$\forall P \in \mathcal{P} : \{\underline{x}_n \in \mathcal{X}^n; \gamma(P) \in C_n(\underline{x}_n)\} \in \mathcal{F}^n$$

Dann heißt $(C_n)_{n\in\mathbb{N}}$ eine Folge von Konfidenzbereichen für $\gamma(P)$ zum Niveau $1-\alpha$, falls

$$\forall P \in \mathcal{P} : \lim_{n\to\infty} \mathbb{P}_P\left(\gamma(P) \in C_n(\underline{X}_n)\right) = 1 - \alpha$$

Bemerkung: Asymptotische Verfahren treten typischerweise in der nicht-parametrischen Statistik auf.

8 Statistische Tests

Erinnerung an Beispiel 1.2

- Qualitätskontrolle bei der Produktion einer bestimmten Ware: $p \in (0,1)$ ist der unbekannte tatsächlicher Ausschußanteil, $p_0 \in (0,1)$ sei zulässiger Ausschußanteil. Zu entscheiden ist zwischen den Aussagen $H_0 : p > p_0$ und $H_1 : p \leq p_0$. Dazu prüfe n Stücke und erhalte $\underline{x}_n = (x_1, \ldots, x_n) \in \{0,1\}^n$ mit $x_i = 1$, falls i-tes Teil schlecht und $x_i = 0$, falls i-tes Teil gut.

- Statisches Modell: $\underline{x}_n$ ist Realisierung von $\underline{X}_n = (X_1, \ldots, X_n)$ mit X_i iid $X_i \sim \beta_p$, $p \in (0,1)$ unbekannt.

- Ziel: Entscheide zwischen $H_0 : p > p_0$ und $H_1 : p \leq p_0$. Notwendig: Entscheidungsvorschrift, die jedem $\underline{x}_n \in \{0,1\}^n$ genau eine von diesen zwei Entscheidungen zuordnet. Formal ist die Entscheidungsvorschrift eine Abbildung $t_n : \mathcal{X}^n \to \{0,1\}$ mit der Interpretation $t_n(\underline{x}_n) = 1 \longleftrightarrow$ Entscheidung gegen H_0 und $t_n(\underline{x}_n) = 0 \longleftrightarrow$ Entscheidung für H_0. Sei $K_n := \{\underline{x}_n \in \mathcal{X}^n; t_n(\underline{x}_n) = 1\} = t_n^{-1}(\{1\})$. Damit

$$t_n(\underline{x}_n) = 1 \Leftrightarrow \underline{x}_n \in K_n \longleftrightarrow \text{Entscheidung gegen } H_0$$
$$t_n(\underline{x}_n) = 0 \Leftrightarrow \underline{x}_n \notin K_n \longleftrightarrow \text{Entscheidung für } H_0$$

  K_n heißt kritischer Bereich.

- Welche t_n bzw. K_n sind sinnvoll? Klar: Alle $\underline{x}_n \in \{0,1\}^n$ sind unter jedem $p \in (0,1)$ möglich, denn

$$\mathbb{P}_p[\underline{X}_n = \underline{x}_n] = \mathbb{P}\left(\bigcap_{i=1}^n [X_i = x_i]\right) = \prod_{i=1}^n \underbrace{\mathbb{P}_p[X_i = x_i]}_{p^{x_i}\cdot(1-p)^{1-x_i}}$$
$$= p^{\sum_{i=1}^n x_i} \cdot (1-p)^{n-\sum_{i=1}^n x_i} > 0 \qquad (p \in (0,1), \underline{x}_n \in \mathcal{X}^n)$$

  Deshalb:

  (i). Fehlentscheidungen sind bei jedem K_n unvermeidbar.

  (ii). Ziel: Mögliche Wahrscheinlichkeiten für Fehlentscheidungen minimieren.

- Mögliche Fehlentscheidungen:

  (i). H_0 gilt und $\underline{x}_n \in K_n$ wird beobachtet (Fehler 1.Art).

  (ii). H_1 gilt und $\underline{x}_n \notin K_n$ wird beobachtet (Fehler 2.Art).

  Die Wahrscheinlichkeiten für den Fehler 1. Art sind $\mathbb{P}_p([\underline{X}_n \in K_n] | p > p_0)$ und für den 2. Art $\mathbb{P}_p([\underline{X}_n \notin K_n] | p \leq p_0)$ und damit insbesondere von p abhängig.

- Idealerweise sollte K_n so beschaffen sein, dass

$$\underbrace{\mathbb{P}_p[\underline{X}_n \in K_n]}_{1-\mathbb{P}_p[\underline{X}_n \notin K_n]} = \begin{cases} 0 & p > p_0 \\ 1 & p \leq p_0 \end{cases}$$

Dann wäre in praktisch jedem Fall jeweils die Wahrscheinlichkeit für eine Fehlentscheidung Null: Der Statistiker träfe mit der Vorschrift $t_n = 1_{K_n}$. Jedoch ist dies bereits in unserem einfachen Beispiel nicht möglich, denn die Funktion

$$p \mapsto \mathbb{P}_p[\underline{X}_n \in K_n] = \sum_{\underline{x}_n \in K_n} \underbrace{\mathbb{P}_p[\underline{X}_n = \underline{x}_n]}_{=p^{\sum_{i=1}^n x_i} \cdot (1-p)^{n-\sum_{i=1}^n x_i}}$$

ist als Polynom in p stetig auf $(0,1)$.

- Typisches Dilemma: Es gibt große Wahrscheinlichkeiten für den Fehler 1.Art und 2.Art (vor allem in der Nähe von p_0). Die beiden Fehlerarten haben aber je nach Problemstellung unterschiedliche Konsequenzen, die im Einzelfall hinsichtlich ihrer Tragweite bewertet werden müssen. (Dies ist ein „außermathematischer“ Vorgang.)
- Im Beispiel 1.2:

  (i). Fehler 1. Art heißt falsche Entscheidung gegen H_0 ($p > p_0$), d.h. Entscheidung für $p \leq p_0$ („alles okay“). Konsequenz: unzufriedene Kunden.

  (ii). Fehler 2. Art bedeutet falsche Entscheidung gegen $p \leq p_0$, d.h. Entscheidung für $p > p_0$ („schlechte Produktion„). Konsequenz: Eingreifen in den Produktionsprozess, z.B. Stoppen

  Deshalb folgendes Prinzip: Die Wahrscheinlichkeiten desjenigen Fehlers, der als „schlimmer“ erachtet wird, werden -falls möglich- kontrolliert unterhalb einer vorgegebenen Schranke $\alpha \in (0,1)$ gehalten. Genauer:

  (i). Auszeichnung einer Aussage als sogenannte Nullhypothese H_0 und der zweiten Aussage als sogenannte Alternative H_1. Im Beispiel etwa $p > p_0$ (Nullhypothese) und $H_1 : p \leq p_0$ (Alternative). Sei $\Theta_0 := (p_0, 1) = \{p \in (0,1); H_0 \text{ gilt }\}$ und $\Theta_1 = (0, p_0] = \{p \in (0,1); H_1 \text{ gilt }\}$. Damit $\Theta = \Theta_0 \oplus \Theta_1$.

  (ii). Kontrolle der Wahrscheinlichkeiten für Fehler 1.Art durch vorgegebene Schranke $\alpha \in (0,1)$, d.h. Wahl von K_n so, dass

  $$\forall p > p_0 : \mathbb{P}_p[\underline{X}_n \in K_n] \leq \alpha \tag{*}$$

  (iii). Zusätzlich zu $(*)$ in (ii): Minimierung der Wahrscheinlichkeiten für den Fehler 2. Art, d.h. K_n mit $(*)$ und

  $$\forall p \leq p_0 : \mathbb{P}_p[\underline{X}_n \notin K_n] = 1 - \mathbb{P}_p[\underline{X}_n \in K_n]$$

  möglichst klein.

  Frage: Ist das Programm (i)-(iii) machbar? Antwort in allgemeinem Rahmen...

8.1 Allgemeine Formulierung eines statistischen Testproblems

Gegeben seien

(i). statistisches Modell $(X, \mathcal{X}, \mathcal{F}, \mathcal{P})$ mit Verteilungsannahme $\mathcal{P} = \{P_\theta; \theta \in \Theta \subseteq \mathbb{R}^d\}$

(ii). $\Theta = \Theta_0 \oplus \Theta_1$ definiert das Testproblem

$$H_0 : \theta \in \Theta_0 \qquad \text{Nullhypothese vs} \qquad H_1 : \theta \in \Theta_1 \qquad \text{Alternative}$$

(iii). $\alpha \in (0,1)$ heißt Signifikanzniveau.

Jedes $K \in \mathcal{F}$ mit der Eigenschaft

$$\forall \theta \in \Theta_0 : \mathbb{P}_\theta[X \in K] \leq \alpha \tag{8.1}$$

definiert dann einen Test zum Niveau α (Niveau-α-Test) für H_0 gegen H_1 vermöge der Entscheidungsregel

$$\begin{aligned} \text{Entscheidung gegen}\, H_0 &:\Leftrightarrow x \in K \\ \text{Entscheidung gegen}\, H_1 &:\Leftrightarrow x \notin K \end{aligned}$$

für $x \in \mathcal{X}$.

Gesucht (falls existent): $K \in \mathcal{F}$ mit (8.1) und

$$\forall K' \in \mathcal{F}\,\text{mit (8.1)}\colon \forall \theta \in \Theta_1\colon \mathbb{P}_\theta[X \in K'] \leq \mathbb{P}_\theta[X \in K] \tag{8.2}$$

Sprechweisen:

- Entscheidung gegen H_0, obwohl H_0 gilt = Fehler 1.Art
- Entscheidung gegen H_1, obwohl H_1 gilt = Fehler 2. Art
- K heißt kritischer Bereich.

Interpretation:

(i). Angenommen $\theta \in \Theta_0$, d.h. H_0 gilt. Eine Beobachtung $x \in K$, d.h. H_0 wird verworfen (Fehler 1.Art), hat wegen (8.1) die Wahrscheinlichkeit von höchstens α, d.h. bei höchstens $\alpha \cdot 100\%$ aller Anwendungen des Tests wird H_0 fälschlicherweise verworfen.

(ii). Angenommen $\theta \in \Theta_1$, d.h. H_1 gilt. Beobachtung $[X \notin K]$, d.h. H_1 wird verworfen (Fehler 2.Art), tritt mit Wahrscheinlichkeit $1 - \mathbb{P}_\theta[X \in K]$ ein. Diese wird durch die Forderung (8.2) möglichst klein, ist aber nicht kontrollierbar und kann nahe bei Eins liegen. Typischerweise falls θ „in der Nähe“ von θ_0 liegt.

Schlussfolgerung: Nur das Verwerfen von H_0 ist eine statistisch gesicherte Entscheidung (in dem Sinne, dass die Wahrscheinlichkeit für ein fälschliches Verwerfen $\leq \alpha$ ist), nicht aber das Beibehalten von H_0 (da eben die Wahrscheinlichkeit für das fälschliche Beibehalten von H_0 mitunter sehr groß sein kann).

Sprechweisen:

- $x \in K$ spricht signifikant (auf dem Niveau α) gegen H_0.
- Die Alternative H_1 ist zum Niveau α signifikant.

Übersicht (unbekannte Realität-Testentscheidung):

	H_0 gilt	H_1 gilt
H_0 verwerfen	Fehler 1. Art	richtige Entscheidung
H_0 beibehalten	richtige Entscheidung	Fehler 2.Art

Fazit: Statistische Tests dienen zur Widerlegung von Nullhypothesen. Soll eine Aussage bestätigt werden, muss sie als Alternative eines Testproblems formuliert werden. Konsequenz für die Anwendungen:

(i). Wahl von H_0 und H_1 ist wesentlich. Im Beispiel: Soll „$p \leq p_0$“ statistisch signifikant belegt werden, so ist zu betrachten $H_0 : p > p_0$ vs. $H_1 : p \leq p_0$. Soll dagegen eine Qualitätsverschlechterung im Produktionsprozess („$p \geq p_0$“) aufgedeckt werden, so ist zu betrachten $H_0 : p < p_0$ vs. $H_1 : p \geq p_0$.

(ii). Beachte: K ist deterministisch und muss deshalb vor Durchführung des Zufallsexperiments festgelegt werden, ansonsten gilt zum Beispiel (8.1) nicht.

Eine sinnvolle Konstruktion von Tests erfolgt häufig durch heuristische Vorüberlegungen. Danach Güteuntersuchung durch Berechnung der Gütefunktion $\theta \mapsto \mathbb{P}_\theta[X \in K]$ $(\theta \in \Theta)$ möglich.

8.2 Beispiel

Seien $X_1, \dots, X_n$ iid mit $X_j \sim \beta_p$, $p \in (0,1)$ unbekannt. Sei $p_0 \in (0,1)$. Testproblem $H_0 : p \le p_0$ vs. $H_1 : p > p_0$. Idee: $\overline{X}_n \to p$ $(n \to \infty)$ $\mathbb{P}$-fast sicher (SLLN). Zumindest für „große " n gilt daher $\overline{X}_n \approx p$. Große Werte von $\overline{X}_n$ sprechen somit für ein Vorliegen von H_1. Also plausibel:

$$H_0 \text{ verwerfen} :\Leftrightarrow \overline{X}_n > c_\alpha \Leftrightarrow S_n := \sum_{i=1}^n X_i > k_\alpha$$

wobei wegen (8.1) k_α derart bestimmt werden muss, dass

$$\forall p \le p_0 : \mathbb{P}_p[S_n > k_\alpha] \le \alpha \tag{8.3}$$

und möglichst klein wegen (8.2). Bekanntlich gilt $\mathbb{P}_p \circ S_n^{-1} = \operatorname{Bin}(n,p)$ für alle $p \in (0,1), n \in \mathbb{N}$. Damit für $k_\alpha \in \mathbb{N}_0$:

$$\mathbb{P}_p[S_n > k_\alpha] = \sum_{k=k_\alpha+1}^n \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} \qquad (p \in (0,1))$$

Diese Gütefunktion ist streng monoton wachsend in p (Übung). Damit geeignet:

$$k_\alpha = \min_{0 \le j \le n} \left\{ \sum_{k=j+1}^n \binom{n}{k} \cdot p_0^k \cdot (1-p_0)^{n-k} \le \alpha \right\} \tag{8.4}$$

$$\Rightarrow K := K_{n,\alpha} = \left\{ \underline{x}_n \in \{0,1\}^n ; \sum_{i=1}^n x_i > k_\alpha \right\}$$

Mit k_α aus (8.4) ist $K_{n,\alpha}$ kritischer Bereich eines Niveau-α-Tests für H_0 gegen H_1.

8.3 Beispiel

$X_1, \dots, X_n$ iid mit $X_j \sim N(\mu, \sigma_0^2)$ mit $\mu \in \mathbb{R}$ unbekannt, $\sigma_0^2 > 0$ bekannt. Testproblem: $H_0 : \mu \le \mu_0$ vs. $H_1 : \mu > \mu_0$ mit μ_0 vorgegeben. Plausibler Test (mit gleicher Motivation wie in 8.2):

$$H_0 \text{ verwerfen} :\Leftrightarrow \overline{X}_n > c_\alpha$$

wobei c_α möglichst klein mit

$$\forall \mu \le \mu_0 : \mathbb{P}_\mu[\overline{X}_n > c_\alpha] \le \alpha \tag{8.5}$$

Da für $\mu \in \mathbb{R}$

$$\mathbb{P}_\mu[\overline{X}_n > c_\alpha] = \mathbb{P}_\mu\left[\underbrace{\sqrt{n} \cdot \frac{\overline{X}_n - \mu}{\sqrt{\sigma_0^2}}}_{\sim N(0,1)} > \frac{c_\alpha - \mu}{\sqrt{\sigma_0^2}} \cdot \sqrt{n} \right] = 1 - \Phi\left(\frac{c_\alpha - \mu}{\sqrt{\sigma_0^2}} \cdot \sqrt{n} \right)$$

und diese Funktion streng monoton wachsend in μ ist, folgt

$$(8.5) \Leftrightarrow 1 - \Phi\left(\frac{c_\alpha - \mu_0}{\sqrt{\sigma_0^2}} \cdot \sqrt{n} \right) \le \alpha \Leftrightarrow c_\alpha \ge \mu_0 + \frac{\sigma_0}{\sqrt{n}} \cdot u_{1-\alpha}$$

Minimales c_α mit (8.5) ist gegeben durch $c_\alpha = \mu_0 + \frac{\sigma_0}{\sqrt{n}} \cdot u_{1-\alpha}$. Damit ist

$$K = K_{n,\alpha} = \left\{ \underline{x}_n \in \mathbb{R}^n ; \frac{1}{n} \cdot \sum_{i=1}^n x_i > c_\alpha \right\}$$

kritischer Bereich eines Niveau-α-Tests für H_0 vs. H_1. Der Test heißt (einseitiger) Gauß-Test.

Tests in 8.2 und 8.3 von der Gestalt

$$K = \{x \in \mathcal{X}; T(x) > c_\alpha\}$$

für eine bestimmte messbare Abbildung $T : \mathcal{X} \to \mathbb{R}$. T heißt (Test-)Statistik und c_α kritischer Wert.

Nächstes Ziel: Konstruktion optimaler Tests. Dazu:

8.4 Definition und Satz

(i). Sei $(X, \mathcal{X}, \mathcal{F}, \mathcal{P})$ parametrisches statistisches Modell über einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P}_\theta)$ mit $\theta \in \Theta \subseteq \mathbb{R}^d$. Jede messbare Abbildung $\varphi : \mathcal{X} \to [0,1]$ heißt (randomisierter) Test.

(ii). Es existiert eine Zufallsvariable $Y \colon (\Omega, \mathcal{A}, \mathbb{P}_\theta) \to (\{0,1\}; \mathcal{P}(\{0,1\}))$ mit $\mathbb{P}_\theta[Y \in \cdot | X = x] = \mathrm{Bin}(1, \varphi(x))$, d.h.

$$\mathbb{P}_\theta[Y = 1 | X = x] = \varphi(x) \qquad \mathbb{P}_\theta[Y = 0 | X = x] = 1 - \varphi(x) \qquad (*)$$

(iii). Das Paar (φ, Y) legt Entscheidungsvorschrift fest durch:

$$H_0 \text{ verwerfen} :\Leftrightarrow Y = 1 \qquad (+)$$

Wird $[X = x]$ beobachtet, so führt man anschließend ein Bernoulli-Experiment mit Trefferwahrscheinlichkeit $p := \varphi(x)$ durch (gekoppeltes Experiment). Das zufällige Ergebnis dieses Bernoulli-Experiments (Hilfsexperiment) wird beschrieben durch die Zufallsvariable Y. Wird $Y = 1$ beobachtet, dann ist H_0 zu verwerfen, sonst H_0 beibehalten.

Beachte: Wegen $(*), (+)$ gilt

$$\varphi(x) = \mathbb{P}_\theta[H_0 \text{ verworfen durch } \varphi | X = x] \qquad (x \in \mathcal{X}) \tag{8.6}$$

Beachte: Linke Seite in (8.6) hängt nicht von θ ab.

8.5 Beispiel

Sei $p_0 = 0.02$, $n = 30$, $\alpha = 0.01$. Aus (8.4) folgt wegen Monotonie

$$k_{0.01} = \min\left\{0 \le k \le 30; \underbrace{\sum_{j=k+1}^{30} \binom{30}{k} \cdot 0.02^j \cdot 0.98^{30-j}}_{=:p(k)=\mathbb{P}_p[S_{30}>k]} \le \alpha\right\}$$

Wegen

k	0	1	2	3
p(k)	0.4545	0.0105	0.0214	0.0029

folgt hier $k_{0.01} = 3$ mit

$$\sup_{p \le 0.02} \mathbb{P}[S_{30} > 3] = \mathbb{P}_{0,02}[S_{30} > 3] = 0.0029 < 0.01$$

Signifikanz-Niveau $\alpha = 0.01$ wird also nicht voll ausgeschöpft (hier nur zu ca. 30%). Dadurch wird Fehler 2.Art unnötig groß. Ausschöpfen des Niveaus durch Randomisierung möglich. Betrachte zum Beispiel

$$\tilde{\varphi}(\underline{x}_{30}) = \begin{cases} 1 & S_{30} \ge 4 \\ 0 & S_{30} \le 2 \\ \gamma & S_{30} = 3 \end{cases} \tag{8.7}$$

Gesucht γ, um Signifikanz-Niveau möglichst gut auszuschöpfen.

$$\begin{aligned}\mathbb{P}_p[H_0 \text{ wird verworfen}] &= \sum_{\underline{x}_{30}\in\{0,1\}^{30}} \mathbb{P}_p[\tilde{\varphi} \text{ verwirft } H_0 | \underline{X}_{30} = \underline{x}_{30}] \cdot \mathbb{P}_p[\underline{X}_{30} = \underline{x}_{30}] \\ &= \sum_{j=0}^{30} \sum_{\underline{x}_{30}\in\{0,1\}^{30}, \sum x_i = j} \tilde{\varphi}(\underline{x}_{30}) \cdot \mathbb{P}_p[\underline{X}_{30} = \underline{x}_{30}] \\ &= \gamma \cdot \underbrace{\mathbb{P}_p[S_n = 3]}_{\text{monoton wachsend für } p\in(0,\frac{1}{10})} + \underbrace{\mathbb{P}_p[S_n \geq 4]}_{\text{monoton wachsend für } p\in(0,1)}\end{aligned}$$

Damit

$$\begin{aligned}\sup_{p\leq 0.02} \mathbb{P}_p[H_0 \text{ ablehnen}] = \gamma \cdot \mathbb{P}_{0.02}[S_{30} = 3] + \mathbb{P}_{0.02}[S_{30} \geq 4] &\overset{!}{=} \alpha \\ \Rightarrow \gamma = \frac{\alpha - \mathbb{P}_{0.02}[S_{30} \geq 4]}{\mathbb{P}_{0.02}[S_{30} = 3]} &= 0.3772\end{aligned}$$

Der zugehörige Test $\tilde{\varphi}$ schöpft das Niveau α voll aus.

Sei $\varphi : \mathcal{X} \to [0,1]$ ein Test mit $\mathcal{X}$ beliebig. Dann

$$\begin{aligned}\mathbb{P}_\theta[\varphi \text{ verwirft } H_0] &= \int_{\mathcal{X}} \mathbb{P}_\theta[\varphi \text{ verwirft } H_0 | X = x]\ (\mathbb{P}_\theta \circ X^{-1})(dx) \\ &= \int_{\mathcal{X}} \varphi(x)\, (\mathbb{P} \circ X^{-1})(dx) = \mathbb{E}_\theta(\varphi(X)) \qquad (\theta \in \Theta)\end{aligned} \tag{8.8}$$

8.6 Definition

Sei $\varphi : \mathcal{X} \to [0,1]$ ein Test im statistischen Modell $(X, \mathcal{X}, \mathcal{F}, \mathcal{P})$. Dann heißt $\beta_\varphi : \Theta \to [0,1]$, $\beta_\varphi(\theta) = \mathbb{E}_\theta(\varphi(X))$ Gütefunktion von φ. Gemäß (8.8) gibt $\beta_\varphi(\theta)$ die Wahrscheinlichkeit an, dass φ die Nullhypothese H_0 bei Vorliegen der Verteilung $\mathbb{P}_\theta$ ablehnt.

8.7 Beispiel: Fortsetzung von 8.5

Betrachte den nichtrandomisierten Test für $H_0 : p \leq p_0$ gegen $H_1 : p > p_0$. H_0 ablehnen $\Leftrightarrow S_n > k_\alpha$ (vgl. 8.5) $\Leftrightarrow \varphi(\underline{X}_n) = 1_{[\sum_{i=1}^n X_i > k_\alpha]}$. Damit

$$\begin{aligned}\beta_\varphi(p) &= \mathbb{E}_p(\varphi(\underline{X}_n)) = \mathbb{P}[S_n > k_\alpha] \\ &= \sum_{k=k_\alpha+1}^{n} \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}\end{aligned}$$

Für $n = 30, \alpha = 0.01, p_0 = 0.02$ erhält man:

$$\beta_\varphi(p) = \sum_{k=4}^{30} \binom{30}{k} \cdot p^k \cdot (1-p)^{30-k} \qquad (p \in (0,1))$$

Vergleich mit randomisierten Test $\tilde{\varphi}$ aus (8.7):

$$\tilde{\varphi}(\underline{x}_{30}) := 1_{[\sum_{i=1}^{30} x_i \geq 4]} + \gamma \cdot 1_{[\sum_{i=1}^{30} x_i = 3]}$$

mit $\gamma = 0.3772$ (vgl. (8.4)). Es folgt

$$\begin{aligned}\beta_{\tilde{\varphi}}(p) &= \beta_\varphi(p) + \gamma \cdot \mathbb{P}_p[S_{30} = 3] = \beta_\varphi(p) + 1540.62 \cdot p^3 \cdot (1-p)^{27} \\ &\geq \beta_\varphi(p)\end{aligned} \tag{8.9}$$

Also $\tilde{\varphi}$ besser als φ.

8.8 Beispiel

Gütefunktion des Gauß-Tests aus 8.2

$$
\begin{aligned}
\varphi(\underline{x}_n) &= 1_{\left[\underline{x}_n > \mu_0 + \frac{\sigma_0}{\sqrt{n}} \cdot u_{1-\alpha}\right]} \\
\Rightarrow \beta_\varphi(\mu) &= \mathbb{E}_\mu(\varphi(\underline{X}_n)) = \mathbb{P}_\mu\left[\overline{X}_n > \mu_0 + \frac{\sigma_0}{\sqrt{n}} \cdot u_{1-\alpha}\right] \\
&= \mathbb{P}_\mu\left[\sqrt{n} \cdot \frac{\overline{X}_n - \mu}{\sigma_0} > \sqrt{n} \cdot \frac{\mu_0 - \mu}{\sigma_0} + u_{1-\alpha}\right] \\
&= 1 - \Phi\left(\sqrt{n} \cdot \frac{\mu_0 - \mu}{\sigma_0} + u_{1-\alpha}\right) = \Phi\left(-\sqrt{n} \cdot \frac{\mu_0 - \mu}{\sigma_0} - u_{1-\alpha}\right) \\
&= \Phi\left(\sqrt{n} \cdot \frac{\mu - \mu_0}{\sigma_0} + u_\alpha\right)
\end{aligned}
$$

Klar: Die Gütefunktion ist entscheidender Begriff zur Beurteilung der Qualität (Güte) von Tests.

8.9 Definition

(i). Sei φ ein Test mit Gütefunktion β_φ. Dann heißt φ Niveau-α-Test für H_0 :⇔ Für alle $\theta \in \Theta_0$: $\beta_\varphi(\theta) \leq \alpha$.

(ii). Seien $\varphi, \tilde{\varphi}$ Niveau-α-Tests für H_0. Dann heißt φ gleichmäßig besser als $\tilde{\varphi}$ für H_1 :⇔ Für alle $\theta \in \Theta_1$: $\beta_\varphi(\theta) \geq \beta_{\tilde{\varphi}}(\theta)$.

(iii). Ein Niveau-α-Test φ heißt gleichmäßig bester Niveau-α-Test für H_0 gegen H_1 :⇔ $\beta_\varphi(\theta) \geq \beta_{\tilde{\varphi}}(\theta)$ für alle $\theta \in \Theta_1$ mit $\tilde{\varphi}$ Niveau-α-Test für H_0. (Mit anderen Worten: φ ist gleichmäßig besser als jeder andere Niveau-α-Test $\tilde{\varphi}$.)

Also: Gleichmäßig beste Niveau-α-Tests halten auf H_0 das Niveau α ein und minimieren unter allen solchen Tests die Wahrscheinlichkeiten für den Fehler 2.Art $(1 - \beta_\varphi(\theta)) = \mathbb{P}[\varphi$ verwirft H_1 nicht$]$ für $\theta \in \Theta_1$. In Beispiel 8.7 sind φ Niveau-α-Tests und $\tilde{\varphi}$ und wegen (8.9) ist $\tilde{\varphi}$ gleichmäßig besser als φ.

Ziel: Konstruktion gleichmäßig bester Niveau-α-Tests in (parametrischen) statistischen Modellen (unter bestimmten Regularitätsannahmen)

9 Das Neyman-Pearson-Lemma

Sei $(X, \mathcal{X}, \mathcal{F}, \mathcal{P})$ mit $\mathcal{P} = \{P_0, P_1\}$ mit $P_0 \neq P_1$, $X \sim P$. Testproblem: $H_0 : P = P_0$ gegen $H_1 : P = P_1$ (sogenannte einfache Hypothese mit einfacher Alternative = einfachste Situation). Es wird sich zeigen, dass hier gleichmäßig beste Niveau-α-Tests konstruierbar sind. Sie bilden die Grundlage für die Konstruktion optimaler Tests in komplexeren Modellen. Erinnere an:

9.1 Definition

Sei $(\mathcal{X}, \mathcal{F})$ ein Meßraum, μ und ν Maße auf $(\mathcal{X}, \mathcal{F})$. Dann heißt ν absolut-stetig bzgl. μ, falls $\mu(A) = 0 \Rightarrow \nu(A) = 0$ für alle $A \in \mathcal{F}$. Notation: $\nu << \mu$.

9.2 Satz von Radon-Nikodým

Seien $(\mathcal{X}, \mathcal{F})$ ein Meßraum und μ, ν σ-endliche Maße auf $(\mathcal{X}, \mathcal{F})$. Dann äquivalent:

(i). $\nu << \mu$

(ii). ν besitzt eine μ-Dichte f mit Werten in $[0, \infty)$.

Gilt (ii), so sind zwei μ-Dichten von ν μ-fast überall gleich.

Notation: $f = \frac{d\nu}{d\mu}$

Beweis: MAST

9.3 Folgerung

Zu P_0 und P_1 Wahrscheinlichkeitsmaße existieren stets μ-Dichten zu z.B. $\mu = P_0 + P_1$.

Beweis:

- $\mu(\mathcal{X}) = P_0(\mathcal{X}) + P_1(\mathcal{X}) = 2$, also ist μ finit (insbesondere σ-endlich). Klar: $P_0 << \mu$ und $P_1 << \mu$, denn $\mu(F) = 0 \Rightarrow 0 = P_0(F) + P_1(F) \Rightarrow P_0(F) = P_1(F) = 0$ wegen $P_i(A) \geq 0$ für alle $A \in \mathcal{F}$ $(i \in \{0,1\})$. Behauptung folgt mit Satz 9.2.

9.4 Satz: Neyman-Pearson-Lemma

Zu $(X, \mathcal{X}, \mathcal{F}, \mathcal{P})$ mit $\mathcal{P} = \{P_0, P_1\}$ seien f_0 bzw. f_1 Dichten von P_0 bzw. P_1 bzgl. eines σ-endlichen Maßes μ auf $(\mathcal{X}, \mathcal{F})$. Sei $\alpha \in (0,1)$ und

$$G_0(t) := P_0\left(\left\{x \in \mathcal{X}; \frac{f_1(x)}{f_2(x)} \leq t\right\}\right) = \mathbb{P}_0\left[\frac{f_1(X)}{f_2(X)} \leq t\right]$$

(d.h. G_0 ist Verteilungsfunktion von $\frac{f_1(X)}{f_0(X)}$ unter $\mathbb{P}_0$). Für das Testproblem $H_0 : P = P_0$ gegen $H_1 : P = P_1$ gilt:

(i). Jeder Test $\varphi : \mathcal{X} \to [0,1]$ mit $\mathbb{E}_0(\varphi(X)) = \alpha$, zu dem ein $k \in [0,\infty)$ existiert mit

$$\varphi(x) = \begin{cases} 1 & \text{falls } f_1(x) > k \cdot f_0(x) \\ 0 & \text{falls} f_1(x) < k \cdot f_0(x) \end{cases} \tag{9.1}$$

ist gleichmäßig bester Niveau-α-Test für H_0 gegen H_1.

(ii). Es existiert ein $k \in [0,\infty)$ und $\gamma \in [0,1]$ mit

$$P_0[f_1 > k \cdot f_0] + \gamma \cdot P_0[f_1 = k \cdot f_0] = \alpha \tag{9.2}$$

nämlich

$$\begin{aligned} k &:= G_0^{-1}(1-\alpha) \\ \gamma &:= \begin{cases} \dfrac{G_0(k) - (1-\alpha)}{G_0(k) - G_0(k-)} & G_0(k) > G_0(k-) \\ \text{beliebig} \in [0,1] & G_0(k) = G_0(k-) \end{cases} \end{aligned}$$

(iii). Für jede Lösung $(k,\gamma) \in [0,\infty) \times [0,1]$ von (9.2) ist

$$\varphi_0 := 1_{[f_1 > k \cdot f_0]} + \gamma \cdot 1_{[f_1 = k \cdot f_0]}$$

gleichmäßig bester Niveau-α-Test für H_0 gegen H_1.

Bemerkung:

(i). Ist G_0 nicht stetig, so ist $G_0(G_0^{-1}(1-\alpha)) \neq 1-\alpha$.

Beweis:

(i). Nach Voraussetzung gilt $\mathbb{E}_0(\varphi(X)) = \alpha$, also ist φ ein Niveau-α-Test für H_0. Sei $\tilde{\varphi}$ ein weiterer Niveau-α-Test für H_0, d.h. $\mathbb{E}_0(\tilde{\varphi}(X)) \leq \alpha$. Zu zeigen: $\mathbb{E}_1(\varphi(X)) \geq \mathbb{E}_1(\tilde{\varphi}(X))$. Es gilt:

$$\tilde{\varphi}(x) \overset{>}{\underset{<}{}} \varphi(x) \Rightarrow \varphi(x) \overset{<1}{\underset{>0}{}} \Rightarrow f_1(x) \overset{\leq}{\underset{\geq}{}} k \cdot f_0(x)$$

Daher folgt:

$$\forall x \in \mathcal{X} : (\tilde{\varphi} - \varphi)(k \cdot f_0 - f_1)(x) \geq 0$$

Man erhält somit

$$\begin{aligned} 0 &\leq \int_{\mathcal{X}} (\tilde{\varphi} - \varphi)(k \cdot f_0 - f_1)\, d\mu \\ &= k \cdot \left(\int \tilde{\varphi} \cdot \underbrace{f_0\, d\mu}_{dP_0} - \int \varphi \cdot \underbrace{f_0\, d\mu}_{dP_0} \right) - \left(\int \tilde{\varphi} \cdot \underbrace{f_1\, d\mu}_{dP_1} - \int \varphi \cdot \underbrace{f_1\, d\mu}_{dP_1} \right) \\ &= \underbrace{k}_{\geq 0} \cdot \Big(\underbrace{\mathbb{E}_0(\tilde{\varphi}(X))}_{\leq \alpha} - \underbrace{\mathbb{E}_0(\varphi(X))}_{\alpha} \Big) - \big(\mathbb{E}_1(\tilde{\varphi}(X)) - \mathbb{E}_1(\varphi(X)) \big) \\ &\leq -\mathbb{E}_1(\tilde{\varphi}(X)) + \mathbb{E}_1(\varphi(X)) \end{aligned} \tag{9.3}$$

also $\mathbb{E}_1(\varphi(X)) \geq \mathbb{E}_1(\tilde{\varphi}(X))$.

(ii). Es gilt $0 \leq \frac{f_1}{f_0} < \infty$ P_0-fast überall, also $G_0(t) = 0$ für alle $t < 0$ und somit $k := G_0^{-1}(1-\alpha) \in [0,\infty)$. Aus den Eigenschaften der Quantilfunktionen (→ Witting, Mathematische Statistik) erhält man

$$G_0(k-) = G_0(G_0^{-1}(1-\alpha)-) \leq 1-\alpha \leq G_0(G_0^{-1}(1-\alpha)) = G_0(k)$$

Somit folgt $\gamma \in [0,1]$ für $G_0(k) \neq G_0(k-)$:

$$\gamma = \frac{G_0(k) - (1-\alpha)}{G_0(k) - G_0(k-)} \leq \frac{G_0(k) - G_0(k-)}{G_0(k) - G_0(k-)} = 1$$

Außerdem:

$$
\begin{aligned}
(9.2) \Leftrightarrow P_0\left[\frac{f_1}{f_0} > k\right] + \gamma \cdot P_0\left[\frac{f_1}{f_0} = k\right] &= \alpha \\
\Leftrightarrow 1 - G_0(k) + \gamma \cdot (G_0(k) - G_0(k-)) &= \alpha \qquad (9.4)
\end{aligned}
$$

Falls $G_0(k) > G_0(k-)$ dann ist (9.4) eindeutig nach γ auflösbar. Falls $G_0(k) = G_0(k-)$, dann folgt $G_0(k) = 1 - \alpha$ und somit die Behauptung.

(iii). Aus der Definition von φ_0 folgt

$$
\begin{aligned}
\mathbb{E}_0(\varphi_0(X)) = \int \varphi_0 \, dP_0 &= P_0[f_1 > k \cdot f_0] + \gamma \cdot P_0[f_1 = k \cdot f_0] \\
&\overset{(ii)}{=} \alpha
\end{aligned}
$$

und φ_0 erfüllt die Voraussetzungen in (i). Es folgt, dass φ_0 gleichmäßig bester Test ist.

Bemerkungen:

(i). Tests der Form (9.1) heißen Tests vom Neyman-Pearson-Typ.

(ii). Ist $\gamma \in \{0,1\}$ wählbar, so ist φ ein nichtrandomisierter Test. Dies ist zum Beispiel stets der Fall, wenn G_0 stetig ist.

9.5 Beispiel

Seien $X_1, \dots, X_n$ iid mit $X_j \sim P \in \{P_0, P_1\}$. $H_0 : P = P_0 := \text{Poi}(1)$ vs. $H_1 : P = P_1 := \text{Poi}(\lambda)$ mit $\lambda > 1$ bekannt. Statistisches Modell $(\underline{X}_n, \mathbb{N}_0^n, \mathcal{P}(\mathbb{N}_0^n), \mathcal{P}^n)$ mit $\mathcal{P}^n = \{P_0^n, P_1^n\} << \mu^n$ mit $\mu = \sum_{j \in \mathbb{N}_0} \delta_j$. Satz 4.5: P_i^n hat μ^n-Dichte

$$
\begin{aligned}
f_{0,n}(\underline{x}_n) &= \prod_{j=1}^n e^{-1} \cdot \frac{1}{x_j!} = e^{-n} \cdot \prod_{j=1}^n \frac{1}{x_j!} \qquad (\underline{x}_n \in \mathbb{N}_0^n) \\
f_{1,n}(\underline{x}_n) &= \prod_{j=1}^n e^{-\lambda} \cdot \frac{\lambda^{x_j}}{x_j!} = e^{-n\lambda} \cdot \lambda^{\sum_{j=1}^n x_j} \cdot \prod_{j=1}^n \frac{1}{x_j!}
\end{aligned}
$$

Es folgt:

$$
\frac{f_{1,n}}{f_{0,n}}(\underline{x}_n) = \lambda^{\sum_{j=1}^n x_i} \cdot e^{-n \cdot (\lambda - 1)}
$$

Für alle $k > 0$ und $\underline{x}_n \in \mathbb{N}_0^n$ gilt:

$$
\frac{f_{1,n}}{f_{0,n}}(\underline{x}_n) \stackrel{>}{=} k \overset{\lambda>1}{\Leftrightarrow} \sum_{j=1}^n x_j \stackrel{>}{=} \frac{\log k + n \cdot (\lambda - 1)}{\log \lambda} =: \tilde{k}
$$

Ansatz für Neyman-Pearson-Test aus 9.4:

$$
\varphi(\underline{x}_n) := 1_{[f_{1,n} > k \cdot f_{0,n}]} + \gamma \cdot 1_{[f_{1,n} = k \cdot f_{0,n}]} = 1_{[\sum_j x_j > \tilde{k}]} + \gamma \cdot 1_{[\sum_j x_j = \tilde{k}]} \qquad (+)
$$

für alle $\underline{x}_n \in \mathbb{N}_0^n, k > 0, \gamma \in [0,1]$. Außerdem:

$$
\mathbb{P}_0 \circ \left(\sum_{j=1}^n X_j\right)^{-1} = \text{Poi}(n) \qquad (++)
$$

gemäß Faltungsgesetz für Poisson-Verteilung. Somit

$$
\begin{aligned}
\mathbb{E}_0(\varphi(\underline{X}_n)) &\overset{(+)}{=} \mathbb{P}_0\left[\sum_j X_j > \tilde{k}\right] + \gamma \cdot \mathbb{P}_0\left[\sum_j X_j = \tilde{k}\right] \qquad (9.5) \\
&\overset{(++)}{=} 1 - \underbrace{\sum_{1 \le j \le \tilde{k}} e^{-n} \cdot \frac{n^j}{j!}}_{=: H_n(\tilde{k})} + \gamma \cdot e^{-n} \frac{n^{\tilde{k}}}{\tilde{k}!} \overset{!}{=} \alpha
\end{aligned}
$$

Damit Lösung $(\tilde{k}, \gamma) \in [0, \infty) \times [0, 1]$ gegeben durch (vgl. Beweis von (ii) in 9.4, ersetze G_0 durch H_n)

$$\tilde{k} := H_n^{-1}(1-\alpha) = \min\{k \in \mathbb{N}_0; H_n(k) \geq 1-\alpha\}$$
$$\gamma := \frac{H_n(\tilde{k}) - (1-\alpha)}{H_n(\tilde{k}) - H_n(\tilde{k}-1)} = \frac{\sum_{j=1}^{\tilde{k}} \frac{e^{-n} \cdot n^j}{j!} - (1-\alpha)}{e^{-n} \cdot \frac{n^{\tilde{k}}}{\tilde{k}!}}$$

Also ist φ ein exakter Niveau-α-Test für H_0 und wegen (+) vom Neyman-Pearson-Typ (9.1). Aus 9.4(i) folgt, dass φ gleichmäßig bester Niveau-α-Test für H_0 gegen H_1 ist.

Beachte: φ hängt nicht von λ ab.

Zahlenbeispiel: $n = 30, \alpha = 0.05$, dann $\tilde{k} = 39$ und $\gamma = 0.20154$. Neyman-Pearson-Test:

$$\varphi(x_1, \ldots, x_{30}) = \begin{cases} 1 & \sum_{i=1}^{30} x_i > 39 \\ 0.20154 & \sum_{i=1}^{30} x_i = 39 \\ 0 & \sum_{i=1}^{30} x_i < 39 \end{cases}$$

10

Einseitige Testprobleme in Modellen mit monotonen Likelihood-Quotienten

Ziel: Anwendung des Neyman-Pearson-Lemmas 9.4 zur Konstruktion gleichmäßig bester Niveau-α-Tests bei zusammengesetzten (d.h. nicht einelementigen) Hypothesen und Alternativen.

Fortsetzung Beispiel 8.2 Seien $X_1, \ldots, X_n$ iid mit $X_j \sim N(\mu, \sigma_0^2)$, $H_0 : \mu \leq \mu_0$ gegen $H_1 : \mu > \mu_0$ mit μ_0 vorgegeben und σ_0^2 bekannt. Solche Probleme heißen einseitige Testprobleme. Ansatz: Fixiere ein $\mu_1 > \mu_0$ und betrachte das Testproblem $H_0' : \mu = \mu_0$ vs. $H_1' : \mu = \mu_1$. Gemäß Neyman-Pearson-Lemma 9.4 gilt für alle $k \geq 0, \gamma \in [0,1]$ mit

$$P_{\mu_0}[f_1 > kf_0] + \gamma \cdot P_{\mu_0}[f_1 = kf_0] = \alpha$$

wobei f_i λ^n-Dichte von $N(\mu_i, \sigma_0^2)^n$, dass

$$\varphi = 1_{[f_1 > kf_0]} + \gamma \cdot 1_{[f_1 = kf_0]}$$

gleichmäßig bester Niveau-α-Test für H_0' gegen H_1'. Idee: Zeige

(i). φ hängt nicht von μ_1 ab. Dann ist φ gleichmäßig bester Niveau-α-Test für H_0' gegen H_1, denn:

Sei $\tilde{\varphi}$ ein Niveau-α-Test für H_0' gegen H_1 und sei $\mu_1 > \mu_0$ beliebig. Dann φ gleichmäßig besser als $\tilde{\varphi}$ für H_0' gegen H_1', d.h. $\beta_\varphi(\mu_1) \geq \beta_{\tilde{\varphi}}(\mu_1)$. Da $\mu_1 > \mu_0$ beliebig gilt $\beta_\varphi(\mu) \geq \beta_{\tilde{\varphi}}(\mu)$ für alle $\mu > \mu_0$.

(ii). φ ist Niveau-α-Test für (sogar) ganz H_0.

Aus (i) und (ii) folgt, dass φ gleichmäßig bester Niveau-α-Test für H_0 gegen H_1 ist, denn sei $\tilde{\varphi}$ ein Niveau-α-Test für H_0, dann ist $\tilde{\varphi}$ insbesondere Niveau-α-Test für H_0'. Aus (i) folgt, dass φ gleichmäßig besser als $\tilde{\varphi}$ auf H_1.

Beweis:

(i). Gemäß 3.7 gilt für $i = 0, 1$:

$$f_{in}(\underline{x}_n) = (2\pi\sigma_0^2)^{-\frac{n}{2}} \cdot \exp\left(-\frac{1}{2\sigma_0^2}\sum_{j=1}^n (x_j - \mu_i)^2\right) \qquad (\underline{x}_n \in \mathbb{R}^n)$$

Damit folgt:

$$f_{1n}(\underline{x}_n) \underset{=}{>} k \cdot f_{0n}(\underline{x}_n) \Leftrightarrow \frac{f_{1n}(\underline{x}_n)}{f_{0n}(\underline{x}_n)} \underset{=}{>} k \Leftrightarrow \exp\left(-\frac{1}{2\sigma_0^2}\sum_{j=1}^n (x_j - \mu_1)^2 - (x_j - \mu_0)^2\right) \underset{=}{>} k$$

$$\Leftrightarrow \exp\left(\frac{1}{\sigma_0^2} \cdot \left((\mu_1 - \mu_0) \cdot \sum_{j=1}^n x_j - \frac{n}{2} \cdot (\mu_1^2 - \mu_0^2)\right)\right) \underset{=}{>} k$$

$$\Leftrightarrow \sqrt{n} \cdot \frac{\overline{x}_n - \mu_0}{\sigma_0} \underset{=}{>} \frac{\sigma_0 \cdot \log(k)}{\sqrt{n} \cdot (\mu_1 - \mu_0)} + \frac{\sqrt{n} \cdot (\mu_1 - \mu_0)}{2\sigma_0} =: \tilde{k}$$

Also

$$\varphi(\underline{x}_n) = 1_{\left[\sqrt{n}\cdot\frac{\overline{x}_n-\mu_0}{\sigma_0}>\tilde{k}\right]} + \gamma\cdot 1_{\left[\sqrt{n}\cdot\frac{\overline{x}_n-\mu_0}{\sigma_0}=\tilde{k}\right]}$$

und dieses φ ist Niveau-α-Test für $H_0' : \mu = \mu_0$ genau dann wenn $(\tilde{k},\tilde{\gamma})$ Lösung von

$$\mathbb{P}_{\mu_0}\left[\sqrt{n}\cdot\frac{\overline{X}_n-\mu_0}{\sigma_0} > \tilde{k}\right] + \gamma\cdot\underbrace{\mathbb{P}_{\mu_0}\left[\sqrt{n}\cdot\frac{\overline{X}_n-\mu_0}{\sigma_0} = \tilde{k}\right]}_{0} = \alpha \qquad (*)$$

Da die Bestimmungsgleichung $(*)$ nicht von μ_1 abhängt, folgt (i).

(ii). Es gilt:

$$(*) \Leftrightarrow 1-\Phi(\tilde{k}) = \alpha \Leftrightarrow \tilde{k} = \Phi^{-1}(1-\alpha) = u_{1-\alpha}$$

und $\gamma\in[0,1]$ beliebig, zum Beispiel $\gamma = 0$. Somit

$$\varphi(\underline{X}_n) = 1_{\left[\sqrt{n}\cdot\frac{\overline{X}_n-\mu_0}{\sigma_0}>u_{1-\alpha}\right]}$$

(einseitiger Gauß-Test aus 8.2). Dieser hat gemäß 8.8 die Gütefunktion

$$\beta_\varphi(\mu) = \Phi\left(\sqrt{n}\cdot\frac{\mu-\mu_0}{\sigma_0}+u_\alpha\right) \qquad (\mu\in\mathbb{R})$$

Somit β_φ monoton wachsend, also $\beta_\varphi(\mu)\le\beta_\varphi(\mu_0) = \alpha$ für alle $\mu\le\mu_0$. Damit ist φ Niveau-α-Test für H_0.

Zusammenfassung: Im Normalverteilungsmodell 8.2 ist der einseitige Gauß-Test gleichmäßig bester Niveau-α-Test für $H_0 : \mu\le\mu_0$ gegen $H_1 : \mu > \mu_0$.

Nächstes Ziel: Verallgemeinerung der obigen Vorgehensweise auf Verteilungsannahmen $\mathcal{P}$, die ein analoges Monotonie-Argument zulassen.

10.1 Definition

Sei $(X,\mathcal{X},\mathcal{F},\mathcal{P})$ ein statistisches Modell mit $\mathcal{P} = \{P_\theta;\theta\in\Theta\} << \mu$ mit μ σ-finit, $\Theta\subseteq\mathbb{R}$ und Likelihood-Funktion $L : \Theta\times\mathcal{X}\to[0,\infty)$. Sei $T : \mathcal{X}\to\mathbb{R}$ eine Statistik. Dann: $\mathcal{P}$ hat monotone Likelihood-Quotienten in T :$\Leftrightarrow$ Für alle $\theta,\theta'\in\Theta$ mit $\theta<\theta'$ existiert eine monoton wachsende Funktion $h(\cdot,\theta,\theta') : \mathbb{R}\to[0,\infty]$ mit

$$\frac{L(\theta',x)}{L(\theta,x)} = h(T(x),\theta,\theta')$$

für $P_\theta + P_{\theta'}$ fast alle $x\in\mathcal{X}$.
Nach Konvention ist dabei für $a>0$ stets $\frac{a}{0} := \infty$ und der nicht-definierte Ausdruck $\frac{0}{0}$ hat $P_\theta+P_{\theta'}$-Maß Null und kann dort beliebig definiert werden.

Diese Forderung bedeutet also: Für alle $\theta<\theta'$ können die Likelihood-Quotienten als Funktion in $T(x)$ geschrieben werden und sind in der Variablen $T(x)$ monoton wachsend.

10.2 Beispiel

$\mathcal{P}$ sei eine einparametrige Exponentialfamilie in q und T. Dann

$$\begin{aligned}\frac{L(\theta',x)}{L(\theta,x)} &= \frac{c(\theta')\cdot h(x)\cdot\exp(q(\theta')\cdot T(x))}{c(\theta)\cdot h(x)\cdot\exp(q(\theta)\cdot T(x))}\\ &= \frac{c(\theta')}{c(\theta)}\cdot\exp\left(T(x)\cdot(q(\theta')-q(\theta))\right)\end{aligned}$$

Falls q monoton wachsend ist, hat $\mathcal{P}$ monotone Likelihood-Quotienten in T,

$$h(u,\theta,\theta') = \frac{c(\theta')}{c(\theta)}\cdot\exp\left(u\cdot(q(\theta')-q(\theta))\right)$$

10.3 Beispiel

Seien $X_1, \dots, X_n$ iid, $X_j \sim \beta_p$ mit $p \in (0,1) =: \Theta$. Aus 4.9 folgt, dass

$$\mathcal{P} = \{(\beta_p)^n; p \in (0,1)\}$$

einparametrige Exponentialfamilie in $q(p) = \log \frac{p}{1-p}, T(\underline{x}_n) = \sum_{i=1}^n x_i$. Da q streng monoton wachsend ist, erhält man aus 10.2, dass $\mathcal{P}$ monotone Likelihood-Quotienten hat.

Es lässt sich die Argumentation bei der Herleitung der Optimalität des Gauß-Tests übertragen:

10.4 Satz

Es sei $(X, \mathcal{X}, \mathcal{F}, \mathcal{P})$ ein statistisches Modell, wobei $\mathcal{P} = \{P_\theta; \theta \in \Theta\}$ mit $\Theta \subseteq \mathbb{R}$ monotone Likelihood-Quotienten in einer Statistik T besitze. Dann gilt für alle $\theta_0 \in \Theta$, $\alpha \in (0,1)$:

(i). Es existieren $k \in \mathbb{R}, \gamma \in [0,1]$ mit

$$P_{\theta_0}[T > k] + \gamma \cdot P_{\theta_0}[T = k] = \alpha \tag{10.1}$$

nämlich, falls G_0 Verteilungsfunktion von $P_{\theta_0} \circ T^{-1}$:

$$\begin{aligned} k &= G_0^{-1}(1-\alpha) \\ \gamma &= \begin{cases} \dfrac{G_0(k) - (1-\alpha)}{G_0(k) - G_0(k-)} & G_0(k) - (1-\alpha) > 0 \\ \in [0,1] & \text{sonst} \end{cases} \end{aligned}$$

(ii). Für jede Lösung $(k, \gamma) \in \mathbb{R} \times [0,1]$ von (10.1) hat der Test

$$\varphi(x) := 1_{[T>k]}(x) + \gamma \cdot 1_{[T=k]}(x) \qquad (x \in \mathcal{X})$$

die folgenden Eigenschaften:

$$\beta_\varphi(\theta_0) = \alpha \tag{10.2}$$
$$\forall \theta \le \theta_0 : \beta_\varphi(\theta) \le \beta_{\tilde{\varphi}}(\theta) \tag{10.3}$$
$$\forall \theta > \theta_0 : \beta_\varphi(\theta) \ge \beta_{\hat{\varphi}}(\theta) \tag{10.4}$$

für alle Tests $\tilde{\varphi}$ mit $\beta_{\tilde{\varphi}}(\theta_0) = \alpha$ bzw. alle Tests $\hat{\varphi}$ mit $\beta_{\hat{\varphi}}(\theta_0) \le \alpha$

(iii). φ aus (ii) ist gleichmäßig bester Niveau-α-Test für $H_0 : \theta \le \theta_0$ gegen $H_1 : \theta > \theta_0$.

10.5 Bemerkung

Die Aussagen (10.2)-(10.4) besagen: φ minimiert die Wahrscheinlichkeiten für den Fehler 2.Art und 1. Art gleichmäßig in θ unter allen Tests $\tilde{\varphi}$ mit $\beta_{\tilde{\varphi}}(\theta_0) = \alpha$.

Beweis von 10.4:

(i). Es gilt gemäß Definition von G_0:

$$(10.1) \Leftrightarrow 1 - G_0(k) + \gamma \cdot (G_0(k) - G_0(k-)) = \alpha$$

Die Behauptung folgt analog zum Beweis von (ii) in Satz 9.4 ($\frac{f_1}{f_0} \longleftrightarrow T$).

(ii). Sei (k, γ) Lösung von (10.1). Zu (10.2):

$$\begin{aligned} \beta_\varphi(\theta_0) &= \mathbb{E}_{\theta_0}(\varphi(X)) = \mathbb{P}_{\theta_0}[T(X) > k] + \gamma \cdot \mathbb{P}_{\theta_0}[T(X) = k] \\ &\overset{(10.1)}{=} \alpha \end{aligned}$$

Zum Nachweis von (10.3) und (10.4) zeige: $\forall \theta_1 > \theta_0 \exists k_{\theta_1} \in [0,\infty), F = F(\theta_0,\theta_1) \in \mathcal{F}, (P_{\theta_0} + P_{\theta_1})(F^c) = 0$:

$$\varphi(x) = \begin{cases} 1 & x \in F \cap [L(\theta_1,\cdot) > k_{\theta_1} \cdot L(\theta_0,\cdot)] \\ 0 & x \in F \cap [L(\theta_1,\cdot) < k_{\theta_1} \cdot L(\theta_0,\cdot)] \end{cases} \tag{10.5}$$

Beweis von (10.5): Sei $k_{\theta_1} := h(k,\theta_0,\theta_1)$. Angenommen $k_{\theta_1} = \infty$. Dann

$$\begin{aligned} \alpha &\overset{(10.1),\gamma\le 1}{\le} P_{\theta_0}[T \ge k] \overset{h(\cdot,\theta_0,\theta_1)\uparrow}{\le} P_{\theta_0}[\underbrace{h(T,\theta_0,\theta_1)}_{\frac{L(\theta_1,\cdot)}{L(\theta_0,\cdot)}} \ge \underbrace{h(k,\theta_0,\theta_1)}_{k_{\theta_1}=\infty}] \\ &= P_{\theta_0}\left[\frac{L(\theta_1,\cdot)}{L(\theta_0,\cdot)} = \infty\right] \le P_{\theta_0}[L(\theta_0,\cdot) = 0] = \int_{[L(\theta_0,\cdot)]=0} L(\theta_0,\cdot)\, d\mu \\ &= 0 \end{aligned}$$

Widerspruch zu $\alpha > 0$. Also $k_{\theta_1} \in [0,\infty)$.
Sei gemäß Definition 10.1 $F = F(\theta_0,\theta_1) \in \mathcal{F}$ mit $(P_{\theta_0} + P_{\theta_1})(F^c) = 0$ und

$$\forall x \in F : \frac{L(\theta_1,x)}{L(\theta_0,x)} = h(T(x),\theta_0,\theta_1)$$

Sei nun $x \in F$ mit $L(\theta_1,x) > k_{\theta_1} \cdot L(\theta_0,x)$, dann

$$h(T(x),\theta_0,\theta_1) = \frac{L(\theta_1,x)}{L(\theta_0,x)} > k_{\theta_1} = h(k,\theta_0,\theta_1)$$

Da $h(\cdot,\theta_0,\theta_1)$ monoton wachsend ist, folgt $T(x) > k$. Also $\varphi(x) = 1$ nach Definition von φ. Anderer Fall folgt analog.
Sei

$$\varphi_0(x) := \begin{cases} 1 & L(\theta_1,x) > k_{\theta_1} \cdot L(\theta_0,x) \\ 0 & L(\theta_1,x) < k_{\theta_1} \cdot L(\theta_0,x) \end{cases}$$

Da $[\varphi \ne \varphi] \subseteq F^c$ nach (10.5) folgt

$$\varphi = \varphi_0 \; P_{\theta_0}, P_{\theta_1} - \text{fast überall} \tag{*}$$

Beachte: φ_0 ist Test vom Neyman-Pearson-Typ (9.1) für das Testproblem $H_0' : \theta = \theta_0$ gegen $H_1' : \theta = \theta_1$ für alle $\theta_1 > \theta_0$. Ferner:

$$\begin{aligned} \beta_{\varphi_0}(\theta_0) &= \mathbb{E}_{\theta_0}(\varphi_0(X)) = \int \varphi_0 \, dP_{\theta_0} \\ &\overset{*}{=} \int \varphi \, dP_{\theta_0} = \beta_\varphi(\theta_0) \overset{(10.2)}{=} \alpha \end{aligned}$$

Gemäß (9.4)(i) ist φ_0 gleichmäßig bester Niveau-α-Test für H_0' gegen H_1', d.h. für alle Tests $\tilde{\varphi}$ mit $\beta_{\tilde{\varphi}}(\theta_0) \le \alpha$ gilt

$$\forall \theta_1 > \theta_0 : \beta_\varphi(\theta_1) \overset{*}{=} \beta_{\varphi_0}(\theta_1) \ge \beta_{\tilde{\varphi}}(\theta_1) \tag{10.6}$$

Dies zeigt (10.4).
Zum Nachweis von (10.3) zeige: Für alle $\theta_1 < \theta_0$ existiert $\bar{k}_{\theta_1} \in (0,\infty]$ und $\bar{F} = \bar{F}(\theta_0,\theta_1) \in \mathcal{F}$ mit $(P_{\theta_0} + P_{\theta_1})(\bar{F}^c) = 0$ derart, dass

$$1 - \varphi(x) = \begin{cases} 1 & x \in \bar{F} \cap \left[L(\theta_1,\cdot) > \frac{1}{\bar{k}_{\theta_1}} \cdot L(\theta_0,\cdot)\right] \\ 0 & x \in \bar{F} \cap \left[L(\theta_1,\cdot) < \frac{1}{\bar{k}_{\theta_1}} \cdot L(\theta_0,\cdot)\right] \end{cases} \tag{10.5(*)}$$

Beweis von (10.5(∗)): Sei $\bar{k}_{\theta_1} = h(k,\theta_1,\theta_0) \in [0,\infty]$. Angenommen $\bar{k}_{\theta_1} = 0$, dann

$$\alpha \overset{(10.1)}{=} P_{\theta_0}[T > k] + \underbrace{\gamma}_{\geq 0} \cdot P_{\theta_0}[T = k] \geq P_{\theta_0}[T > k] = 1 - P_{\theta_0}[T \leq k]$$

$$\begin{aligned}\Rightarrow 0 < 1-\alpha \leq P_{\theta_0}[T \leq k] &\overset{h\uparrow}{\leq} P_{\theta_0}[h(T,\theta_1,\theta_0) \leq \underbrace{h(k,\theta_1,\theta_0)}_{\bar{k}_{\theta_1}=0}] \\ &= P_{\theta_0}[h(T,\theta_1,\theta_0) = 0] = P_{\theta_0}\left[\frac{L(\theta_0,\cdot)}{L(\theta_1,\cdot)} = 0\right] = P_{\theta_0}\left[L(\theta_0,\cdot) = 0\right] \\ &= 0\end{aligned}$$

Widerspruch! Damit $\bar{k}_{\theta_1} \in (0,\infty]$.
Gemäß Definition 10.1 gilt: Es existiert $\bar{F} = \bar{F}(\theta_0,\theta_1) \in \mathcal{F}$ mit $(P_{\theta_0} + P_{\theta_1})(F^c) = 0$ und

$$\forall x \in F : \frac{L(\theta_0,x)}{L(\theta_1,x)} = h(T(x),\theta_1,\theta_0)$$

Sei $x \in \bar{F}$ mit $L(\theta_1,x) > \bar{k}_{\theta_1}^{-1} \cdot L(\theta_0,x)$. Dann

$$h(T(x),\theta_1,\theta_0) = \frac{L(\theta_0,x)}{L(\theta_1,x)} < \bar{k}_{\theta_1} = h(k,\theta_1,\theta_0)$$

also $T(x) < k$ und somit $\varphi(x) = 0$. Analog anderer Fall.
Sei

$$\varphi_0^* := \begin{cases} 1 & x \in \left[L(\theta_1,\cdot) > \frac{1}{\bar{k}_{\theta_1}} \cdot L(\theta_0,\cdot)\right] \\ 0 & x \in \left[L(\theta_1,\cdot) < \frac{1}{\bar{k}_{\theta_1}} \cdot L(\theta_0,\cdot)\right] \end{cases}$$

Da $[1-\varphi \neq \varphi_0^*] \subseteq \bar{F}^c$ nach (10.5(∗)) folgt nochmals mit (10.5(∗))

$$1 - \varphi = \varphi_0^* \; P_{\theta_0}, P_{\theta_1} - \text{fast überall} \tag{+}$$

Ferner ist φ_0^* ein Test vom Neyman-Pearson-Typ (9.1) für $H_0' : \theta = \theta_0$ gegen $H_1' : \theta = \theta_1$ für jedes $\theta_1 < \theta_0$ fest. Außerdem gilt

$$\begin{aligned}\beta_{\varphi_0^*}(\theta_0) = \mathbb{E}_{\theta_0}(\varphi_0^*(X)) &= \int \varphi_0^* \, dP_{\theta_0} \\ &\overset{\pm}{=} \int (1-\varphi)\, dP_{\theta_0} = 1 - \beta_\varphi(\theta_0) = 1-\alpha\end{aligned}$$

also ist φ_0^* ein exakter Niveau-$(1-\alpha)$-Test für H_0' gegen H_1'. Nach 9.4: φ_0^* ist gleichmäßig bester Niveau-$(1-\alpha)$-Test für H_0' gegen H_1'. Insbesondere folgt: Für jeden Test $\tilde{\varphi}$ mit $\beta_{\tilde{\varphi}}(\theta_0) = \alpha$ gilt $\beta_{1-\tilde{\varphi}}(\theta_0) = 1 - \beta_{\tilde{\varphi}}(\theta_0) = 1-\alpha$ und somit

$$\begin{aligned}1 - \beta_\varphi(\theta_1) = \beta_{1-\varphi}(\theta_1) &\overset{\pm}{=} \beta_{\varphi_0^*}(\theta_1) \geq \beta_{1-\tilde{\varphi}}(\theta_1) \\ &= 1 - \beta_{\tilde{\varphi}}(\theta_1) \\ \Leftrightarrow \beta_\varphi(\theta_1) &\leq \beta_{\tilde{\varphi}}(\theta_1)\end{aligned}$$

für alle $\theta_1 < \theta_0$. Damit ist (10.3) gezeigt.

(iii). Sei $\varphi_\alpha(x) := \alpha$ für alle $x \in \mathcal{X}$. Dann $\beta_{\varphi_\alpha}(\theta) = \alpha$. Nach (10.3) gilt $\beta_\varphi(\theta) \leq \beta_{\varphi_\alpha}(\theta) = \alpha$ für alle $\theta < \theta_0$. Also ist mit (10.2) φ Niveau-α-Test für H_0 gegen H_1. Wegen (10.4) ist φ optimal.

10.6 Beispiel: Fortsetzung von 9.5

Seien $X_1,\dots,X_n$ iid mit $X_j \sim \text{Poi}(\lambda)$, $\lambda > 0$. $H_0: \lambda \le \lambda_0$ vs. $H_1: \lambda > \lambda_0$ mit λ_0 vorgegeben. Man zeigt leicht (unter Verwendung von Satz 4.8), dass die zugehörige Verteilungsannahme

$$\mathcal{P} = \{(\text{Poi}(\lambda))^n; \lambda > 0\}$$

eine einparametrige Exponentialfamilie in $q(\lambda) = \log\lambda$ und $T(\underline{x}_n) = \sum_{i=1}^n x_i$ ist. Nach 10.2: $\mathcal{P}$ hat monotone Likelihood-Quotienten in T. Gemäß Faltungsgesetz für die Poisson-Verteilung gilt

$$\mathbb{P}_{\lambda_0} \circ (T(X))^{-1} = \mathbb{P}_{\lambda_0} \circ \left(\sum_{i=1}^n X_i\right)^{-1} = \text{Poi}(n\cdot\lambda_0)$$

Damit folgt, dass $k = G_0^{-1}(1-\alpha) = (1-\alpha)$-Quantil von $\text{Poi}(n\cdot\lambda_0)$. Aus Diskretheit folgt:

$$k = \min\left\{\ell \in \mathbb{N}_0; e^{-n\cdot\lambda_0}\cdot\sum_{j=0}^{\ell}\frac{(n\cdot\lambda_0)^j}{j!} \ge 1-\alpha\right\}$$

$$\gamma = \frac{G_0(k)-(1-\alpha)}{G_0(k)-G_0(k-)} = \frac{e^{-n\cdot\lambda_0}\cdot\sum_{j=0}^k\frac{(n\cdot\lambda_0)^j}{j!}-(1-\alpha)}{e^{-n\cdot\lambda_0}\cdot\frac{(n\cdot\lambda_0)^k}{k!}}$$

$$\overset{10.4}{\Rightarrow} \varphi(\underline{x}_n) = 1_{\{\sum_{i=1}^k x_i > k\}} + \gamma\cdot 1_{\{\sum_{i=1}^k x_i = k\}}$$

Aus 10.4 folgt, dass φ optimal ist.

10.7 Beispiel

Sei $\mathcal{X} = [0,\infty)$ und seien $X_1,\dots,X_n$ iid mit $X_j \sim U(0,\theta)$ mit $\theta > 0$. $H_0: \theta \le \theta_0$ vs. $H_1: \theta > \theta_0$ mit θ_0 vorgegeben,

$$\mathcal{P} = \{(U(0,\theta))^n; \theta > 0\}$$

Behauptung: $\mathcal{P}$ hat monotone Likelihood-Quotienten in $T(\underline{x}_n) = \max\{x_i; i = 1,\dots,n\}$. Dazu:

- Für $\underline{x}_n \in \mathcal{X}^n$:

$$\begin{aligned} L_n(\theta,\underline{x}_n) &= \prod_{i=1}^n \frac{1}{\theta}\cdot 1_{[0,\theta]}(x_i) = \theta^{-n}\cdot\prod_{i=1}^n 1_{\{x_i\le\theta\}} \\ &= \theta^{-n}1_{[0,\theta]}\left(\max_{1\le i\le n} x_i\right) \end{aligned}$$

- Es folgt für alle $\theta < \theta'$:

$$\frac{L_n(\theta',\underline{x}_n)}{L_n(\theta,\underline{x}_n)} = \begin{cases} \left(\frac{\theta}{\theta'}\right)^n & \max\limits_{1\le i\le n} \le \theta \\ \infty & \theta < \max\limits_{1\le i\le n} \le \theta' \\ \text{nicht definiert} & \max\limits_{1\le i\le n} x_i > \theta' \end{cases}$$

- Sei $F := F(\theta') := \{\underline{x}_n \in \mathcal{X}^n; \max_{1\le i\le n} x_i > \theta'\}$. Es folgt $P_\theta(F) = P_{\theta'}(F) = 0$, denn

$$\begin{aligned} P_{\theta'}(F) &= \mathbb{P}_{\theta'}\left[\max_{1\le i\le n} X_i > \theta'\right] = \mathbb{P}_{\theta'}\left(\bigcup_{i=1}^n [X_i > \theta']\right) \\ &\le \sum_{i=1}^n \mathbb{P}_{\theta'}[X_i > \theta'] = \sum_{i=1}^n (1 - \underbrace{\mathbb{P}_{\theta'}[X_i \le \theta']}_{1}) = 0 \end{aligned}$$

  Analog: $P_\theta(F) = 0$.

- Sei nun
$$h(y,\theta,\theta') := \begin{cases} \left(\dfrac{\theta}{\theta'}\right)^n & y \le \theta \\ \infty & y > \theta \end{cases}$$
Dann
$$\forall \underline{x}_n \notin F : \frac{L_n(\theta', \underline{x}_n)}{L_n(\theta, \underline{x}_n)} = h\left(\max_{1\le i\le n} x_i, \theta, \theta'\right)$$
Klar: $h(y,\theta,\theta')$ ist monoton wachsend in y.

Man erhält für $t \in \mathbb{R}$:

$$\begin{aligned} G_0(t) &:= \mathbb{P}_{\theta_0}\left[T(\underline{X}_n) \le t\right] = \mathbb{P}_{\theta_0}\left[\max_{1\le i\le n} X_i \le t\right] \\ &= \mathbb{P}_{\theta_0}\left(\bigcap_{i=1}^{n}[X_i \le t]\right) = \prod_{i=1}^{n} \mathbb{P}_{\theta_0}[X_i \le t] \\ &= \prod_{i=1}^{n}\left(\frac{t}{\theta_0}\cdot 1_{[0,\theta_0]}(t) + 1_{(\theta_0,\infty)}(t)\right) = \left(\frac{t}{\theta_0}\right)^n \cdot 1_{[0,\theta_0]}(t) + 1_{(\theta_0,\infty)}(t) \end{aligned}$$

Somit

$$\begin{aligned} k = G_0^{-1}(1-\alpha) \Leftrightarrow G_0(k) = 1-\alpha \Leftrightarrow \left(\frac{k}{\theta_0}\right)^n = 1-\alpha \\ \Leftrightarrow k = \theta_0 \cdot (1-\alpha)^{\frac{1}{n}} \end{aligned}$$

Da G_0 stetig auf $\mathbb{R}$, kann γ beliebig gewählt werden, zum Beispiel $\gamma = 0$. Wegen 10.4(iii) ist

$$\varphi(\underline{x}_n) = 1_{\left\{\max_{1\le i\le n} x_i > \theta_0\cdot(1-\alpha)^{\frac{1}{n}}\right\}}$$

optimal.