

# Mathematische Statistik

Vorlesungsmitschrieb zur Vorlesung  
„Mathematische Statistik (Statistik I)“  
Dr. Klar / Universität Karlsruhe / Sommersemester 2007 <sup>1</sup>

geT<sub>E</sub>Xed von  
Tobias Baust und Tobias Flaig

Stand: 7. September 2017

<sup>1</sup>Dieser inoffizielle Mitschrieb der Vorlesung wurde mit ausdrücklicher Genehmigung von Herrn Dr. Klar auf <http://mitschriebwiki.nomeata.de/> veröffentlicht, *Herr Dr. Klar ist für den Inhalt dieser Seiten nicht verantwortlich.* Der Mitschrieb erhebt keinen Anspruch auf Vollständigkeit und Richtigkeit!

## Inhaltsverzeichnis

1	Grundbegriffe, Motivation	1
2	Klassische statistische Verfahren unter Normalverteilungs-Annahme	7
3	Schätzer und ihre Eigenschaften	15
4	Schätzmethoden	21
5	Optimale erwartungstreue Schätzer	33
6	Exponentialfamilien	43
7	Suffizienz und Vollständigkeit	51
8	Asymptotik von Schätzfehlern	61
9	Robuste Schätzer	67
10	Grundbegriffe der Testtheorie	77
11	Neyman-Pearson-Tests (NP-Tests)	80
12	UMPU Tests („UMP unbiased“)	89
13	Konfidenzbereiche	99
14	Lineare statistische Modelle	107

# 1 Grundbegriffe, Motivation

Sei  $(\Omega, \mathcal{A}, \mathbb{P})$  Wahrscheinlichkeitsraum,  $(\mathfrak{X}, \mathcal{B})$  messbarer Raum<sup>1</sup> (sogenannter Stichprobenraum).

$X : \Omega \rightarrow \mathfrak{X}$  Zufallsvariable

$$P(B) := P^X(B) := \mathbb{P}(X^{-1}(B)), \quad B \in \mathcal{B}$$

Verteilung von X ( $\leftrightarrow$  Wahrscheinlichkeitsraum  $(\mathfrak{X}, \mathcal{B}, P)$ )

Statistischer Entscheidung liegt Datenmaterial (Beobachtung)  $x \in \mathfrak{X}$  zugrunde.

Grundannahme:

- 1)  $x = X(\omega)$  für ein  $\omega \in \Omega$ , d.h. x ist Realisierung von X
- 2) P ist (teilweise) unbekannt

Ziel: Aufgrund von x Aussagen über P machen!

Sei  $\mathcal{M}^1(\mathfrak{X}, \mathcal{B}) := \{P : P \text{ ist Wahrscheinlichkeitsmaß auf } \mathcal{B}\}$ .

## 1.1 Definition

Eine Verteilungsannahme ist eine Teilmenge  $\wp \subset \mathcal{M}^1(\mathfrak{X}, \mathcal{B})$ .

Das Tripel  $(\mathfrak{X}, \mathcal{B}, \wp)$  heißt statistischer Raum (statistisches Modell).

## 1.2 Beispiel

$$(\mathfrak{X}, \mathcal{B}) = (\mathbb{R}^n, \mathcal{B}^n)$$

$$\wp := \{P : \exists \text{ Wahrscheinlichkeitsmaß } Q \text{ auf } \mathcal{B}^1 \text{ mit } P = \underbrace{Q \otimes \dots \otimes Q}_{n \text{ Faktoren}}\}$$

Mit anderen Worten  $X = (X_1, \dots, X_n)$ ,  $X_1, \dots, X_n$  stochastisch unabhängig mit gleicher Verteilung Q,  $X_1, \dots, X_n \stackrel{\text{i.i.v.}}{\sim} Q$ .

<sup>1</sup> $\mathcal{B}$  steht hier für eine beliebige  $\sigma$ -Algebra, die Borelsche  $\sigma$ -Algebra wird mit  $\mathcal{B}^d$  bezeichnet, wobei d die Dimension angibt

### 1.3 Beispiel

$$(\mathfrak{X}, \mathcal{B}) = (\mathbb{R}^n, \mathcal{B}^n)$$

$$\wp := \{P : \exists (\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}_{>0} \text{ mit } P = \mathcal{N}(\mu, \sigma^2) \otimes \dots \otimes \mathcal{N}(\mu, \sigma^2)\}$$

Also  $X_1, \dots, X_n \stackrel{\text{uiv}}{\sim} \mathcal{N}(\mu, \sigma^2)$ .

Ein-Stichproben-Normalverteilungs-Annahme

### 1.4 Beispiel

$$(\mathfrak{X}, \mathcal{B}) = (\mathbb{R}^{m+n}, \mathcal{B}^{m+n})$$

$$\wp := \{P : \exists \text{ W'ma\ss e } Q_1, Q_2 : P = \underbrace{Q_1 \otimes \dots \otimes Q_1}_{m \text{ Faktoren}} \otimes \underbrace{Q_2 \otimes \dots \otimes Q_2}_{n \text{ Faktoren}}\}$$

Also  $X = (X_1, \dots, X_m, Y_1, \dots, Y_n)$ ,  $X_1, \dots, X_m, Y_1, \dots, Y_n$  unabhangig,  
 $X_1, \dots, X_m \stackrel{\text{uiv}}{\sim} Q_1, Y_1, \dots, Y_n \stackrel{\text{uiv}}{\sim} Q_2$ .

### 1.5 Beispiel

$(\mathfrak{X}, \mathcal{B})$  wie in 1.4

$$\wp := \{P : \exists (\mu, \nu, \sigma^2) \in \mathbb{R} \times \mathbb{R} \times \mathbb{R}_{>0} : P = \mathcal{N}(\mu, \sigma^2) \otimes \dots \otimes \mathcal{N}(\mu, \sigma^2) \otimes \mathcal{N}(\nu, \sigma^2) \otimes \dots \otimes \mathcal{N}(\nu, \sigma^2)\}$$

$X_1, \dots, X_m, Y_1, \dots, Y_n$  unabhangig

$X_i \stackrel{\text{uiv}}{\sim} \mathcal{N}(\mu, \sigma^2), Y_j \stackrel{\text{uiv}}{\sim} \mathcal{N}(\nu, \sigma^2)$

2 unabhangige normalverteilte Stichproben mit gleicher Varianz

### 1.6 Definition

Eine **Parametrisierung** von  $\wp \subset \mathcal{M}^1(\mathfrak{X}, \mathcal{B})$  ist eine bijektive Abbildung  
 $\Theta \ni \vartheta \rightarrow P_\vartheta \in \wp$ .

Ist  $X$  eine Zufallsvariable mit Verteilung  $P_\vartheta$ , so schreibt man auch

$$\left. \begin{array}{l} E_\vartheta(X) \\ \text{Var}_\vartheta(X) \\ (*) F_\vartheta(t) := P_\vartheta(X \leq t) = P_\vartheta((-\infty, t]) \end{array} \right\} \text{ falls } X \text{ reellwertig}$$

$$(**) P_{\vartheta}(B) = P_{\vartheta}(X \in B), \quad B \in \mathcal{B}$$

Schreibweisen (\*), (\*\*) unterstellen

$$(\Omega, \mathcal{A}, \mathbb{P}) = (\mathfrak{X}, \mathcal{B}, P), \quad X = \text{id}_{\Omega}$$

[eigentlich:  $P_{\vartheta}(B) := \mathbb{P}(X^{-1}(B)) = \mathbb{P}(X \in B)$ ]

## 1.7 Definition

Eine Verteilungsklasse  $\wp = \{P_{\vartheta} : \vartheta \in \Theta\}$  heißt  $\vartheta$ -parametrisch, wenn sie sich „zwanglos“ durch einen Parameterraum  $\Theta \subset \mathbb{R}^k$  parametrisieren lässt. Ist  $\vartheta = (\vartheta_1, \vartheta_2)$  und interessiert nur  $\vartheta_1$ , so heißt  $\vartheta_1$  Hauptparameter und  $\vartheta_2$  Nebenparameter oder Störparameter.

## 1.8 Beispiele

- a) In Beispiel 1.3:  
2-parametrische Verteilungsannahme, wobei  $\vartheta = (\mu, \sigma^2)$ ,  $\Theta = \mathbb{R} \times \mathbb{R}_{>0}$ .
- b) In Beispiel 1.5:  
3-parametrisch,  $\vartheta = (\mu, \nu, \sigma^2)$ ,  $\Theta = \mathbb{R} \times \mathbb{R} \times \mathbb{R}_{>0}$   
Hier meistens:  $(\mu, \nu)$  Hauptparameter

Häufig interessiert von  $\wp$  der Wert eines reellwertigen Funktionals  $\gamma : \wp \rightarrow \mathbb{R}$  anstelle von  $P$ , z.B. (falls  $P$  Wahrscheinlichkeitsmaß auf  $\mathcal{B}^1$ )

$$\gamma(P) := \int_{\mathbb{R}} x dP(x)$$

(Erwartungswert von  $X$ )

Falls  $\wp = \{P_{\vartheta} : \vartheta \in \Theta\}$ , so schreibt man auch  $\gamma(\vartheta) := \gamma(P_{\vartheta})$ , fasst also  $\gamma$  als Abbildung  $\gamma : \Theta \rightarrow \mathbb{R}$  auf.

### Problem:

„Enge“ Verteilungsannahme täuscht oft nicht vorhandene Genauigkeit vor.  
 $\wp$  sollte das wahre  $P$  enthalten. (realistisch?)

Bei diskreten Zufallsvariablen ergibt sich  $\wp$  manchmal zwangsläufig; bei stetigen Zufallsvariablen ist  $\wp$  häufig nicht vorgezeichnet.

## 1.9 Typische Fragestellungen der Statistik

- a) Punktschätzung  
Schätze aufgrund von  $x \in \mathfrak{X}$  den Wert  $\gamma(\vartheta) \in \mathbb{R}$  möglichst „gut“.
- b) Konfidenzbereiche  
Konstruiere „möglichst kleinen“, von  $x$  abhängigen Bereich, der  $\gamma(\vartheta)$  mit „großer Wahrscheinlichkeit“ enthält.
- c) Testprobleme  
Es sei  $\Theta = \Theta_0 + \Theta_1$  eine Zerlegung von  $\Theta$ . Teste die Hypothese  $H_0 : \vartheta \in \Theta_0$  gegen die Alternative  $H_1 : \vartheta \in \Theta_1$ .

## 1.10 Asymptotische Betrachtungen

Häufig liegt Folge  $(X_j)_{j \in \mathbb{N}}$  unabhängiger Zufallsvariablen zugrunde (alle auf nicht interessierenden Wahrscheinlichkeitsräume  $(\Omega, \mathcal{A}, \mathbb{P})$  definiert) mit Werten in einem Messraum  $(\mathfrak{X}_0, \mathcal{B}_0)$ .

Häufig:  $P^{X_j} = P \ \forall j$  (identische Verteilung)

Unter der Verteilungsannahme  $P \in \wp_0 \subset \mathcal{M}^1(\mathfrak{X}_0, \mathcal{B}_0)$  nimmt dann die Folge  $(X_j)_{j \in \mathbb{N}}$  Werte im statistischen Raum

$$(\mathfrak{X}, \mathcal{B}, \wp) := \left( \times_{j=1}^{\infty} \mathfrak{X}_0, \bigotimes_{j=1}^{\infty} \mathcal{B}_0, \left\{ \bigotimes_{j=1}^{\infty} P : P \in \wp_0 \right\} \right)$$

an. Also:  $X_1, X_2, \dots$  unabhängig,  $\mathfrak{X}_0$ -wertig mit gleicher Verteilung  $P \in \wp_0$

## 1.11 Statistiken

Es seien  $(\mathfrak{X}, \mathcal{B})$  Stichprobenraum und  $(\mathcal{T}, \mathcal{D})$  Messraum. Eine messbare Abbildung  $T : \mathfrak{X} \rightarrow \mathcal{T}$  heißt Statistik (Stichprobenfunktion).

Häufig:  $(\mathcal{T}, \mathcal{D}) = (\mathbb{R}, \mathcal{B}^1)$ .

Wichtigstes Beispiel:

$$\mathfrak{X} = \mathbb{R}^n, \mathcal{T} = \mathbb{R}$$

$$T(x_1, \dots, x_n) = \frac{1}{n} \sum_{i=1}^n x_i$$

Stichproben-Funktionen bewirken eine **Datenkompression**.

Statistische Entscheidungen wie Ablehnung von Hypothesen hängen von  $x \in \mathfrak{X}$  im Allgemeinen durch den Wert  $T(x)$  einer geeigneten Statistik ab.

Bei Tests: Statistik  $\hat{=}$  Testgröße  $\hat{=}$  Prüfgröße

Sind  $X_1, \dots, X_n \stackrel{iid}{\sim} P$ , so nennt man  $X_1, \dots, X_n$  eine Stichprobe vom Umfang  $n$  aus der Verteilung  $P$ .

Ist  $T(X_1, \dots, X_n)$  eine mit  $X_1, \dots, X_n$  operierende Statistik, so schreib man auch  $T_n := T_n(X_1, \dots, X_n) := T(X_1, \dots, X_n)$ .

Insbesondere bei asymptotischen Betrachtungen ist  $(T_n)_{n \geq 1}$  dann eine Folge von Statistiken.



## 2 Klassische statistische Verfahren unter Normalverteilungs-Annahme

### 2.1 $\chi_s^2, t_s, F_{r,s}$ -Verteilung

- a)  $N_1, \dots, N_s \stackrel{uiv}{\sim} \mathcal{N}(0, 1)$   
 Verteilung von  $Y := N_1^2 + \dots + N_s^2$  heißt Chi-Quadrat-Verteilung mit  $s$  Freiheitsgraden ( $\chi_s^2$ -Verteilung).  
 $Y$  besitzt die Dichte

$$f(y) = \frac{1}{2^{\frac{s}{2}} \Gamma(\frac{s}{2})} e^{-\frac{y}{2}} y^{\frac{s}{2}-1}, \quad y > 0$$

Dabei:

$$\Gamma(t) = \int_0^\infty e^{-x} x^{t-1} dx, \quad t > 0$$

(Gamma-Funktion)

$$[\Gamma(x+1) = x\Gamma(x) \quad (x > 0), \quad \Gamma(n+1) = n! \quad (n \in \mathbb{N}), \quad \Gamma(\frac{1}{2}) = \sqrt{\pi}]$$

Es gilt:

$$EY = s, \quad \text{Var}(Y) = 2s$$

- b) Seien  $N_0, N_1, \dots, N_s \stackrel{uiv}{\sim} \mathcal{N}(0, 1)$ .  
 Die Verteilung von

$$y := \frac{N_0}{\sqrt{\sum_{j=1}^s N_j^2 / s}}$$

heißt (studentsche) t-Verteilung mit  $s$  Freiheitsgraden ( $t_s$ -Verteilung).  
 Dichte:

$$f(y) = \frac{1}{\sqrt{\pi s}} \frac{\Gamma(\frac{s+1}{2})}{\Gamma(\frac{s}{2})} \left(1 + \frac{y^2}{s}\right)^{-\frac{s+1}{2}}, \quad y \in \mathbb{R}$$

$$E(Y) = 0 \quad (s \geq 2), \quad \text{Var}(Y) = \frac{s}{s-2} \quad (s \geq 3)$$

$s=1^2$ :

$$f(y) = \frac{1}{\sqrt{\pi}} \frac{1}{\sqrt{\pi}} \frac{1}{1+y^2} = \frac{1}{\pi(1+y^2)}, \quad y \in \mathbb{R}$$

- c) Es seien  $R, S$  unabhängig,  $R \sim \chi_r^2$ ,  $S \sim \chi_s^2$ . Die Verteilung von

$$Y := \frac{\frac{1}{r}R}{\frac{1}{s}S}$$

---

<sup>2</sup>Cauchy-Verteilung

heißt F(isher)-Verteilung mit  $r$  Zähler- und  $s$  Nenner-Freiheitsgraden (kurz:  $Y \sim F_{r,s}$ ).

Dichte:

$$f(y) = \frac{\Gamma(\frac{r+s}{2}) (\frac{r}{s})^{\frac{r}{2}} y^{\frac{r}{2}-1}}{\Gamma(\frac{r}{2}) \Gamma(\frac{s}{2}) (1 + \frac{r}{s}y)^{\frac{r+s}{2}}}, \quad y > 0$$

Es gilt:

$$E(Y) = \frac{s}{s-2}, \quad s > 2, \quad \text{Var}(Y) = \frac{2s^2(r+s-2)}{r(s-2)^2(s-4)}, \quad s > 4$$

## 2.2 Satz

Es seien  $X_1, \dots, X_n \stackrel{uiv}{\sim} \mathcal{N}(\mu, \sigma^2)$ .

$$\bar{X}_n := \frac{1}{n} \sum_{i=1}^n X_i, \quad \hat{\sigma}_n^2 := \frac{1}{n} \sum_{j=1}^n (X_j - \bar{X}_n)^2$$

Dann gilt:

- a)  $\bar{X}_n \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$ .
- b)  $\frac{n}{\sigma^2} \hat{\sigma}_n^2 \sim \chi_{n-1}^2$
- c)  $\bar{X}_n$  und  $\hat{\sigma}_n^2$  sind unabhängig.

### Beweis

a) klar

b),c) Sei  $Z_j := X_j - \mu$ ,  $Z := (Z_1, \dots, Z_n)^T$ .

Es gilt:  $Z \sim \mathcal{N}_n(0, \sigma^2 \cdot I_n)$ .

Sei  $H = (h_{ij})$  orthogonale  $n \times n$ -Matrix mit  $h_{nj} = \frac{1}{\sqrt{n}}$ ,  $1 \leq j \leq n$ .

$$Y = (Y_1, \dots, Y_n)^T := HZ$$

$Y \sim \mathcal{N}_n(0, \sigma^2 H H^T) = \mathcal{N}(0, \sigma^2 I_n)$ , d.h.  $Y_1, \dots, Y_n \stackrel{uiv}{\sim} \mathcal{N}(0, \sigma^2)$ .

Es gilt:

$$\sum_{j=1}^n Y_j^2 = \|Y\|^2 = \|Z\|^2 = \sum_{j=1}^n Z_j^2$$

$$Y_n = \frac{1}{\sqrt{n}} \sum_{j=1}^n Z_j = \sqrt{n}(\bar{X}_n - \mu)$$

Mit  $\bar{Z}_n := \frac{1}{n} \sum_{j=1}^n Z_j$  folgt:

$$\begin{aligned} \sum_{i=1}^n (X_j - \bar{X}_n)^2 &= \sum_{j=1}^n (Z_j - \bar{Z}_n)^2 = \sum_{j=1}^n Z_j^2 - n\bar{Z}_n^2 \\ &= \sum_{j=1}^n Y_j^2 - Y_n^2 \\ &= \sum_{j=1}^{n-1} Y_j^2 \\ &= \sigma^2 \underbrace{\sum_{j=1}^{n-1} \left( \frac{Y_j}{\sigma} \right)^2}_{\substack{\sim \mathcal{N}(0,1) \\ \sim \chi_{n-1}^2}} \end{aligned}$$

$\Rightarrow$  b)

Wegen  $\sum_{j=1}^n (X_j - \bar{X}_n)^2 =$  Funktion von  $Y_1, \dots, Y_{n-1}$  und  $\bar{X}_n =$  Funktion von  $Y_n$  sind  $\hat{\sigma}_n^2$  und  $\bar{X}_n$  unabhängig (Blockungslemma). ■

## 2.3 Korollar

In der Situation von 2.2 sei<sup>3</sup>

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

Dann gilt:

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{S_n} \sim t_{n-1}$$

Beweis:

$$\sqrt{n} \frac{\bar{X}_n - \mu}{S_n} = \frac{\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma}}{\sqrt{\frac{1}{\sigma^2} \cdot \underbrace{\sum_i (X_i - \bar{X}_n)^2}_{\sim \chi_{n-1}^2} / n - 1}}$$

Zähler und Nenner sind unabhängig und der Zähler ist  $\mathcal{N}(0, 1)$ -verteilt.

<sup>2.1b)</sup>  $\Rightarrow$  Behauptung

---

<sup>3</sup>Stichprobenvarianz

## 2.4 Korollar

Sei  $0 < \alpha \leq 1$  und sei  $t_{n-1, 1-\frac{\alpha}{2}}$  das  $(1 - \frac{\alpha}{2})$ -Quantil der  $t_{n-1}$ -Verteilung.<sup>4</sup>  
Dann gilt:

$$P_{\mu, \sigma^2} \left( \left| \frac{\sqrt{n}(\bar{X}_n - \mu)}{S_n} \right| \leq t_{n-1, 1-\frac{\alpha}{2}} \right) = 1 - \alpha$$

Oder äquivalent:

$$P_{\mu, \sigma^2} \left( \bar{X}_n - \frac{S_n}{\sqrt{n}} t_{n-1, 1-\frac{\alpha}{2}} \leq \mu \leq \bar{X}_n + \frac{S_n}{\sqrt{n}} t_{n-1, 1-\frac{\alpha}{2}} \right) = 1 - \alpha$$

Sprechweise:

$[\bar{X}_n - \frac{S_n}{\sqrt{n}} t_{n-1, 1-\frac{\alpha}{2}}, \bar{X}_n + \frac{S_n}{\sqrt{n}} t_{n-1, 1-\frac{\alpha}{2}}]$  ist ein Konfidenzintervall (Vertrauensintervall) zur Konfidenzwahrscheinlichkeit (Vertrauenswahrscheinlichkeit)  $1 - \alpha$  für  $\mu$ . (Störparameter  $\sigma^2$  tritt hier nicht auf!)

## 2.5 Ein-Stichproben-t-Test (1-SP-t-Test)

Seien  $X_1, \dots, X_n \stackrel{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$ ,  $\vartheta = (\mu, \sigma^2)$ ,  $\Theta := \mathbb{R} \times \mathbb{R}_{>0}$ .

Zu testen sei  $H_0 : \mu = \mu_0$  gegen  $H_1 : \mu \neq \mu_0$  ( $\mu_0$  gegebener Wert).

$$H_0/H_1 \leftrightarrow \Theta_0/\Theta_1$$

$$\Theta_0 := \{\vartheta = (\mu, \sigma^2) \in \Theta : \mu = \mu_0\}$$

$$\Theta_1 := \{\vartheta = (\mu, \sigma^2) \in \Theta : \mu \neq \mu_0\}$$

Sei  $(x_1, \dots, x_n) =: x$  Realisierung von  $X_1, \dots, X_n$ .

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

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

$$T(x_1, \dots, x_n) := \frac{\sqrt{n}(\bar{x}_n - \mu_0)}{s_n}$$

<sup>4</sup>→ Abbildung 2.1

(Zweiseitiger) 1-SP-t-Test zum Niveau  $\alpha$ :

$H_0$  ablehnen, falls

$$|T(x_1, \dots, x_n)| \geq t_{n-1, 1-\frac{\alpha}{2}}$$

Kein Widerspruch zu  $H_0$ , falls

$$|T(x_1, \dots, x_n)| < t_{n-1, 1-\frac{\alpha}{2}}$$

Sei  $\vartheta \in \Theta_0$ , also  $\mu = \mu_0$ .

$\Rightarrow T(X_1, \dots, X_n) \sim t_{n-1}$

$$\Rightarrow P(|T(X_1, \dots, X_n)| \geq t_{n-1, 1-\frac{\alpha}{2}}) = \alpha \quad \forall \vartheta \in \Theta_0$$

[Das bedeutet, wenn  $H_0$  gilt, ist die Wahrscheinlichkeit  $H_0$  trotzdem abzulehnen  $\alpha$ . ( $\rightarrow$  Niveau)]

Bemerkungen:

1) Entsprechend einseitige Tests<sup>5</sup>, z.B.  $H_0 : \mu \geq \mu_0$  gegen  $H_1 : \mu < \mu_0$ .  
 $H_0$  ablehnen, falls  $|T(x_1, \dots, x_n)| \geq t_{n-1, \alpha}$ <sup>6</sup>

2) Sei

$$f(x, \vartheta) := \prod_{j=1}^n f_1(x_j, \vartheta) = \left( \frac{1}{\sigma\sqrt{2\pi}} \right)^n \exp\left(-\frac{1}{2\sigma^2} \sum_{j=1}^n (x_j - \mu)^2\right)$$

Die Prüfgröße  $T(x_1, \dots, x_n)$  des t-Tests ergibt sich aus einem allgemeinen „Rezept“:

Bilde den (verallgemeinerten) Likelihood-Quotienten

$$\lambda(x) := \frac{\sup_{\vartheta \in \Theta_0} f(x, \vartheta)}{\sup_{\vartheta \in \Theta} f(x, \vartheta)}$$

und lehne  $H_0 : \vartheta \in \Theta_0$  für zu „kleine“ Werte von  $\lambda(x)$  ab.

Es gilt:

$$(n-1)(\lambda(x)^{-\frac{2}{n}} - 1) = T^2(x_1, \dots, x_n)$$

(Blatt 2, Aufgabe 6)

<sup>5</sup>vgl. Stochstik I ( $\rightarrow$  Wahl der Nullhypothese)

<sup>6</sup>etc.; Niveau  $\alpha$

## 2.6 Satz

Seien  $X_1, \dots, X_m, Y_1, \dots, Y_n$  unabhängig.

$$X_i \sim \mathcal{N}(\mu, \sigma^2) \quad \forall i, \quad Y_j \sim \mathcal{N}(\nu, \sigma^2) \quad \forall j$$

$$\bar{X}_m := \frac{1}{m} \sum_{i=1}^m X_i, \quad \bar{Y}_n := \frac{1}{n} \sum_{j=1}^n Y_j$$

Dann gilt:

$$\bar{X}_m \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{m}\right), \quad \bar{Y}_n \sim \mathcal{N}\left(\nu, \frac{\sigma^2}{n}\right)$$

$$\bar{X}_m - \bar{Y}_n - (\mu - \nu) \sim \mathcal{N}\left(0, \frac{\sigma^2}{m} + \frac{\sigma^2}{n}\right) = \mathcal{N}\left(0, \frac{m+n}{mn} \sigma^2\right)$$

$$\frac{1}{\sigma^2} \sum_{i=1}^m (X_i - \bar{X}_m)^2 \sim \chi_{m-1}^2$$

$$\frac{1}{\sigma^2} \sum_{j=1}^n (Y_j - \bar{Y}_n)^2 \sim \chi_{n-1}^2$$

$\bar{X}_m, \bar{Y}_n$  und die letzten beiden Größen sind stochastisch unabhängig!

Sei

$$S_{m,n}^2 := \frac{1}{m+n-2} \left( \sum_{i=1}^m (X_i - \bar{X}_m)^2 + \sum_{j=1}^n (Y_j - \bar{Y}_n)^2 \right)$$

Dann gilt weiter:

$$(m+n-2) \cdot S_{m,n}^2 / \sigma^2 \sim \chi_{m+n-2}^2$$

## 2.7 Korollar

In der Situation von 2.6 gilt:

$$\frac{\sqrt{\frac{mn}{m+n}} (\bar{X}_m - \bar{Y}_n - (\mu - \nu))}{S_{m,n}} \sim t_{m+n-2}$$

Beweis:

Wie Korollar 2.3.

## 2.8 Korollar

$$P_{\mu, \nu, \sigma^2} \left( \sqrt{\frac{mn}{m+n}} \frac{|\bar{X}_m - \bar{Y}_n - (\mu - \nu)|}{S_{m,n}} \leq t_{m+n-2, 1-\frac{\alpha}{2}} \right) = 1 - \alpha$$

D.h.

$$\bar{X}_m - \bar{Y}_n \pm \frac{S_{m,n}}{\sqrt{\frac{mn}{m+n}}} t_{m+n-2, 1-\frac{\alpha}{2}}$$

ist Konfidenzintervall für  $\mu - \nu$  zur Konfidenzwahrscheinlichkeit  $1 - \alpha$ .

## 2.9 (Zweiseitiger) 2-SP-t-Test

Situation von 2.6.

$$H_0 : \mu = \nu \ (\mu - \nu = 0), \quad H_1 : \mu \neq \nu \ (\mu - \nu \neq 0)$$

Mit

$$\Theta := \{\vartheta = (\mu, \nu, \sigma^2) : -\infty < \mu, \nu < \infty, \sigma^2 > 0\}$$

$$\Theta_0 := \{(\mu, \nu, \sigma^2) \in \Theta : \mu = \nu\}$$

$$\Theta_1 := \{(\mu, \nu, \sigma^2) \in \Theta : \mu \neq \nu\}$$

gilt:  $H_0 \hat{=} \vartheta \in \Theta_0, \quad H_1 \hat{=} \vartheta \in \Theta_1$ .

Prüfgröße:

$$T_{m,n}(x_1, \dots, x_m, y_1, \dots, y_n) := \sqrt{\frac{mn}{m+n}} \frac{\bar{x}_m - \bar{y}_n}{s_{m,n}}$$

Testentscheidung:

$H_0$  ablehnen, falls  $|T_{m,n}| \geq t_{m+n-2, 1-\frac{\alpha}{2}}$ .

Kein Widerspruch zu  $H_0$ , falls  $|T_{m,n}| < t_{m+n-2, 1-\frac{\alpha}{2}}$ .

Es gilt:

$$P(|T_{m,n}(X_1, \dots, X_m, Y_1, \dots, Y_n)| \geq t_{m+n-2, 1-\frac{\alpha}{2}}) = \alpha \quad \forall \vartheta \in \Theta_0$$

D.h. Test hat Niveau  $\alpha$ .

### 2.10 F-Test für den Varianz-Quotienten

Situation wie in 2.6, aber  $Y_j \sim \mathcal{N}(\nu, \tau^2)$  ( $\tau^2 \neq \sigma^2$  möglich).

Zu testen:  $H_0 : \sigma^2 = \tau^2$  ( $\frac{\sigma^2}{\tau^2} = 1$ ) gegen  $H_1 : \sigma^2 \neq \tau^2$  ( $\frac{\sigma^2}{\tau^2} \neq 1$ ).

Prüfgröße des F-Tests für Varianzquotienten ist

$$Q_{m,n} = \frac{\frac{1}{m-1} \sum_{i=1}^m (X_i - \bar{X}_m)^2}{\frac{1}{n-1} \sum_{j=1}^n (Y_j - \bar{Y}_n)^2}$$

Unter  $H_0$  gilt  $Q_{m,n} \sim F_{m-1, n-1}$ .

Ablehnung von  $H_0$  erfolgt für große und kleine Werte von  $Q_{m,n}$

[Meist<sup>7</sup>: Ablehnung für  $Q_{m,n} \geq F_{m-1, n-1, 1-\frac{\alpha}{2}}$ ,  $Q_{m,n} \leq F_{m-1, n-1, \frac{\alpha}{2}}$ ]

<sup>7</sup>→ Abbildung 2.2

### 3 Schätzer und ihre Eigenschaften

Es seien  $(\mathfrak{X}, \mathcal{B}, \{P_\vartheta : \vartheta \in \Theta\})$  ein statistischer Raum,  $\gamma : \Theta \rightarrow \Gamma$  ein Funktional, wobei  $\Gamma \supset \gamma(\Theta)$ ,  $A_\Gamma$  eine  $\sigma$ -Algebra auf  $\Gamma$ .

#### 3.1 Definition

Ein **Schätzer** für  $\gamma(\vartheta)$  ist eine messbare Abbildung  $S : (\mathfrak{X}, \mathcal{B}) \rightarrow (\Gamma, A_\Gamma)$ .  $S(x)$  heißt **Schätzwert** für  $\gamma(\vartheta)$  zur Beobachtung  $x \in \mathfrak{X}$ .

#### 3.2 Beispiel

$(\mathfrak{X}, \mathcal{B}) = (\{0, 1\}^n, \mathcal{P}(\{0, 1\}^n))$ ,  $\Theta = (0, 1)$ ,  $P_\vartheta = \bigotimes_{j=1}^n \text{Bin}(1, \vartheta)$

$$\gamma(\vartheta) = \vartheta$$

$$S(x_1, \dots, x_n) := \frac{1}{n} \sum_{j=1}^n X_j$$

(relative Trefferhäufigkeit)

$$\Gamma = [0, 1] = \bar{\Theta}, \quad A_\Gamma = \mathcal{B}^1 \cap [0, 1]$$

[Beachte:  $\gamma(\Theta) = \Theta \subset \Gamma$ ]

Die Güte eines Schätzers wird über die Verteilung  $P_\vartheta^{S(X)}$  von  $S(X)$  unter  $\vartheta$  beurteilt. Für jedes  $\vartheta \in \Theta$  sollte  $P_\vartheta^{S(x)}$  „stark um  $\gamma(\vartheta)$  konzentriert“ sein.

#### 3.3 Definition (Sei $\Gamma \subset \mathbb{R}^k$ .)

- S erwartungstreu** (unbiased) für  $\gamma(\vartheta)$  :  $\Leftrightarrow E_\vartheta S(X) = \gamma(\vartheta) \forall \vartheta \in \Theta$
- $b_S(\vartheta) := E_\vartheta S(X) - \gamma(\vartheta)$  heißt Verzerrung (bias) von S an der Stelle  $\vartheta$ .
- Ist  $S_n = S_n(X_1, \dots, X_n)$ ,  $n \geq 1$  eine Schätzfolge, so heißt  $(S_n)$  **asymptotisch erwartungstreu** für  $\gamma(\vartheta)$  :  $\Leftrightarrow$

$$\lim_{n \rightarrow \infty} E_\vartheta S_n = \gamma(\vartheta) \forall \vartheta \in \Theta$$

Erwartungstreue:  $\forall \vartheta \in \Theta$ : Schwerpunkt von  $P_\vartheta^{S(X)}$  ist  $\gamma(\vartheta)$

### 3.4 Definition (Sei $\Gamma \subset \mathbb{R}$ .)

S mediantreu für  $\gamma(\vartheta) : \Leftrightarrow \text{med}_\vartheta S(X) = \gamma(\vartheta) \forall \vartheta \in \Theta$ .

Dabei:

Sei Y Zufallsvariable mit Verteilungsfunktion F.

$$F^{-1}(q) := \inf\{x \in \mathbb{R} : F(x) \geq q\}, \quad 0 < q < 1$$

$$\text{med } Y := \text{med } F := \frac{1}{2} \left( F^{-1}\left(\frac{1}{2}\right) + \underbrace{F^{-1}\left(\frac{1}{2} + 0\right)}_{=\lim_{x \rightarrow \frac{1}{2}^+} F^{-1}(x)} \right)$$

→ Median<sup>8</sup>

Mediantreue:  $\forall \vartheta \in \Theta$ :

$$P_\vartheta(S(X) \leq \gamma(\vartheta)) = P_\vartheta(S(X) \geq \gamma(\vartheta)) \geq \frac{1}{2}$$

(In jeweils 50% der Fälle Unter- bzw. Überschätzung.)

Beispiele:

- a)  $X_1, \dots, X_n$  reellwertig,  $X_1, \dots, X_n \stackrel{i.i.d.}{\sim} P_\vartheta$ ,  $\mu(\vartheta) := E_\vartheta X_1$   
 $(E_\vartheta |X_1| < \infty)$ .

$\bar{X}_n$  ist erwartungstreu für  $\mu(\vartheta)$

$X_1$  ist erwartungstreu für  $\mu(\vartheta)$

- b) Wie a),  $E_\vartheta X_1^2 < \infty$ .  $\sigma^2(\vartheta) := \text{Var}_\vartheta(X_1)$ .

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

$S_n^2$  ist erwartungstreu für  $\sigma^2(\vartheta)$ .

---

<sup>8</sup>→ Abbildung 3.1

Beweis:

$$\begin{aligned}
 E_{\vartheta} S_n^2 &= \frac{1}{n-1} E_{\vartheta} \left[ \sum_{i=1}^n ((X_i - \mu(\vartheta)) - (\bar{X}_n - \mu(\vartheta)))^2 \right] \\
 &= \frac{1}{n-1} \left[ \sum_{i=1}^n \underbrace{E_{\vartheta} (X_i - \mu(\vartheta))^2}_{=\text{Var}_{\vartheta}(X_i)} - n \underbrace{E_{\vartheta} (\bar{X}_n - \mu(\vartheta))^2}_{=\text{Var}_{\vartheta}(\bar{X}_n) = \frac{\sigma^2(\vartheta)}{n}} \right] \\
 &= \frac{1}{n-1} (n\sigma^2(\vartheta) - \sigma^2(\vartheta)) \\
 &= \sigma^2(\vartheta)
 \end{aligned}$$

c)  $X_1, \dots, X_n \stackrel{uiv}{\sim} \mathcal{N}(\mu, \sigma^2)$ ,  $\vartheta = (\mu, \sigma^2)$ ,  $\gamma(\vartheta) = E_{\vartheta} X_1 = \mu$

$$\bar{X}_n \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right) \Rightarrow \text{med}_{\vartheta} \bar{X}_n = \mu$$

$\Rightarrow \bar{X}_n$  ist mediantreu für  $\gamma(\vartheta)$ .

### 3.5 Definition (Sei $\Gamma \subset \mathbb{R}^k$ .)

Schätzfolge  $S_n = S_n(X_1, \dots, X_n)$ ,  $n \geq 1$ , heißt (schwach) **konsistent** für  $\gamma(\vartheta) \Leftrightarrow$

$$P_{\vartheta}(\|S_n(X_1, \dots, X_n) - \gamma(\vartheta)\| \geq \varepsilon) \rightarrow 0 \quad \forall \varepsilon > 0 \quad \forall \vartheta \in \Theta$$

( $\forall \vartheta \in \Theta : S_n \xrightarrow{P_{\vartheta}} \gamma(\vartheta)$ )

### 3.6 Bemerkung (Sei $\Gamma \subset \mathbb{R}$ .)

( $S_n$ ) asymptotisch erwartungstreu für  $\gamma(\vartheta)$  und  $\text{Var}_{\vartheta} S_n \rightarrow 0 (n \rightarrow \infty) \Rightarrow$   
 ( $S_n$ ) konsistent für  $\gamma(\vartheta)$ .

[Beweis:]

$$\begin{aligned}
 P_{\vartheta}(|S_n - \gamma| \geq \varepsilon) &\leq P_{\vartheta}(|S_n - ES_n| + |ES_n - \gamma| \geq \varepsilon) \\
 &\leq P_{\vartheta}(|S_n - ES_n| \geq \frac{\varepsilon}{2} \text{ oder } |ES_n - \gamma| \geq \frac{\varepsilon}{2}) \\
 &\leq P_{\vartheta}(|S_n - ES_n| \geq \frac{\varepsilon}{2}) + P_{\vartheta}(|ES_n - \gamma| \geq \frac{\varepsilon}{2}) \\
 &\stackrel{(*)}{\leq} \frac{\text{Var}(S_n)}{(\frac{\varepsilon}{2})^2} + P_{\vartheta}(|ES_n - \gamma| \geq \frac{\varepsilon}{2}) \\
 &\rightarrow 0 \quad (n \rightarrow \infty)
 \end{aligned}$$

(\*): Tschebyscheff

Kurz:

$$|S_n - \gamma| \leq \underbrace{|S_n - E_\vartheta S_n|}_{\xrightarrow{P_\vartheta} 0 \text{ (1)}} + \underbrace{|E_\vartheta S_n - \gamma|}_{\xrightarrow{P_\vartheta} 0 \text{ (2)}} \xrightarrow{P_\vartheta} 0$$

(1): Tschebyscheff, (2): asymptotisch erwartungstreu

Obiges Beispiel a):

$\bar{X}_n$  konsistent für  $\mu(\vartheta)$ , falls  $E_\vartheta X_1^2 < \infty$  nach 3.6.

Starkes Gesetz der großen Zahlen (SGGZ):

$$\bar{X}_n \xrightarrow{P_\vartheta\text{-f.s.}} E_\vartheta X_1 \text{ (ohne weitere Voraussetzung)} \Rightarrow \bar{X}_n \xrightarrow{P_\vartheta} E_\vartheta X_1$$

### 3.7 Bemerkung und Definition (Sei $\Gamma \subset \mathbb{R}$ .)

$$\text{MQA}_S(\vartheta) := E_\vartheta(S(X) - \gamma(\vartheta))^2$$

heißt **mittlere quadratische Abweichung** von S an der Stelle  $\vartheta$ .

Es gilt:

$$\text{MQA}_S(\vartheta) = E_\vartheta(S(X) - E_\vartheta S(X))^2 + (E_\vartheta S(X) - \gamma(\vartheta))^2 = \text{Var}_\vartheta S(X) + b_s^2(\vartheta)$$

### 3.8 Beispiel

$X_1, \dots, X_n \stackrel{i.i.d.}{\sim} \mathcal{N}(\mu, \sigma^2)$ ,  $\mu, \sigma^2$  unbekannt

Schätzer von  $\sigma^2$ :

$$\tilde{\sigma}_n^2(c) = c \cdot \sum_{j=1}^n (X_j - \bar{X}_n)^2$$

Ziel: c so wählen, dass MQA von  $\tilde{\sigma}_n^2(c)$  minimal wird.

$$E_\vartheta(\tilde{\sigma}_n^2(c)) = c\sigma^2 \cdot E_\vartheta \left[ \underbrace{\frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \bar{X}_n)^2}_{\sim \chi_{n-1}^2} \right] = c\sigma^2 \cdot (n-1)$$

$$\text{Var}_\vartheta(\tilde{\sigma}_n^2(c)) = c^2 \sigma^4 \cdot 2(n-1)$$

Damit:

$$\begin{aligned}\text{MQA}_{\hat{\sigma}_n^2(c)}(\vartheta) &= 2(n-1)c^2\sigma^4 + (c\sigma^2(n-1) - \sigma^2)^2 \\ &= \dots \\ &= \sigma^4(n^2-1) \left[ \left(c - \frac{1}{n+1}\right)^2 + \frac{2}{(n-1)(n+1)^2} \right] \\ &\stackrel{!}{=} \min\end{aligned}$$

Dies führt offensichtlich auf  $c = \frac{1}{n+1}$ .



## 4 Schätzmethoden

a) Maximum-Likelihood-Schätzer (ML-Schätzer)

ML-Methode (R. A. Fisher) setzt dominierte Verteilungsklasse  
 $\mathcal{P} := \{P_\vartheta : \vartheta \in \Theta\}$  voraus. ( $\Theta \subset \mathbb{R}^k$ )

Im Folgenden:  $(\mathcal{X}, \mathcal{B}) = (\mathbb{R}^n, \mathcal{B}^n)$

### 4.1 Grundannahmen

1)  $\exists \sigma$ -endliches Maß  $\mu$  auf  $\mathcal{B}$  mit:

$$\forall N \in \mathcal{B} : \mu(N) = 0 \Rightarrow P_\vartheta(N) = 0 \quad \forall \vartheta \in \Theta$$

d.h.  $P_\vartheta$  stetig bzgl.  $\mu \quad \forall \vartheta$ .

( $\Rightarrow P_\vartheta$  besitzt Dichte bzgl.  $\mu$ )

2) Im Folgenden stets

(i)  $\mu = \lambda^n$  (Borel-Lebesgue-Maß)  
 ( $\rightarrow$  stetige Verteilung)

oder

(ii)  $\mu =$  Zählmaß auf einer abzählbaren Menge  $A \subset \mathbb{R}^n$ .  
 ( $\rightarrow$  diskrete Verteilung)

Im Falle (i) bezeichne  $f(x, \vartheta) = \frac{dP_\vartheta}{d\lambda^n}(x)$  die Lebesgue-Dichte<sup>9</sup> von X, also

$$P_\vartheta(X \in B) = \int_B f(x, \vartheta) d\lambda^n(x), \quad B \in \mathcal{B}$$

Im Falle (ii) bezeichne  $f(x, \vartheta) = \frac{dP_\vartheta}{d\mu}(x)$  die Zähldichte von X, also

$$f(x, \vartheta) = P_\vartheta(X = x), \quad x \in A$$

$$P_\vartheta(X \in B) = \sum_{x \in B \cap A} f(x, \vartheta)$$

---

<sup>9</sup>Beachte Schreibweise aus Stochastik II!

## 4.2 Definition und Bemerkung

Für jedes  $x \in \mathbb{R}^n$  heißt die Abbildung

$$L_x : \begin{cases} \Theta & \rightarrow [0, \infty) \\ \vartheta & \mapsto L_x(\vartheta) := f(x, \vartheta) \end{cases}$$

die Likelihood-Funktion zur Stichprobe  $x$ .

Jeder Wert  $\hat{\vartheta}(x) \in \Theta$ , der Lösung  $t$  von

$$L_x(t) = \sup_{\vartheta \in \Theta} L_x(\vartheta) \quad (*)$$

ist, heißt (ein) ML-Schätzwert für  $\vartheta \in \Theta$

- (i) Im Allgemeinen Existenz gesichert, falls  $\Theta$  abgeschlossen ist.
- (ii) Falls  $\Theta$  nicht abgeschlossen, so häufig  $\vartheta \mapsto f(x, \vartheta)$  auf  $\bar{\Theta}$  fortsetzbar. Dann sieht man  $\hat{\vartheta}(x)$  auch als Lösung an, wenn  $\sup$  in  $(*)$  im Punkt  $\hat{\vartheta}(x) \in \bar{\Theta} \setminus \Theta$  angenommen wird.

Eine messbare Funktion  $\hat{\vartheta} : (\mathbb{R}^n, \mathcal{B}^n) \rightarrow (\bar{\Theta}, \bar{\Theta} \cap \mathcal{B}^k)$  heißt **ML-Schätzer** für  $\vartheta$ , wenn für jedes  $x \in \mathfrak{X}$  gilt:  $\hat{\vartheta}(x)$  ist Lösung von  $(*)$  im obigen Sinn<sup>10</sup>.

## 4.3 Bemerkungen

- (i) Oft ist  $L_x(\cdot) = f(x, \cdot)$  differenzierbar. Dann liefert  $\frac{\partial}{\partial \vartheta} f(x, \vartheta) \stackrel{!}{=} 0 \in \mathbb{R}^k$  die lokalen Maximalstellen von  $L_x$  im Inneren  $\Theta^0$  von  $\Theta$ .
- (ii) Oft:  $X = (X_1, \dots, X_n)$ ,  $X_1, \dots, X_n \stackrel{uiv}{\sim} f_1(\xi, \vartheta)$  [Dichte von  $X_1$ .]  
Dann:

$$f(x, \vartheta) = \prod_{j=1}^n f_1(x_j, \vartheta), \quad x = (x_1, \dots, x_n)$$

Log-Likelihood-Funktion

$$\log L_x(\vartheta) = \sum_{j=1}^n \log f_1(x_j, \vartheta)$$

$$\frac{\partial}{\partial \vartheta} \log L_x(\vartheta) \stackrel{!}{=} 0 \rightarrow \text{Maximalstellen von } L_x \text{ in } \Theta^0$$

<sup>10</sup>siehe Punkt (ii)

#### 4.4 Satz (Invarianzprinzip für ML-Schätzer)

Sei  $g : \mathbb{R}^k \rightarrow \mathbb{R}^l$  messbar und

$$M_x(\gamma) := \sup_{\vartheta: g(\vartheta)=\gamma} L_x(\vartheta)$$

(sogenannte von  $g$  induzierte Likelihood-Funktion)

Ist  $\hat{\vartheta}$  ML-Schätzer für  $\vartheta \in \Theta$ , so ist  $\hat{\gamma} := g(\hat{\vartheta})$  der ML-Schätzer für  $\gamma = g(\vartheta) \in \Gamma := g(\Theta)$ , es gilt also  $M(\hat{\gamma}) \geq M(\gamma) \forall \gamma \in \Gamma$ .  
(Plug-In-Methode)

Beweis:<sup>11</sup>

Aus

$$M_x(g(\hat{\vartheta})) = \sup_{\vartheta: g(\vartheta)=g(\hat{\vartheta})} L_x(\vartheta) \geq L_x(\hat{\vartheta})$$

und

$$M_x(g(\hat{\vartheta})) \leq \sup_{\gamma \in \Gamma} M_x(\gamma) = L_x(\hat{\vartheta})$$

folgt

$$M_x(g(\hat{\vartheta})) = L_x(\hat{\vartheta}) \geq M_x(\gamma) \quad \forall \gamma \in \Gamma$$

#### 4.5 Beispiel

$X_1, \dots, X_n \stackrel{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$ ,  $\vartheta = (\mu, \sigma^2)$

$$\hat{\vartheta}(x) = (\bar{X}_n, \hat{\sigma}_n^2)$$

$\Rightarrow \bar{X}_n$  ist ML-Schätzer für  $\mu$

$\hat{\sigma}_n^2$  ist ML-Schätzer für  $\sigma^2$

$\hat{\sigma}_n = +\sqrt{\hat{\sigma}_n^2}$  ist ML-Schätzer für  $\sigma$

b) Minimum-Quadrat-Schätzer (MQ-Schätzer)

#### 4.6 Situation

Seien  $X_1, \dots, X_n$  stochastisch unabhängig.

Annahme:

$EX_j = \mu_j(\vartheta)$ , wobei  $\vartheta \in \mathbb{R}^p$  **unbekannt**,  $\mu_j : \mathbb{R}^p \rightarrow \mathbb{R}$  ( $j = 1, \dots, n$ )  
**bekannte Regressionsfunktionen.**

<sup>11</sup>In der 1. Zeile gilt eigentlich bereits Gleichheit.

Für  $\varepsilon_j := X_j - EX_j$  gilt dann:

$\varepsilon_1, \dots, \varepsilon_n$  unabhängig,  $E(\varepsilon_j) = 0 \forall j$ ,  $X_j = \mu_j(\vartheta) + \varepsilon_j$  ( $j = 1, \dots, n$ ) bzw.

$$X = \mu(\vartheta) + \varepsilon$$

wobei

$$X = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}, \mu(\vartheta) = \begin{pmatrix} \mu_1(\vartheta) \\ \vdots \\ \mu_n(\vartheta) \end{pmatrix}, \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

Schätzmethode von  $\vartheta$  durch Methode der **kleinsten Quadrate**, d.h. durch Minimierung der Fehlerquadratsumme

$$Q(\vartheta) := \sum_{j=1}^n (X_j - \mu_j(\vartheta))^2 = \|X - \mu(\vartheta)\|^2$$

Sind  $\mu_1, \dots, \mu_n$  stetig differenzierbar, so gilt mit

$$M(\vartheta) := \begin{pmatrix} \frac{\partial}{\partial \vartheta_1} \mu_1(\vartheta) & \cdots & \frac{\partial}{\partial \vartheta_p} \mu_1(\vartheta) \\ \vdots & & \vdots \\ \frac{\partial}{\partial \vartheta_1} \mu_n(\vartheta) & \cdots & \frac{\partial}{\partial \vartheta_p} \mu_n(\vartheta) \end{pmatrix}$$

$$\frac{\partial}{\partial \vartheta} Q(\vartheta) = -2 \cdot M^T(\vartheta) \cdot (X - \mu(\vartheta))$$

Beweis:

Sei allgemein  $f, g: \mathbb{R}^p \rightarrow \mathbb{R}^q$ .

Jacobi-Matrix

$$J_f = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_p} \\ \vdots & & \vdots \\ \frac{\partial f_q}{\partial x_1} & \cdots & \frac{\partial f_q}{\partial x_p} \end{pmatrix} \in \mathbb{R}^{q \times p}$$

$$\Rightarrow \underbrace{J_{f \cdot g}^T}_{\in \mathbb{R}^p} = \underbrace{J_f^T}_{\in \mathbb{R}^{p \times q}} \cdot \underbrace{g}_{\in \mathbb{R}^q} + \underbrace{J_g^T}_{\in \mathbb{R}^p} \cdot \underbrace{f}_{\in \mathbb{R}^q} \quad (*)$$

[Beachte: f,g vektorwertig!]

Hier speziell:  $f(\vartheta) = g(\vartheta) = X - \mu(\vartheta)$

$$\Rightarrow Q(\vartheta) = f^T(\vartheta) \cdot f(\vartheta)$$

$$\Rightarrow J_f = -M(\vartheta) = J_g$$

$$\begin{aligned} \stackrel{(*)}{\Rightarrow} \frac{\partial}{\partial \vartheta} [f^T(\vartheta) \cdot f(\vartheta)] &= -M^T(\vartheta)g - M^T(\vartheta)f \\ &= -2M^T(\vartheta)(X - \mu(\vartheta)) \end{aligned}$$

Die Lösungen  $\hat{\vartheta}$  von

$$Q(\hat{\vartheta}) = \min_{\vartheta \in \mathbb{R}^p} Q(\vartheta)$$

(sogenannte MQ-Schätzer) befinden sich also unter den Lösungen  $\vartheta$  der sogenannten **Normalengleichung**

$$M^T(\vartheta) \cdot \mu(\vartheta) = M^T(\vartheta) \cdot X$$

#### 4.7 Beispiel (Einfach lineare Regression)

$\vartheta = (\vartheta_0, \vartheta_1) \in \mathbb{R}$

$$\mu_i(\vartheta) = \vartheta_0 + \vartheta_1 t_i \quad (i = 1, \dots, n)$$

$t_i$  bekannt, nicht alle gleich.

$$Q(\vartheta) = \sum_{i=1}^n (X_i - \vartheta_0 - \vartheta_1 t_i)^2 = \min_{\vartheta_0, \vartheta_1}!$$

$$M^T(\vartheta) = \begin{pmatrix} 1 & \dots & 1 \\ t_1 & \dots & t_n \end{pmatrix}$$

Normalengleichung:

$$\begin{pmatrix} 1 & \dots & 1 \\ t_1 & \dots & t_n \end{pmatrix} \begin{pmatrix} \vartheta_0 + \vartheta_1 t_1 \\ \vdots \\ \vartheta_0 + \vartheta_1 t_n \end{pmatrix} = \begin{pmatrix} 1 & \dots & 1 \\ t_1 & \dots & t_n \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

$$\Leftrightarrow n\vartheta_0 + \vartheta_1 \sum_{i=1}^n t_i = \sum_{i=1}^n x_i$$

$$\vartheta_0 \sum_{i=1}^n t_i + \vartheta_1 \sum_{i=1}^n t_i^2 = \sum_{i=1}^n t_i x_i$$

Mit  $\bar{t} = \frac{1}{n} \sum_{i=1}^n t_i$ ,  $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$  folgt

$$\begin{aligned} \hat{\vartheta}_0 &= \bar{x} - \hat{\vartheta}_1 \bar{t} \\ \hat{\vartheta}_1 &= \frac{\sum_{i=1}^n t_i x_i - n \bar{t} \bar{x}}{\sum_{i=1}^n (t_i - \bar{t})^2} \end{aligned}$$

Wegen<sup>12</sup>

$$\sum_i a_i b_i - n \bar{a} \bar{b} = \sum_i (a_i - \bar{a})(b_i - \bar{b}) = \sum_i (a_i - \bar{a}) b_i$$

folgt

$$\hat{\vartheta}_1 = \hat{\vartheta}_1(X) = \frac{\sum_{i=1}^n (t_i - \bar{t}) x_i}{\sum_{i=1}^n (t_i - \bar{t})^2}$$

und somit

$$E(\vartheta_1) = \frac{1}{\sum_i (t_i - \bar{t})^2} \sum_i (t_i - \bar{t})(\vartheta_0 + \vartheta_1 t_i) = \vartheta_1$$

Falls  $\text{Var}(X_i) = \sigma^2 \forall i$ , so gilt:

$$\text{Var}(\hat{\vartheta}_1) = \frac{1}{(\sum_i (t_i - \bar{t})^2)^2} \sum_i (t_i - \bar{t})^2 \sigma^2 = \frac{\sigma^2}{\sum_{i=1}^n (t_i - \bar{t})^2}$$

[ $\text{Var}(\hat{\vartheta}_1) = \text{MQA}$ , da erwartungstreu;  $t_i$  so wählen, dass  $\text{Var}(\hat{\vartheta}_1)$  klein wird, also möglichst weit auseinander.]

Weiter gilt

$$E(\hat{\vartheta}_0) = E\bar{X} - \bar{t}E(\hat{\vartheta}_1) = \frac{1}{n} \sum_{i=1}^n (\vartheta_0 + \vartheta_1 t_i) - \bar{t}\vartheta_1 = \vartheta_0 + \vartheta_1 \bar{t} - \vartheta_1 \bar{t} = \vartheta_0$$

Bemerkungen:

- (i) Falls  $\text{Var}(X_i) = \sigma^2 \forall i$  ( $\text{Cov}(X_i, X_j) = 0 \forall i \neq j$  wegen Unabhängigkeit<sup>13</sup>), so gilt mit  $\bar{t}^2 = \frac{1}{n} \sum_{i=1}^n t_i^2$

$$\text{Var}(\hat{\vartheta}_0) = \frac{\sigma^2 \bar{t}^2}{n(\bar{t}^2 - (\bar{t})^2)}$$

- (ii) Falls  $X_i \sim \mathcal{N}(\mu, \sigma^2) \forall i$ , so ist der MQ-Schätzer auch ML-Schätzer für  $\vartheta$

<sup>12</sup> $\sum (a_i - \bar{a})\bar{b} = \bar{b} \sum (a_i - \bar{a}) = 0$

<sup>13</sup>Voraussetzung!

c) Momentenmethode

#### 4.8 Definition

Es seien  $X_1, \dots, X_n \stackrel{uiv}{\sim} X$ ,  $X$  reellwertig,  $P^X \in \{P_\vartheta : \vartheta \in \Theta\}$ ,  $\Theta \subset \mathbb{R}^k$

$$\vartheta = (\vartheta_1, \dots, \vartheta_k)$$

Annahme

(i)  $E|X^k| < \infty$

(ii) Es gibt Funktionen  $g_1, \dots, g_k : \mathbb{R}^k \rightarrow \mathbb{R}$  mit

$$\vartheta_1 = g_1(EX, \dots, EX^k)$$

⋮

$$\vartheta_k = g_k(EX, \dots, EX^k)$$

Sei  $\bar{X}_n^l = \frac{1}{n} \sum_{j=1}^n X_j^l$  ( $l = 1, \dots, k$ ).

Dann ist der Momentenschätzer für  $\vartheta$

$$\hat{\vartheta} := \begin{pmatrix} g_1(\bar{X}_n^1, \dots, \bar{X}_n^k) \\ \vdots \\ g_k(\bar{X}_n^1, \dots, \bar{X}_n^k) \end{pmatrix}$$

Beispiel:

$$X_i \stackrel{uiv}{\sim} \mathcal{N}(\mu, \sigma^2), \mu = EX, \sigma^2 = EX^2 - (EX)^2 \\ \Rightarrow \hat{\mu}_n = \bar{X}_n^1, \hat{\sigma}_n^2 = \bar{X}_n^2 - (\bar{X}_n^1)^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$$

Probleme:

$g_1, \dots, g_k$  nicht explizit gegeben.

Ausreißeranfälligkeit.

Momentenschätzer sind nicht „robust“.

Beachte

Momentenschätzer sind konsistent, falls  $g_1, \dots, g_k$  stetig sind an der Stelle  $(EX, \dots, EX^k)$ .

### 4.9 Beispiel (Gamma-Verteilung)

$X_1, \dots, X_n \stackrel{uiv}{\sim} \Gamma(\alpha, \beta)$ , Dichte

$$f(x, \alpha, \beta) = \frac{\beta^\alpha x^{\alpha-1} e^{-\beta x}}{\Gamma(\alpha)} \quad (x > 0)$$

$\vartheta = (\alpha, \beta) \in \Theta = \mathbb{R}_{>0} \times \mathbb{R}_{>0}$

$X \sim f(x, \alpha, \beta) \Rightarrow EX = \frac{\alpha}{\beta}, EX^2 = \frac{\alpha(\alpha+1)}{\beta^2} \Rightarrow$

$$\alpha = \frac{(EX)^2}{EX^2 - (EX)^2} =: g_1(EX, EX^2)$$

$$\beta = \frac{EX}{EX^2 - (EX)^2} =: g_2(EX, EX^2)$$

$\Rightarrow$  Momentenschätzer

$$\hat{\alpha} = \frac{(\bar{X}_n^1)^2}{\bar{X}_n^2 - (\bar{X}_n^1)^2} = \frac{\bar{X}_n^2}{\hat{\sigma}_n^2}$$

$$\hat{\beta} = \frac{\bar{X}_n}{\hat{\sigma}_n^2}$$

#### d) Ein nichtparametrisches Schätzprinzip

Seien  $X_1, \dots, X_n \stackrel{uiv}{\sim} F$ ,  $F(t) = P(X \leq t)$ ,  $t \in \mathbb{R}$ ,  $F$  unbekannt

### 4.10 Definition

Die durch

$$\hat{F}_n(t) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}\{X_i \leq t\}, \quad t \in \mathbb{R}$$

definierte Funktion heißt **empirische Verteilungsfunktion** (EVF) von  $X_1, \dots, X_n$ .

Die Realisierungen von  $\hat{F}_n$  sind Treppenfunktionen.

$$\hat{F}_n(t_0) \xrightarrow{f.s.} E[\mathbf{1}\{X_1 \leq t_0\}] = P(X_1 \leq t_0) = F(t_0)$$

### 4.11 Satz von Glivenko-Cantelli

Sei  $\hat{F}_n^\omega(t) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}\{X_i(\omega) \leq t\}$ ,  $\omega \in \Omega$ .

Falls  $X_1, \dots, X_n \stackrel{i.i.d.}{\sim} F$  auf Wahrscheinlichkeitsraum  $(\Omega, \mathcal{A}, \mathbb{P})$ , so gilt

$$\mathbb{P}(\{\omega \in \Omega : \lim_{n \rightarrow \infty} \sup_{t \in \mathbb{R}} \underbrace{|\hat{F}_n^\omega(t) - F(t)|}_{=:\|\hat{F}_n^\omega - F\|_\infty} = 0\}) = 1$$

Kurz:  $\|\hat{F}_n^\omega - F\|_\infty \rightarrow 0 \quad \mathbb{P} - f.s.$   
(Stochastik II, Henze)

### 4.12 $\hat{F}_n$ als nichtparametrischer ML-Schätzer

Sei  $\mathfrak{F}$  die Menge aller Verteilungsfunktionen auf  $\mathbb{R}$ ,  $X_1, \dots, X_n \stackrel{i.i.d.}{\sim} F \in \mathfrak{F}$ .

Sei  $P_F$  das zu  $F$  gehörende Wahrscheinlichkeitsmaß auf  $\mathcal{B}^1$ , also

$$P_F([a, b]) = F(b) - F(a), \quad a < b$$

$$P_F(\{x\}) = F(x) - F(x-0), \quad x \in \mathbb{R}$$

Sei  $(x_1, \dots, x_n)$  Realisierung von  $(X_1, \dots, X_n)$ . Die durch

$$L_x : \begin{array}{ll} \mathfrak{F} & \rightarrow [0, \infty) \\ G & \mapsto L_x(G) := \prod_{i=1}^n P_G(\{x_i\}) \end{array}$$

definierte Funktion heißt nichtparametrische Likelihood-Funktion zu  $x = (x_1, \dots, x_n)$ .

Beachte:  $L_x(G) = 0$ , falls  $P_G(\{x_i\}) = 0$  für ein  $i$ .<sup>14</sup>

Behauptung:

$L_x(\cdot)$  wird maximal für  $G(t) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}\{x_i \leq t\}$ .

Beweis:

Seien  $z_1, \dots, z_k$  die unterschiedlichen Werte unter  $x_1, \dots, x_n$ ,  $n_1, \dots, n_k$  die entsprechenden Vielfachheiten.

$$L_x(G) = \prod_{i=1}^n P_G(\{x_i\}) = \prod_{j=1}^k \underbrace{P_G(\{z_j\})}_{=: p_j}^{n_j} = \prod_{j=1}^k p_j^{n_j}$$

Setze  $\hat{p}_j := \frac{n_j}{n}$ ,  $j = 1, \dots, k$ , Verteilungsfunktion ist  $\hat{F}_n$ .

<sup>14</sup>z.B. für  $G$  stetig

$F$  sei beliebige Verteilungsfunktion mit  $p_j := F(z_j) - F(z_j - 0) > 0, j = 1, \dots, k$  mit  $p_j \neq \hat{p}_j$  für mindestens ein  $j$ .

Es gilt für  $x > 0$ :

$$\log x \leq x - 1 \quad (*)$$

$\log x = x - 1$  nur für  $x = 1$ .

$$\begin{aligned} \log \left( \frac{L_x(F)}{L_x(\hat{F}_n)} \right) &= \sum_{j=1}^k n_j \cdot \log \left( \frac{p_j}{\hat{p}_j} \right) \\ &= n \sum_{j=1}^k \hat{p}_j \cdot \log \left( \frac{p_j}{\hat{p}_j} \right) \\ &\stackrel{(*)}{<} n \sum_{j=1}^k \hat{p}_j \left( \frac{p_j}{\hat{p}_j} - 1 \right) \\ &= n \left( \underbrace{\sum_{j=1}^k p_j}_{\leq 1} - \underbrace{\sum_{j=1}^k \hat{p}_j}_{=1} \right) \\ &\leq 0 \end{aligned}$$

$\Rightarrow L_x(F) < L_x(\hat{F}_n)$  ■

### 4.13 Nichtparametrisches Schätzprinzip

Seien  $X_1, \dots, X_n \stackrel{uiv}{\sim} F, F \in \mathfrak{F}, \mathfrak{F}$  Menge von Verteilungsfunktionen (Verteilungsannahme).

Sei  $\gamma : \mathfrak{F} \rightarrow \mathbb{R}$  Funktional.

Interessierender Parameter sei  $\gamma(F)$ .

„Rezept“: Schätze  $\gamma(F)$  durch  $\gamma(\hat{F}_n)$

### 4.14 Beispiele

$$\text{a) } \mathfrak{F} := \left\{ F : \underbrace{\int |x| F(dx)}_{=E|X_1|} < \infty \right\}$$

$$\gamma(F) := \int x F(dx) (= EX_1)$$

$$\gamma(\hat{F}_n) = \int x \hat{F}_n(dx) = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X}_n$$

b)  $\mathfrak{F} := \{F : \int x^2 F(dx) < \infty\}$

$$\gamma(F) := \int (x - \int y dF(y))^2 dF(x) = \text{Var}(X_1)$$

$$\gamma(\hat{F}_n) = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2$$

c)  $\mathfrak{F} := \{F : F \text{ hat Lebesgue-Dichte } f\}$

$$\gamma(F) = F'(t_0) = f(t_0)$$

$$\gamma(\hat{F}_n) = ?$$



## 5 Optimale erwartungstreue Schätzer

### 5.1 Definition

Seien  $X_1, \dots, X_n$  reelle Zufallsvariablen,  $T = T(X_1, \dots, X_n)$  reellwertige Statistik.

T heißt **linear**  $\Leftrightarrow \exists c_1, \dots, c_n \in \mathbb{R}$  mit  $T = \sum_{j=1}^n c_j X_j$

### 5.2 Satz

Seien  $X_1, \dots, X_n \stackrel{uiv}{\sim} X$ ,  $EX^2 < \infty$ ,  $\mu := EX$ ,  $\sigma^2 := \text{Var}(X)$ ,  $(\mu, \sigma^2)$  unbekannt. Es sei T ein beliebiger linearer erwartungstreuer Schätzer für  $\mu$ . Dann gilt:

$$\text{Var}(T) \geq \text{Var}(\bar{X}_n) = \frac{\sigma^2}{n}$$

Beweis:

Sei  $T = \sum_{j=1}^n c_j X_j$ .

T erwartungstreu

$$\Rightarrow \mu = E(T) = \mu \sum_{j=1}^n c_j$$

$$\Rightarrow \sum_{j=1}^n c_j = 1$$

$$\text{Var}(T) = \sigma^2 \sum_{j=1}^n c_j^2$$

$$\underbrace{\left( \sum_{j=1}^n c_j \cdot 1 \right)^2}_{=1} \leq \sum_{j=1}^n c_j^2 \underbrace{\sum_{j=1}^n 1^2}_{=n}$$

(Cauchy-Schwarz)

$$\sum_{j=1}^n c_j^2 \geq \frac{1}{n}$$

$$\sum_{j=1}^n c_j^2 = \frac{1}{n} \Leftrightarrow c_j = \frac{1}{n} \quad \forall j$$

$\Rightarrow T = \bar{X}_n$ . ■

### 5.3 Situation

Sei  $(\mathfrak{X}, \mathcal{B}, \{P_\vartheta : \vartheta \in \Theta\})$ ,  $\Theta \subset \mathbb{R}^k$ , ein statistischer Raum.  $X_1, \dots, X_n \stackrel{uiv}{\sim} P_\vartheta$ .

$g : \Theta \rightarrow \mathbb{R}$  Funktional

$g(\vartheta)$  interessierender Parameter.

Sei

$$U_g = \{T \mid T : \mathfrak{X}^n \rightarrow \mathbb{R} \text{ messbar, } E_\vartheta T = g(\vartheta) \forall \vartheta \in \Theta, E_\vartheta T^2 < \infty \forall \vartheta \in \Theta\}$$

die Menge aller erwartungstreuen Schätzer für  $g(\vartheta)$  mit endlicher Varianz.

Annahme:  $U_g \neq \emptyset$

Sei

$$m(\vartheta) := \inf\{\text{Var}_\vartheta(T) : T \in U_g\}$$

### 5.4 Definition

Ein  $T_0 \in U_g$  mit  $\text{Var}_\vartheta(T_0) = m(\vartheta) \forall \vartheta \in \Theta$  heißt **UMVUE**.  
(Uniformly Minimum Variance Unbiased Estimator)

### 5.5 Satz

Falls  $T_1$  und  $T_2$  UMVUE, so gilt

$$P_\vartheta(T_1 = T_2) = 1 \forall \vartheta \in \Theta$$

Beweis:

$U_g$  ist konvex, d.h.

$$S, T \in U_g \Rightarrow \lambda S + (1 - \lambda)T \in U_g \forall \lambda \in [0, 1]$$

Seien  $T_1, T_2$  UMVUE.

$$\Rightarrow \frac{1}{2}(T_1 + T_2) \in U_g$$

$$\begin{aligned} \Rightarrow \underbrace{\text{Var}_\vartheta\left(\frac{1}{2}(T_1 + T_2)\right)} &\geq \text{Var}_\vartheta(T_1) (= m(\vartheta) = \text{Var}_\vartheta(T_2)) \\ &= \frac{1}{4}(\text{Var}_\vartheta(T_1) + \text{Var}_\vartheta(T_2) + 2\text{Cov}_\vartheta(T_1, T_2)) \end{aligned}$$

$$\Rightarrow \text{Var}_\vartheta(T_1) \leq \text{Cov}_\vartheta(T_1, T_2) \stackrel{\text{CSU}}{\leq} \sqrt{\text{Var}_\vartheta(T_1) \text{Var}_\vartheta(T_2)} = \text{Var}_\vartheta(T_1)$$

$$\begin{aligned} \Rightarrow \text{Var}_{\vartheta}(T_1) &= \text{Cov}_{\vartheta}(T_1, T_2) \\ \Rightarrow \text{Var}_{\vartheta}(T_1 - T_2) &= \text{Var}_{\vartheta}(T_1) + \text{Var}_{\vartheta}(T_2) - 2 \text{Cov}_{\vartheta}(T_1, T_2) = 0 \\ E_{\vartheta}(T_1 - T_2) &= 0 \\ \Rightarrow P_{\vartheta}(T_1 = T_2) &= 1. \blacksquare \end{aligned}$$

## 5.6 Definition und Satz

Sei

$$\mathcal{S}_n := \{\pi = (\pi(1), \dots, \pi(n)) : \pi \text{ Permutation von } \{1, \dots, n\}\}$$

Für Statistik  $T : \mathfrak{X}^n \rightarrow \mathbb{R}$  sei  $T^{\pi}(X_1, \dots, X_n) = T(X_{\pi(1)}, \dots, X_{\pi(n)})$ .  
In der Situation von 5.3 heißt  $T$  (im wesentlichen) symmetrisch  $:\Leftrightarrow$

$$P_{\vartheta}(T^{\pi} = T) = 1 \quad \forall \vartheta \in \Theta \forall \pi \in \mathcal{S}_n$$

$T_0 \in U_g$  UMVUE  $\Rightarrow T$  symmetrisch.

Beweis:

Sei  $\pi \in \mathcal{S}_n$ ,  $\vartheta \in \Theta$  beliebig.

Wegen  $X_1, \dots, X_n \stackrel{uiv}{\sim} P_{\vartheta}$  folgt  $T_0^{\pi} \sim T_0$  unter  $P_{\vartheta}$

$$\left. \begin{aligned} \Rightarrow E_{\vartheta}(T_0^{\pi}) &= E_{\vartheta}(T_0) = g(\vartheta) \\ \Rightarrow \text{Var}_{\vartheta}(T_0^{\pi}) &= \text{Var}_{\vartheta}(T_0) = m(\vartheta) \end{aligned} \right\} \Rightarrow T_0^{\pi} \in U_g, \text{ UMVUE}$$

Satz 5.5  $\Rightarrow P_{\vartheta}(T_0^{\pi} = T_0) = 1. \blacksquare$

## 5.7 Reguläre Verteilungsklassen

Situation:

Sei  $(\mathfrak{X}, \mathcal{B}, \{P_{\vartheta} : \vartheta \in \Theta\})$  statistischer Raum mit  $(\mathfrak{X}, \mathcal{B}) = (\mathbb{R}^n, \mathcal{B}^n)$ ,  $\Theta \subset \mathbb{R}^k$ ,  $\Theta$  offen.

$X = (X_1, \dots, X_n)$  Zufallsvektor mit Verteilung  $P_{\vartheta}$  ( $\vartheta \in \Theta$ ),  $P_{\vartheta}$  besitze Dichte  $f(x, \vartheta)$  bezüglich  $\mu$ , dabei sei  $\mu$  entweder das Lebesgue-Maß oder das Zählmaß auf einer abzählbaren Teilmenge des  $\mathbb{R}^n$ .

$T : \mathbb{R}^n \rightarrow \mathbb{R}^s$  sei Statistik mit  $E_{\vartheta}\|T\|^2 < \infty$ , Kovarianzmatrix<sup>15</sup> von  $T$ :<sup>16</sup>

$$\text{Var}_{\vartheta}(T) := E_{\vartheta}[(T - E_{\vartheta}T)(T - E_{\vartheta}T)^T]$$

<sup>15</sup>Schreibweise für Kovarianzmatrix hier nicht  $\text{Cov}_{\vartheta}$ , sondern  $\text{Var}_{\vartheta}$ . Beachte dazu die Fälle  $s = 1$  und  $s > 1$ !

<sup>16</sup>Bei Vektoren manchmal Schreibweise  $x'$  für  $x^T$ .

Folgende Regularitätsbedingungen sollen gelten:

- a)  $\forall x \in \mathfrak{X}$  existiert  $\frac{\partial}{\partial \vartheta_j} f(x, \vartheta)$  und ist stetig. ( $j = 1, \dots, k$ )  
 b)

$$\frac{d}{d\vartheta} \int f(x, \vartheta) \mu(dx) = \int \frac{d}{d\vartheta} f(x, \vartheta) \mu(dx)$$

wobei hier  $\frac{d}{d\vartheta} := \left( \frac{\partial}{\partial \vartheta_1}, \dots, \frac{\partial}{\partial \vartheta_k} \right)^T$ .

Der  $k$ -dimensionale Zufallsvektor

$$\mathcal{U}_n(\vartheta) := \frac{d}{d\vartheta} \log f(X, \vartheta) = \frac{\frac{d}{d\vartheta} f(X, \vartheta)}{f(X, \vartheta)}$$

heißt Score-Vektor.

Die  $k \times k$ -Matrix

$$I_n(\vartheta) := E_{\vartheta}[\mathcal{U}_n(\vartheta) \cdot \mathcal{U}_n(\vartheta)^T] = \left( E_{\vartheta} \left[ \frac{\partial}{\partial \vartheta_i} \log f(X, \vartheta) \frac{\partial}{\partial \vartheta_j} \log f(X, \vartheta) \right] \right)_{i,j=1,\dots,k}$$

heißt **Fisher-Informationsmatrix** (von  $f$  an der Stelle  $\vartheta$ ):

- c)  $I_n(\vartheta)$  existiert und ist positiv definit.

Eine Verteilungsklasse  $\{P_{\vartheta} : \vartheta \in \Theta\}$ , die die Bedingungen (a)-(c) erfüllt, heißt **regulär**.

## 5.8 Lemma

In der Situation von 5.7 gilt:

$E_{\vartheta} \mathcal{U}_n(\vartheta) = 0 \forall \vartheta \in \Theta$  und somit  $I_n(\vartheta) = \text{Var}_{\vartheta}(\mathcal{U}_n(\vartheta))$ ,  $\vartheta \in \Theta$ , d.h. die Fisher-Informationsmatrix ist Kovarianzmatrix des Score-Vektors.

Beweis:

$$E_{\vartheta} \mathcal{U}_n(\vartheta) \stackrel{(*)}{=} \int \frac{\frac{d}{d\vartheta} f(x, \vartheta)}{f(x, \vartheta)} f(x, \vartheta) d\mu(x) \stackrel{(b)}{=} \frac{d}{d\vartheta} \underbrace{\int f(x, \vartheta) d\mu(x)}_{=1} = 0$$

(\*): Integration bezüglich  $P_{\vartheta}$ ;  $P_{\vartheta}$  hat aber Dichte  $f(x, \vartheta)$  bezüglich  $\mu$

## 5.9 Bemerkung

Gelegentlich werden die weiteren Voraussetzungen

d)  $\forall x \in \mathfrak{X}$  existiert  $\frac{\partial^2}{\partial \vartheta_i \partial \vartheta_j} f(x, \vartheta)$  und ist stetig. ( $i, j = 1, \dots, k$ )

e)

$$\frac{\partial^2}{\partial \vartheta_i \partial \vartheta_j} \int f(x, \vartheta) \mu(dx) = \int \frac{\partial^2}{\partial \vartheta_i \partial \vartheta_j} f(x, \vartheta) \mu(dx) \quad \forall i, j = 1, \dots, k$$

benötigt.

Wir führen noch die folgenden Notationen ein:

$$W_n(\vartheta) := \left( \frac{\partial^2}{\partial \vartheta_i \partial \vartheta_j} \log f(X, \vartheta) \right)_{1 \leq i, j \leq k} =: \frac{d^2}{d\vartheta d\vartheta^T} \log f(X, \vartheta)$$

## 5.10 Lemma

Unter (d) und (e) gilt:

$$I_n(\vartheta) = -E_\vartheta W_n(\vartheta)$$

Beweis:

Wegen

$$\frac{\partial^2}{\partial \vartheta_i \partial \vartheta_j} \log f = \frac{\frac{\partial^2}{\partial \vartheta_i \partial \vartheta_j} f}{f} - \frac{(\frac{\partial}{\partial \vartheta_i} f)(\frac{\partial}{\partial \vartheta_j} f)}{f^2}$$

folgt

$$\begin{aligned} E_\vartheta(W_n(\vartheta)) &= \int \frac{d^2}{d\vartheta d\vartheta^T} \log f(x, \vartheta) \cdot f(x, \vartheta) d\mu(x) \\ &= \underbrace{\left( \int \frac{\partial^2}{\partial \vartheta_i \partial \vartheta_j} f(x, \vartheta) \mu(dx) \right)}_{=0 \text{ nach (e) [vgl. 5.7]}}_{i,j} \\ &\quad - \left( \int \frac{\partial}{\partial \vartheta_i} \log f(x, \vartheta) \cdot \frac{\partial}{\partial \vartheta_j} \log f(x, \vartheta) \cdot f(x, \vartheta) d\mu(x) \right)_{i,j} \\ &= -E_\vartheta[\mathcal{U}_n(\vartheta) U_n n(\vartheta)^T] \\ &= -I_n(\vartheta) \end{aligned}$$

### 5.11 Reguläre Statistiken (Schätzer)

In der Situation von 5.7 heißt eine Statistik  $T : \mathbb{R}^n \rightarrow \mathbb{R}^s$  **regulär**, falls gilt:

- f) Die Funktion  $\Theta \ni \vartheta \mapsto E_\vartheta T \in \mathbb{R}^s$  ist stetig differenzierbar.
- g) Differenziation und Integration können vertauscht werden:

$$\frac{\partial}{\partial \vartheta_j} \int T(x) f(x, \vartheta) \mu(dx) = \int T(x) \frac{\partial}{\partial \vartheta_j} f(x, \vartheta) \mu(dx) \quad j = 1, \dots, k$$

Mit

$$C_n(\vartheta) := \begin{bmatrix} \frac{\partial}{\partial \vartheta_1} E_\vartheta T_1 & \cdots & \frac{\partial}{\partial \vartheta_1} E_\vartheta T_s \\ \vdots & \ddots & \vdots \\ \frac{\partial}{\partial \vartheta_k} E_\vartheta T_1 & \cdots & \frac{\partial}{\partial \vartheta_k} E_\vartheta T_s \end{bmatrix}_{k \times s} = \frac{d}{d\vartheta} E_\vartheta T^T$$

wird Bedingung (g) zu

$$C_n(\vartheta) = E_\vartheta[\mathcal{U}_n(\vartheta) T^T]$$

Wegen  $E_\vartheta[\mathcal{U}_n(\vartheta)] = 0$  folgt

$$C_n(\vartheta) = E_\vartheta[\mathcal{U}_n(\vartheta)(T - E_\vartheta T)^T]$$

### 5.12 Strukturlemma

Vorbemerkung:

Seien  $A, B$   $n \times n$ -Matrizen.

$A \geq B$   $:\Leftrightarrow A - B$  positiv semidefinit<sup>17</sup> ( $\Leftrightarrow x^T A x \geq x^T B x \quad \forall x \in \mathbb{R}^n$ )

(„ $\geq$ “ definiert Loewner-Halbordnung)

Es seien  $T : \mathbb{R}^n \rightarrow \mathbb{R}^s$  eine Statistik,  $P_\vartheta$  Verteilung auf  $\mathcal{B}^n$ ,  $V(\vartheta)$  ein  $k$ -dimensionaler Zufallsvektor mit  $E_\vartheta V(\vartheta) = 0$  und positiv definiten Kovarianzmatrix

$$J(\vartheta) = E_\vartheta[V(\vartheta) \cdot V(\vartheta)^T]$$

Definiert man

$$D(\vartheta) := E_\vartheta[V(\vartheta) \cdot (T - E_\vartheta T)^T]$$

( $k \times s$ -Matrix), so gilt<sup>18</sup>:

$$\text{Var}_\vartheta(T) \geq D^T(\vartheta) \cdot J^{-1}(\vartheta) \cdot D(\vartheta)$$

<sup>17</sup>  $A - B \geq 0$

<sup>18</sup>  $\text{Var}_\vartheta(T)$  ist Kovarianzmatrix, da  $T$  vektorwertig; im Folgenden wird diese Schreibweise bei (Zufalls-)Vektoren meistens angewandt (...)

„=“ gilt genau dann, wenn  $T = E_{\vartheta}T + D^T(\vartheta) \cdot J^{-1}(\vartheta) \cdot V(\vartheta)$   $P_{\vartheta}$ -f.s.

Beweis:

Für jeden Zufallsvektor  $Y_{k \times 1}$  gilt:

$$(i) \quad E[YY^T] \geq 0$$

$$(ii) \quad E[YY^T] = 0 \Leftrightarrow Y = 0 \text{ P-f.s.}$$

[zu (i):

$$\forall a \in \mathbb{R}^k : a^T E[YY^T] a = E[a^T YY^T a] = E[(a^T Y)^2] \geq 0$$

zu (ii): „ $\Rightarrow$ “

$$EYY^T = 0 \Rightarrow \forall j : EY_j^2 = 0 \Rightarrow Y = 0 \text{ P-f.s. } ]$$

Setze  $Y := T - E_{\vartheta}T - D^T(\vartheta) \cdot J^{-1}(\vartheta) \cdot V(\vartheta)$ .

Dann gilt:

$$\begin{aligned} 0 \stackrel{(i)}{\leq} E_{\vartheta}[YY^T] &\stackrel{(*)}{=} E_{\vartheta}[(T - E_{\vartheta}T)(T - E_{\vartheta}T)^T] \\ &\quad - \underbrace{E_{\vartheta}[(T - E_{\vartheta}T)V^T(\vartheta)]}_{=D^T(\vartheta)} J^{-1}(\vartheta) D(\vartheta) \\ &\quad - D^T(\vartheta) J^{-1}(\vartheta) \underbrace{E_{\vartheta}[V(\vartheta)(T - E_{\vartheta}T)^T]}_{=D(\vartheta)} \\ &\quad + D^T(\vartheta) J^{-1}(\vartheta) \underbrace{E_{\vartheta}[V(\vartheta) \cdot V^T(\vartheta)]}_{=J(\vartheta)} J^{-1}(\vartheta) D(\vartheta) \\ &= \text{Var}_{\vartheta}(T) - D^T(\vartheta) J^{-1}(\vartheta) D(\vartheta) \end{aligned}$$

(\*): Beachte:  $J$  symmetrisch,  $J = E_{\vartheta}[\cdot]$ ,  $D = E_{\vartheta}[\cdot]$ .

$[Y = (T - E_{\vartheta}T) - (D^T(\vartheta) \cdot J^{-1}(\vartheta) \cdot V(\vartheta))]$

„=“  $\stackrel{(ii)}{\Leftrightarrow} Y = 0$  P-f.s. ■

### 5.13 Satz (Cramér-Rao-Ungleichung)

Es seien  $\{P_\vartheta : \vartheta \in \Theta\}$  reguläre Verteilungsklasse und  $T : \mathbb{R}^n \rightarrow \mathbb{R}^s$  reguläre Statistik. Dann gilt:

$$(1) \quad \text{Var}_\vartheta(T) \geq \left(\frac{d}{d\vartheta} E_\vartheta T^T\right)^T \cdot I_n^{-1}(\vartheta) \cdot \left(\frac{d}{d\vartheta} E_\vartheta T^T\right) \quad (\vartheta \in \Theta)$$

„=“ in (1) gilt  $\Leftrightarrow T = E_\vartheta T + \left(\frac{d}{d\vartheta} E_\vartheta T^T\right)^T \cdot I_n^{-1}(\vartheta) \cdot \mathcal{U}_n(\vartheta)$

Beweis:

5.12 mit  $V(\vartheta) := \mathcal{U}_n(\vartheta)$ ,  $E_\vartheta \mathcal{U}_n(\vartheta) = 0$  (Lemma 5.8),  $J(\vartheta) = I_n(\vartheta)$ ,  
 $D(\vartheta) = E_\vartheta[\mathcal{U}_n(\vartheta)(T - E_\vartheta T)^T] = C_n(\vartheta) = \frac{d}{d\vartheta} E_\vartheta T^T$  (5.11).

### 5.14 Bemerkungen

- a) Ist  $T$  erwartungstreu für  $g(\vartheta)$ , so gilt

$$E_\vartheta T = g(\vartheta) \quad \forall \vartheta \in \Theta$$

$\Rightarrow$  rechte Seite von 5.13(1) ist nicht von  $T$  abhängig.

- b) Falls  $k = s$  und  $T$  erwartungstreu für  $\vartheta$ , so gilt  $E_\vartheta T = \vartheta \quad \forall \vartheta \in \Theta$  und somit  $\frac{d}{d\vartheta} E_\vartheta T^T = I_k \Rightarrow$

$$\text{Var}_\vartheta T \geq I_n^{-1}(\vartheta)$$

$$\text{„=“} \Leftrightarrow T = \vartheta + I_n^{-1}(\vartheta) \frac{d}{d\vartheta} \log f(X, \vartheta) \quad P_\vartheta - f.s.$$

- c) Falls  $X = (X_1, \dots, X_n)$  und  $X_1, \dots, X_n \stackrel{uiv}{\sim} f_1(\xi, \vartheta)$ , so gilt:

$$f(x, \vartheta) = \prod_{j=1}^n f_1(x_j, \vartheta)$$

$$\mathcal{U}_n(\vartheta) = \frac{d}{d\vartheta} \sum_{j=1}^n \log f_1(X_j, \vartheta) = \sum_{j=1}^n \underbrace{\frac{d}{d\vartheta} \log f_1(X_j, \vartheta)}_{\text{uiv mit } E_\vartheta(\cdot)=0}$$

$$\begin{aligned}
\Rightarrow I_n(\vartheta) &= E_\vartheta[\mathcal{U}_n(\vartheta)\mathcal{U}_n^T(\vartheta)] \\
&= \sum_{i=1}^n \sum_{j=1}^n E_\vartheta[\underbrace{\frac{d}{d\vartheta} \log f_1(X_i, \vartheta) \frac{d}{d\vartheta} \log f_1(X_j, \vartheta)^T}_{=0 \text{ für } i \neq j}] \\
&= n \cdot E_\vartheta[\underbrace{\frac{d}{d\vartheta} \log f_1(X_1, \vartheta) \cdot \frac{d}{d\vartheta} \log f_1(X_1, \vartheta)^T}_{=: I_1(\vartheta)}] \\
&= n \cdot I_1(\vartheta)
\end{aligned}$$

$\Rightarrow$  Schranke in 5.13(1) geht mit  $\frac{1}{n}$  gegen 0.

- d) Ist  $\Theta \subset \mathbb{R}^1$ ,  $T : \mathbb{R}^1 \rightarrow \mathbb{R}^1$ ,  $\gamma(\vartheta) := E_\vartheta(T)$ ,  $\vartheta \in \Theta$ ,  
 $X_1, \dots, X_n \stackrel{uiv}{\sim} f_1(\xi, \vartheta)$  wie in (c), so folgt:

$$\text{Var}_\vartheta(T) \geq \frac{(\gamma'(\vartheta))^2}{n \cdot I_1(\vartheta)}, \quad \vartheta \in \Theta$$

- e) T heißt **CR-effizient**, falls in 5.13(1) Gleichheitszeichen gilt.  
Achtung: CR-effizienter Schätzer muss nicht existieren.

## 5.15 Beispiel

$X_1, \dots, X_n \stackrel{uiv}{\sim} \text{Bin}(1, \vartheta)$ ,  $\vartheta \in \Theta = (0, 1)$ ,  $\mu$  = Zählmaß auf  $\{0, 1\}^n$ .  
 $f_1(\xi, \vartheta) = \vartheta^\xi \cdot (1 - \vartheta)^{1-\xi}$ ,  $\xi \in \{0, 1\}$

$$f(x, \vartheta) = \prod_{j=1}^n f_1(x_j, \vartheta) = \vartheta^{\sum_j x_j} (1 - \vartheta)^{n - \sum_j x_j}, \quad x \in A$$

$$\log f(x, \vartheta) = \sum_j x_j \log \vartheta + (n - \sum_j x_j) \log(1 - \vartheta)$$

$$\frac{d}{d\vartheta} \log f(x, \vartheta) = \frac{\sum_j x_j}{\vartheta} - \frac{n - \sum_j x_j}{1 - \vartheta} = \frac{\sum_j x_j - n\vartheta}{\vartheta(1 - \vartheta)}$$

$$\begin{aligned}
\Rightarrow I_n(\vartheta) &= E_\vartheta\left[\left(\frac{d}{d\vartheta} \log f(X, \vartheta)\right)^2\right] \\
&= \frac{1}{\vartheta^2(1 - \vartheta)^2} E_\vartheta\left[\underbrace{\left(\sum_{j=1}^n X_j - n\vartheta\right)^2}_{\sim \text{Bin}(n, \vartheta)}\right] \\
&= \frac{1}{\vartheta(1 - \vartheta)}
\end{aligned}$$

[Erwartungswert von  $\text{Bin}(n, \vartheta) = n\vartheta$ , also ist in der vorletzten Zeile die Varianz von  $\text{Bin}(n, \vartheta)$  gesucht.]

(1) „Raten“

$$\text{Sei } T(x) := \frac{1}{n} \sum_{j=1}^n x_j.$$

$$E_{\vartheta} T = \vartheta$$

$\Rightarrow T$  erwartungstreu

5.14(d)  $\Rightarrow$

$$\underbrace{\text{Var}_{\vartheta} T}_{= \frac{1}{n} \text{Var}_{\vartheta}(X_1) = \frac{1}{n} \vartheta(1-\vartheta)} \geq I_n^{-1}(\vartheta) = \frac{\vartheta(1-\vartheta)}{n}$$

$\Rightarrow T$  ist UMVUE

(2) Konstruktion nach 5.13 durchführen

$$T(X) \stackrel{5.14(b)}{=} \vartheta + \underbrace{\frac{\vartheta(1-\vartheta)}{n}}_{I_n(\vartheta)^{-1}} \cdot \underbrace{\frac{\sum_{j=1}^n X_j - n\vartheta}{\vartheta(1-\vartheta)}}_{\frac{d}{d\vartheta} \log f(X, \vartheta)} = \bar{X}_n$$

## 6 Exponentialfamilien

Es sei  $(\mathfrak{X}, \mathcal{B})$  Messraum,  $\mathcal{M}^1(\mathfrak{X}, \mathcal{B})$  Menge aller Wahrscheinlichkeitsmaße auf  $\mathcal{B}$ .

### 6.1 Definition

Eine Verteilungsklasse  $\wp = \{P_\vartheta : \vartheta \in \Theta\} \subset \mathcal{M}^1(\mathfrak{X}, \mathcal{B})$  heißt **Exponentialfamilie** : $\Leftrightarrow$  es existiert ein  $\sigma$ -endliches dominierendes Maß  $\mu$  auf  $\mathcal{B}$ , für ein  $k \in \mathbb{N}$  existieren  $q_1, \dots, q_k, c : \Theta \rightarrow \mathbb{R}$  und messbare Funktionen  $T_1, \dots, T_k : \mathfrak{X} \rightarrow \mathbb{R}, h : \mathfrak{X} \rightarrow \mathbb{R}_{\geq 0}$  mit

$$f(x, \vartheta) := \frac{dP_\vartheta}{d\mu}(x) = c(\vartheta) \cdot e^{\sum_{j=1}^k q_j(\vartheta) T_j(x)} \cdot h(x) \quad \mu\text{-f.ü.}$$

### 6.2 Bemerkungen

a) Mit  $q(\vartheta) := (q_1(\vartheta), \dots, q_k(\vartheta))^T$  und  $T(x) := (T_1(x), \dots, T_k(x))^T$  ist

$$f(x, \vartheta) = c(\vartheta) e^{q(\vartheta)^T T(x)} h(x)$$

b)  $c$  ist Normierungskonstante:

$$c(\vartheta) = \left[ \int e^{q(\vartheta)^T T(x)} h(x) \mu(dx) \right]^{-1} > 0$$

c) Der Träger  $\{x : f(x, \vartheta) > 0\}$  hängt nicht von  $\vartheta$  ab, insbesondere gilt

$$\forall N \in \mathcal{B} : P_{\vartheta_1}(N) = 0 \Leftrightarrow P_{\vartheta_2}(N) = 0 \quad (\vartheta_1, \vartheta_2 \in \Theta)$$

(d.h. es gilt  $P_{\vartheta_1} \ll P_{\vartheta_2}, P_{\vartheta_2} \ll P_{\vartheta_1}$ ).

d) Im Folgenden gelte immer:

- (i) Die Funktionen  $1, q_1, \dots, q_k$  sind linear unabhängig
- (ii) Die Funktionen  $1, T_1, \dots, T_k$  sind linear unabhängig auf dem Komplement jeder  $\mu$ -Nullmenge

(sogenannte (strikt) **k-parametrische Exponentialfamilie**).

Dann ist  $k$  kleinstmöglich gewählt, und  $q$  sowie  $T$  sind bis auf nicht ausgeartete affine Transformationen  $q \mapsto Aq + a, T \mapsto BT + b$  ( $\mu$ -f.ü.) eindeutig bestimmt.

### 6.3 Beispiele

- a)  $P_\vartheta := \mathcal{N}(\mu, \sigma^2)$ ,  $\vartheta := (\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}_{>0} =: \Theta$ .

Die Lebesguedichte ist

$$\begin{aligned} f(x, \vartheta) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \\ &= \underbrace{\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\mu^2}{2\sigma^2}\right)}_{=:c(\vartheta)} \exp\left(\frac{\mu}{\sigma^2}x - \frac{1}{2\sigma^2}x^2\right) \cdot \underbrace{1}_{=:h(x)} \end{aligned}$$

Mit  $q(\vartheta) := (\frac{\mu}{\sigma^2}, -\frac{1}{2\sigma^2})$ ,  $T(x) := (x, x^2)$  folgt, dass hier eine (strikt) zweiparametrische Exponentialfamilie vorliegt.

- b)  $P_\vartheta := \mathcal{N}(\vartheta, \vartheta^2)$ ,  $\vartheta \in \mathbb{R}_{>0} =: \Theta$ .

Die Lebesguedichte ist

$$\begin{aligned} f(x, \vartheta) &= \frac{1}{\sqrt{2\pi\vartheta^2}} \exp\left(-\frac{(x-\vartheta)^2}{2\vartheta^2}\right) \\ &= \underbrace{\frac{1}{\sqrt{2\pi\vartheta^2}} e^{-1/2}}_{=:c(\vartheta)} \exp\left(\frac{1}{\vartheta}x - \frac{1}{2\vartheta^2}x^2\right) \cdot \underbrace{1}_{=:h(x)} \end{aligned}$$

Mit  $q(\vartheta) := (\frac{1}{\vartheta}, -\frac{1}{2\vartheta^2})$ ,  $T(x) := (x, x^2)$  folgt wieder, dass eine (strikt) zweiparametrische Exponentialfamilie vorliegt (obwohl der Parameter-raum  $\Theta$  eindimensional ist!)

- c)  $P_\vartheta := \text{Bin}(n, \vartheta)$ ,  $\vartheta \in (0, 1) =: \Theta$ .

Die Zähl-dichte ist

$$f(x, \vartheta) = \binom{n}{x} \vartheta^x (1-\vartheta)^{n-x} = (1-\vartheta)^n \exp\left(x \log \frac{\vartheta}{1-\vartheta}\right) \binom{n}{x}.$$

Mit  $c(\vartheta) := (1-\vartheta)^n$ ,  $q(\vartheta) := \log \frac{\vartheta}{1-\vartheta}$ ,  $T(x) := x$  und  $h(x) := \binom{n}{x}$  folgt, dass  $\wp := \{\text{Bin}(n, \vartheta) : \vartheta \in \Theta\}$  eine einparametrische Exponentialfamilie ist.

- d) Die Menge aller Gleichverteilungen  $\{U(0, \vartheta), \vartheta \in \mathbb{R}_{>0}\}$  ist nach 6.2(c) keine Exponentialfamilie.

### 6.4 Satz

Es seien  $X_1, \dots, X_n \stackrel{uiv}{\sim} P_\vartheta$ , wobei  $P_\vartheta$  Element einer  $k$ -parametrischen Exponentialfamilie  $\{P_\vartheta : \vartheta \in \Theta\}$  ist. Dann gehört auch die Verteilung von  $X := (X_1, \dots, X_n)$  zu einer  $k$ -parametrischen Exponentialfamilie mit

$$q(\vartheta) \quad \text{und} \quad T_{(n)}(x) := \sum_{j=1}^n T(x_j).$$

Beweis:

Sei  $\mu^n := \mu \otimes \dots \otimes \mu$  das  $n$ -fache Produktmaß auf  $\mathcal{B}^n := \mathcal{B} \otimes \dots \otimes \mathcal{B}$  und

$$P_\vartheta^n := P_\vartheta \otimes \dots \otimes P_\vartheta$$

die Verteilung von  $X$  unter  $P_\vartheta$ . Wir erhalten mit  $x := (x_1, \dots, x_n)$ :

$$\begin{aligned} \frac{dP_\vartheta^n}{d\mu^n}(x) &= \prod_{j=1}^n \frac{dP_\vartheta}{d\mu}(x_j) \quad \mu\text{-f.ü.} \\ &= \prod_{j=1}^n [c(\vartheta) \exp(q^T(\vartheta)T(x_j)) h(x_j)] \quad \mu\text{-f.ü.} \\ &= c(\vartheta)^n \exp\left(q^T(\vartheta) \sum_{j=1}^n T(x_j)\right) \prod_{j=1}^n h(x_j) \quad \mu\text{-f.ü.} \end{aligned}$$

Bemerkung:

In der Situation von Satz 6.4 ist der ML-Schätzer  $\hat{\vartheta}_n$  für  $\vartheta$  eine Funktion von  $\sum_{j=1}^n T(X_j)$ .

In der Darstellung

$$f(x, \vartheta) = c(\vartheta) \exp(q^T(\vartheta)T(x))h(x)$$

hängt  $c(\cdot)$  von  $\vartheta$  nur über  $q := q(\vartheta) \in Q := q(\Theta) \subset \mathbb{R}^k$  ab, das heißt es gilt

$$c(\vartheta) = C(q(\vartheta))$$

für ein geeignetes  $C : Q \rightarrow \mathbb{R}$ .

$q$  heißt **natürlicher Parameter**. Somit lässt sich  $f$  ausdrücken als

$$f(x, q) = \frac{dP_q}{d\mu}(x) = C(q)e^{q^T \cdot T(x)} h(x)$$

Die Menge

$$Q_* := \{q \in \mathbb{R}^k : 0 < \int e^{q^T T(x)} h(x) \mu(dx) < \infty\}$$

heißt **natürlicher Parameterraum** der Exponentialfamilie. Es gilt

$$Q = q(\Theta) \subset Q_*.$$

### 6.5 Satz

$Q_*$  ist konvex und enthält ein nicht-ausgeartetes  $k$ -dimensionales Intervall.

Beweis:

Für  $q, r \in Q_*$  und  $\lambda \in [0, 1]$  gilt

$$\begin{aligned} 0 &< \int e^{(\lambda q^T + (1-\lambda)r^T)T} h d\mu \\ &= \int \left(e^{q^T T}\right)^\lambda \left(e^{r^T T}\right)^{1-\lambda} h d\mu \\ &\leq \int \max\left(e^{q^T T}, e^{r^T T}\right) h d\mu \\ &= \int \left(e^{q^T T} + e^{r^T T}\right) h d\mu < \infty \end{aligned}$$

Die zweite Aussage folgt dann aus der linearen Unabhängigkeit von  $1, q_1, \dots, q_k$ .

Bemerkung:

Im Folgenden setzen wir  $\vartheta := q$ , betrachten also Exponentialfamilien

$$f(x, \vartheta) = \frac{dP_\vartheta}{d\mu}(x) = C(\vartheta) e^{\vartheta^T T(x)} h(x) \quad (1)$$

mit  $\vartheta \in \Theta := \left\{ \vartheta \in \mathbb{R}^k : 0 < \int e^{\vartheta^T T(x)} h(x) \mu(dx) < \infty \right\}$ .

Weiter sei

$$b(\vartheta) := -\log C(\vartheta).$$

### 6.6 Lemma

Es sei  $\varphi : \mathfrak{X} \rightarrow \mathbb{R}$  eine messbare Abbildung mit

$$E_\vartheta |\varphi| = \int |\varphi(x)| f(x, \vartheta) \mu(dx) < \infty$$

Sei

$$A_\varphi(\vartheta) := \int \varphi(X) e^{\vartheta^T T(x)} h(x) \mu(dx), \quad \vartheta \in \Theta^0 \quad (2)$$

Dann ist  $A_\varphi : \Theta^0 \rightarrow \mathbb{R}$  beliebig oft differenzierbar und die Differentiation in (2) kann unter dem Integralzeichen vorgenommen werden beziehungsweise Integration und Differentiation können vertauscht werden.

Beweis:

Witting, 1985, S. 151f.

### 6.7 Satz

- a) Die Funktion  $b(\vartheta)$ ,  $\vartheta \in \Theta^0$ , ist beliebig oft differenzierbar.  
 b) Besitzt  $X$  die Dichte  $f(x, \vartheta)$  aus (1), so gilt:

$$E_\vartheta T(X) = \frac{d}{d\vartheta} b(\vartheta)$$

$$\text{Var}_\vartheta T(X) = \frac{d^2}{d\vartheta d\vartheta^T} b(\vartheta)$$

Beweis:

- a)  $\varphi \equiv 1$  in 6.6  $\Rightarrow A_\varphi(\vartheta) = C(\vartheta)^{-1} = e^{b(\vartheta)}$   
 6.6  $\Rightarrow$  Behauptung

b)

$$\begin{aligned} E_\vartheta T(X) &= e^{-b(\vartheta)} \int T(x) e^{\vartheta^T T(x)} h(x) \mu(dx) \\ &= e^{-b(\vartheta)} \int \frac{d}{d\vartheta} e^{\vartheta^T T(x)} h(x) \mu(dx) \\ &\stackrel{6.6}{=} e^{-b(\vartheta)} \frac{d}{d\vartheta} \underbrace{\int e^{\vartheta^T T(x)} h(x) \mu(dx)}_{=e^{b(\vartheta)}} \\ &= \frac{d}{d\vartheta} b(\vartheta) \end{aligned}$$

$$\begin{aligned}
E_{\vartheta}[T(X) \cdot T(X)^T] &= e^{-b(\vartheta)} \int T(x) e^{\vartheta^T T(x)} T(x)^T h(x) \mu(dx) \\
&= e^{-b(\vartheta)} \int \frac{d^2}{d\vartheta d\vartheta^T} e^{\vartheta^T T(x)} h(x) \mu(dx) \\
&= e^{-b(\vartheta)} \frac{d^2}{d\vartheta d\vartheta^T} e^{b(\vartheta)} \\
&= \frac{d^2}{d\vartheta d\vartheta^T} b(\vartheta) + \underbrace{\left( \frac{d}{d\vartheta} b(\vartheta) \right) \left( \frac{d}{d\vartheta} b(\vartheta) \right)^T}_{= E_{\vartheta} T(X) \cdot (E_{\vartheta} T(X))^T}
\end{aligned}$$

$$\Rightarrow \text{Var}_{\vartheta} T(X) = \frac{d^2}{d\vartheta d\vartheta^T} b(\vartheta)$$

## 6.8 CR-Effizienz in Exponentialfamilien

Seien  $X_1, \dots, X_n \stackrel{uiv}{\sim} f_1(\xi, \vartheta) = e^{-b(\vartheta)} e^{\vartheta^T T(\xi)} h(\xi)$  wie in (1).  
 $\Rightarrow X = (X_1, \dots, X_n)$  besitzt die Dichte

$$f(x, \vartheta) = e^{-nb(\vartheta)} \cdot \exp(\vartheta^T \sum_{i=1}^n T(x_j)) \prod_{j=1}^n h(x_j)$$

Sei  $S(X) = \frac{1}{n} \sum_{j=1}^n T(X_j)$ .

$$\Rightarrow E_{\vartheta} S(X) = E_{\vartheta} T(X_1) \stackrel{6.7}{=} \frac{d}{d\vartheta} b(\vartheta), \quad \vartheta \in \Theta$$

$\Rightarrow$  S erwartungstreu für  $\frac{d}{d\vartheta} b(\vartheta)$ .

Behauptung:  $S(X)$  ist CR-effizient.

Beweis:

$$\text{Var}_{\vartheta} S(X) = \frac{1}{n} \text{Var}_{\vartheta} T(X_1) \stackrel{6.7}{=} \frac{1}{n} \frac{d^2}{d\vartheta d\vartheta^T} b(\vartheta)$$

CR-Ungleichung:

$$\text{Var}_{\vartheta} S(X) \geq C_n(\vartheta)^T I_n(\vartheta)^{-1} C_n(\vartheta)$$

wobei

$$C_n(\vartheta) = \frac{d}{d\vartheta} E_{\vartheta}[S(X)^T] = \frac{d}{d\vartheta} \left[ \frac{d}{d\vartheta} b(\vartheta) \right]^T = \frac{d^2}{d\vartheta d\vartheta^T} b(\vartheta)$$

$$I_n(\vartheta) = n \cdot I_1(\vartheta) = n \cdot E_{\vartheta} \left[ \frac{d}{d\vartheta} \log f_1(X_1, \vartheta) \cdot \frac{d}{d\vartheta} \log f_1(X_1, \vartheta)^T \right]$$

$$\log f_1(X_1, \vartheta) = -b(\vartheta) + \vartheta^T T(X_1) + \log h(X_1)$$

$$\frac{d}{d\vartheta} \log f_1(X_1, \vartheta) = -\frac{d}{d\vartheta} b(\vartheta) + T(X_1) = T(X_1) - E_{\vartheta} T(X_1)$$

$$\begin{aligned}\Rightarrow I_n(\vartheta) &= n \cdot \text{Var}_{\vartheta} T(X_1) = n \cdot \frac{d^2}{d\vartheta d\vartheta^T} b(\vartheta) \\ \Rightarrow C_n(\vartheta)^T I_n(\vartheta)^{-1} C_n(\vartheta) &= \frac{1}{n} \frac{d^2}{d\vartheta d\vartheta^T} b(\vartheta)\end{aligned}$$

## 6.9 Beispiel

$$f_1(\xi, \vartheta) = \underbrace{\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\mu^2}{2\sigma^2}\right)}_{=C(\vartheta)} \exp\left(\frac{\mu}{\sigma^2} \cdot \xi - \frac{1}{2\sigma^2} \xi^2\right)$$

$$\vartheta = (\vartheta_1, \vartheta_2) := \left(\frac{\mu}{\sigma^2}, -\frac{1}{2\sigma^2}\right)$$

$$b(\vartheta) = -\log C(\vartheta) = \frac{\mu^2}{2\sigma^2} + \frac{1}{2} \log(2\pi\sigma^2) = -\frac{1}{4} \frac{\vartheta_1^2}{\vartheta_2} + \frac{1}{2} \log\left(\frac{-\pi}{\vartheta_2}\right)$$

$$\frac{d}{d\vartheta} b(\vartheta) = \left(-\frac{1}{2} \frac{\vartheta_1}{\vartheta_2}, \frac{1}{4} \frac{\vartheta_1^2}{\vartheta_2^2} - \frac{1}{2\vartheta_2}\right)^T = (\mu, \sigma^2 + \mu^2)^T$$

Fazit:

$$S(X) = \left(\frac{1}{n} \sum_{j=1}^n X_j, \frac{1}{n} \sum_{j=1}^n X_j^2\right)$$

ist erwartungstreu und CR-effizient für  $(E_{\vartheta} X_1, E_{\vartheta} X_1^2)$ .

Frage:

Ist  $S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$  CR-effizient für  $\sigma^2$ ?



## 7 Suffizienz und Vollständigkeit

### 7.1 Wiederholung

#### Bedingte Verteilungen

Sei  $(\Omega, \mathcal{A}, P)$  Wahrscheinlichkeitsraum,  $X : \Omega \rightarrow \mathbb{R}^k$ ,  $Y : \Omega \rightarrow \mathbb{R}^s$  Zufallsvektoren.

Stochastik:

Es existiert Übergangswahrscheinlichkeit  $P^{Y|X}$  mit

$$P^{(X,Y)} = P^X \otimes P^{Y|X} \quad (1)$$

$$P^{Y|X} : \begin{cases} \mathbb{R}^k \times \mathcal{B}^s \rightarrow \mathbb{R} \\ (x, B) \rightarrow P^{Y|X}(x, B) =: P^{Y|X=x}(B) \end{cases}$$

$$\begin{aligned} \text{mit } \forall x \in \mathbb{R}^k : & \quad P^{Y|X=x}(\cdot) \text{ Wahrscheinlichkeitsmaß auf } \mathcal{B}^s \\ \forall B \in \mathcal{B}^s : & \quad P^{Y|X=x}(B) \text{ } \mathcal{B}^k \text{ - messbar} \end{aligned}$$

$P^{Y|X}$  heißt (eine) bedingte Verteilung von Y bei gegebenem X.

$P^{Y|X=x}$  heißt (eine) bedingte Verteilung von Y bei gegebenem  $X = x$ .

Schreibweise:

$$P(Y \in B | X = x) := P^{Y|X=x}(B)$$

Dann (1) äquivalent zu

$$P^{(X,Y)}(A \times B) = P(X \in A, Y \in B) = \int_A P(Y \in B | X = x) P^X(dx)$$

$$\forall A \in \mathcal{B}^k, B \in \mathcal{B}^s$$

Insbesondere:

$$P(Y \in B) = \int P(Y \in B | X = x) P^X(dx)$$

Falls  $(X, Y)$  Dichte  $f(x, y)$  bezüglich  $\lambda \times \nu$  besitzt, so definiert man bedingte Dichte von Y gegeben  $X = x$  durch

$$f_{Y|X}(y|x) = \frac{f(x, y)}{f_X(x)}$$

$$f_X(x) := \int f(x, y) \nu(dy) > 0$$

Damit:

$$P(Y \in B | X = x) = \int_B f_{Y|X}(y|x) \nu(dy)$$

$$\left[ \begin{aligned} P(X \in A, Y \in B) &\stackrel{!}{=} \int_A \left[ \int_B f_{Y|X}(y|x) \nu(dy) \right] f_X(x) \lambda(dx) \\ &= \int_A \int_B f(x, y) d(\lambda \times \nu)(x, y) \end{aligned} \right]$$

## 7.2 Definition

Sei  $(\Omega, \mathcal{A}, \mathbb{P})$  Wahrscheinlichkeitsraum,  $(\mathbb{R}^n, \mathcal{B}^n, \varphi)$  statistischer Raum.

$X : \Omega \rightarrow \mathbb{R}^n$  Zufallsvektor,  $T : \mathbb{R}^n \rightarrow \mathbb{R}^s$  Statistik.

T heißt **suffizient** für  $\varphi : \Leftrightarrow P^{X|T(X)}$  hängt nicht von  $P \in \varphi$  ab.

„Die bedingte Verteilung von X gegeben T ist bekannt.“

Falls  $\varphi = \{P_\vartheta : \vartheta \in \Theta\}$ , so T **suffizient für**  $\vartheta : \Leftrightarrow P^{X|T(X)}$  hängt nicht von  $\vartheta$  ab.

## 7.3 Bemerkungen

(i) Wegen

$$\begin{aligned} \underbrace{P(X \in A, X \in B)}_{= \int_A P(X \in B | X=x) P^X(dx)} &= \int \mathbf{1}_{A \cap B}(x) P^X(dx) \\ &= \int_A \mathbf{1}_B(x) P^X(dx) \end{aligned}$$

gilt  $P^{X|X=x}(B) = \mathbf{1}_B(x)$   
 $\Rightarrow X$  suffizient für  $\varphi$

(ii) T suffizient für  $\varphi \Leftrightarrow \forall A \in \mathcal{B}^n : P(X \in A | T(X) = t)$  ist unabhängig von  $\varphi$  für alle t (im Wertebereich von T)

(iii) Sei g bijektiv,  $g, g^{-1}$  messbar. Dann:

$$T \text{ suffizient} \Leftrightarrow g(T) \text{ suffizient}$$

## 7.4 Beispiel

$X = (X_1, \dots, X_n)$ ,  $X_1, \dots, X_n \stackrel{iid}{\sim} \text{Bin}(1, \vartheta)$ ,  $\vartheta \in (0, 1)$ ,  $T(x) = \sum_{j=1}^n x_j$ .  
Sei  $t \in \{0, 1, \dots, n\}$ ,  $x \in \{0, 1\}^n$ .

$$\begin{aligned} P_\vartheta(X = x | T = t) &= \frac{P_\vartheta(X = x, T = t)}{P_\vartheta(T(x) = t)} \\ &= \begin{cases} 0 & , \sum_{j=1}^n x_j \neq t \\ \frac{P_\vartheta(X=x)}{P_\vartheta(T(x)=t)} = \frac{\prod_{j=1}^n \vartheta^{x_j} (1-\vartheta)^{1-x_j}}{\binom{n}{t} \vartheta^t (1-\vartheta)^{n-t}} = \frac{1}{\binom{n}{t}}, & \sum_{j=1}^n x_j = t \end{cases} \end{aligned}$$

Also:

$$P_\vartheta^{X|T(X)=t} = U(\{(s_1, \dots, s_n) : s_j \in \{0, 1\} \forall j, \sum_{j=1}^n s_j = t\})$$

Insbesondere ist T suffizient für  $\vartheta$ .<sup>19</sup>

7.3(ii)  $\Rightarrow$

$$P_\vartheta(X \in A) = \int \underbrace{P(X \in A | T = t)}_{\text{unabhängig von } \vartheta} P_\vartheta^T(dt)$$

Hier:

$$\begin{aligned} P_\vartheta(X = x) &= \sum_{t=0}^n P(X = x | T = t) P_\vartheta(T = t) \\ &= \sum_{t=0}^n \frac{1}{\binom{n}{t}} \mathbf{1}\{\sum_{j=1}^n x_j = t\} \cdot \binom{n}{t} \vartheta^t (1-\vartheta)^{n-t} \\ &= \vartheta^{\sum x_j} (1-\vartheta)^{n-\sum x_j} \end{aligned}$$

„In Verteilung von T ist alle Information bezüglich  $\vartheta$  enthalten.“

$\Leftrightarrow$  Datenreduktion **ohne Informationsverlust**

## 7.5 Faktorisierungssatz

In der Situation von 7.2 existiere  $\sigma$ -endliches Maß  $\mu$  auf  $\mathcal{B}^n$  mit  $P \ll \mu$   
 $\forall P \in \varphi$ . Dann sind äquivalent:

- (i) T(X) ist suffizient für  $\varphi$ .

<sup>19</sup>  $P_\vartheta^{X|T(X)=t}$  Gleichverteilung (auf Menge)

(ii)  $\exists h : \mathbb{R}^n \rightarrow \mathbb{R}_{\geq 0}$  messbar,  $\forall P \in \varphi$  existiert  $g_P : \mathbb{R}^k \rightarrow \mathbb{R}_{\geq 0}$  messbar mit

$$\frac{dP}{d\mu}(x) = g_P(T(x)) \cdot h(x), \quad x \in \mathbb{R}$$

Ist speziell  $\varphi = \{P_\vartheta : \vartheta \in \Theta\}$ ,  $f(x, \vartheta) := \frac{dP_\vartheta}{d\mu}(x)$ ,  $g(T(x), \vartheta) = g_{P_\vartheta}(T(x))$ , so gilt also:

$$T \text{ suffizient} \Leftrightarrow f(x, \vartheta) = g(T(x), \vartheta) \cdot h(x) \quad \forall x \in \mathbb{R}^n$$

Beweis:

z.B. Shao, Mathematische Statistik, S.104-106 oder Pruscha, S. 77-80

## 7.6 Beispiel (Ordnungsstatistik)

Sei  $X = (X_1, \dots, X_n)$ ,  $X_1, \dots, X_n$  unabhängig identisch verteilt mit Verteilung  $P \in \varphi$ ,  $\varphi$  die Familie aller Verteilungen auf  $\mathbb{R}$  mit Lebesgue-Dichte.

$$T(X_1, \dots, X_n) := (X_{(1)}, \dots, X_{(n)})$$

geordnete Stichprobe (Ordnungsstatistik).

$$\frac{dP^n}{d\lambda^n}(x) = \prod_{j=1}^n f(x_j) = \underbrace{\prod_{j=1}^n f(x_{(j)})}_{=g_P(T(x))} \cdot \underbrace{1}_{=h(x)}$$

$\stackrel{7.5}{\Rightarrow}$  T suffizient für  $\varphi$ .

Bemerkung:

Es gilt

$$P^{X|T(x)=(x_{(1)}, \dots, x_{(n)})} = U(\{(x_{\pi_1}, \dots, x_{\pi_n}) : (\pi_1, \dots, \pi_n) \in \mathcal{S}_n\})$$

## 7.7 Beispiel (Exponentialfamilien)

In der Situation von Satz 6.4 ist  $T_{(n)}(X)$  suffizient für  $\vartheta$ .  
[Aufgabe 21(b)]

## 7.8 Satz von Rao-Blackwell

Sei  $(\mathfrak{X}, \mathcal{B}, \{P_\vartheta : \vartheta \in \Theta \subset \mathbb{R}^s\})$  statistischer Raum,  $g : \Theta \rightarrow \mathbb{R}$ ,  $X : \Omega \rightarrow \mathfrak{X}$ ,  
 $U_g = \{S \mid S : \mathfrak{X} \rightarrow \mathbb{R} \text{ messbar, } E_\vartheta S = g(\vartheta) \forall \vartheta \in \Theta, E_\vartheta S^2 < \infty \forall \vartheta \in \Theta\}$ .

Annahme:  $U_g \neq \emptyset$

Sei  $T : \mathfrak{X} \rightarrow \mathbb{R}^k$  suffizient für  $\vartheta$ ,  $S \in U_g$ .

Sei  $\tilde{S}(X) := E[S(X)|T(X)]$ .<sup>20</sup>

Dann gilt:

$$\tilde{S} \in U_g \text{ und } \text{Var}_\vartheta \tilde{S}(X) \leq \text{Var}_\vartheta S(X) \forall \vartheta \in \Theta$$

(Verbesserung erwartungstreuer Schätzer durch suffiziente Statistiken)

Beweis:

$$E_\vartheta \tilde{S}(X) = E_\vartheta E[S(X)|T(X)] = E_\vartheta S(X) = g(\vartheta) \forall \vartheta \in \Theta$$

$$\begin{aligned} \text{Var}_\vartheta S(X) &= E_\vartheta [(S(X) - \tilde{S}(X) + \tilde{S}(X) - \underbrace{E_\vartheta S(X)}_{=g(\vartheta)})^2] \\ &= \underbrace{E_\vartheta (S(X) - \tilde{S}(X))^2}_{\geq 0} + \text{Var}_\vartheta \tilde{S}(X) \\ &\quad + 2E_\vartheta \underbrace{[E_\vartheta [(S(X) - \tilde{S}(X))(\tilde{S}(X) - g(\vartheta))|T(X)]]}_{=(\tilde{S}(X)-g(\vartheta)) \cdot \underbrace{E_\vartheta [S(X) - \tilde{S}(X)|T(X)]}_{=\tilde{S}(X)-\tilde{S}(X)=0}} \\ &\geq \text{Var}_\vartheta \tilde{S}(X) \end{aligned}$$

[Beachte:  $E_\vartheta \tilde{S}(X) = E_\vartheta S(X) = g(\vartheta)$ ; Regeln bedingter Erwartungswert<sup>21</sup>]

<sup>20</sup>Nicht von  $\vartheta$  abhängig, da T suffizient. (Sonst wäre  $\tilde{S}$  kein Schätzer!)

<sup>21</sup>insbesondere einmal ohne Auswirkung Erwartungswert in Erwartungswert eines bedingten Erwartungswertes umgeschrieben

### 7.9 Beispiel

$X_1, \dots, X_n \stackrel{iid}{\sim} U(0, \vartheta), \vartheta \in \Theta = (0, \infty), X = (X_1, \dots, X_n)$

$$S(X) = \frac{2}{n} \sum_{j=1}^n X_j$$

$$\Rightarrow E_{\vartheta} S(X) = 2E_{\vartheta} X_1 = \vartheta$$

d.h. S erwartungstreu für  $\vartheta$ .

$$\text{Var}_{\vartheta} S(X) = \frac{4}{n} \text{Var}_{\vartheta} X_1 = \frac{4}{n} \cdot \frac{\vartheta^2}{12} = \frac{\vartheta^2}{3n}$$

$$T(X) := \max_{1 \leq j \leq n} X_j$$

Wegen

$$f(x, \vartheta) = \prod_{j=1}^n \frac{1}{\vartheta} \mathbf{1}_{(0, \vartheta)}(x_j) = \underbrace{\frac{1}{\vartheta^n} \cdot \mathbf{1}_{(0, \vartheta)}(\max x_j)}_{=g(T(x), \vartheta)} \cdot \underbrace{1}_{=h(x)}$$

ist  $T(X)$  suffizient für  $\vartheta$ .

Wegen

$$P^{X_1 | \max X_j} = \frac{1}{n} \delta_{\max X_j} + \frac{n-1}{n} U(0, \max X_j)$$

folgt

$$\begin{aligned} \tilde{S}(X) &= E[S(X) | \max_j X_j] \\ &= \frac{2}{n} \sum_{i=1}^n E[X_i | \max_j X_j] \\ &= 2E[X_1 | \max_j X_j] \\ &= 2\left(\frac{1}{n} \cdot \max_j X_j + \frac{n-1}{n} \frac{\max_j X_j}{2}\right) \\ &= \frac{n+1}{n} \max_j X_j \end{aligned}$$

$$\text{Var}_{\vartheta} \tilde{S}(X) = \dots = \frac{\vartheta^2}{n(n+2)} < \text{Var}_{\vartheta} S(X) \quad \text{für } n \geq 2$$

$$\text{Var}_{\vartheta} \tilde{S}(X) = \dots = \frac{\vartheta^2}{n(n+2)} = \text{Var}_{\vartheta} S(X) \quad \text{für } n = 1$$

### 7.10 Definition

In der Situation von 7.2 heißt  $T : \mathbb{R}^n \rightarrow \mathbb{R}^k$  vollständig für  $P \in \wp$  (bzw.  $\vartheta \in \Theta$ , falls  $\wp = \{P_\vartheta : \vartheta \in \Theta\}$ ), falls gilt:

Für jede messbare Funktion  $\Psi : \mathbb{R}^k \rightarrow \mathbb{R}$  mit  $E_P[\Psi(T)] = 0 \forall P \in \wp$  (bzw.  $E_\vartheta[\Psi(T)] = 0 \forall \vartheta \in \Theta$ ) folgt  $\Psi(T) = 0$  P-f.s.  $\forall P \in \wp$  (bzw.  $P_\vartheta$ -f.s.  $\forall \vartheta \in \Theta$ ).

### 7.11 Beispiel (Fortsetzung von 7.9)

Behauptung:

$\overline{T(X)} := \max_j X_j$  vollständig

Beweis:

Sei  $\Psi : \mathbb{R} \rightarrow \mathbb{R}$  messbar.

$$E_\vartheta \Psi(T) = \int_0^\vartheta \Psi(t) \cdot \frac{n}{\vartheta} \left(\frac{t}{\vartheta}\right)^{n-1} dt = \frac{n}{\vartheta^n} \underbrace{\int_0^\vartheta \Psi(t) \cdot t^{n-1} dt}_{=: G(\vartheta)}$$

$$E_\vartheta \Psi(T) = 0 \forall \vartheta > 0 \Rightarrow G(\vartheta) = 0 \forall \vartheta > 0$$

$$\Rightarrow \Psi(\vartheta) \cdot \vartheta^{n-1} = 0 \quad \lambda^1|_{[0,\infty)}\text{-f.s.}$$

$$\Rightarrow \Psi(\vartheta) = 0 \quad \lambda^1|_{[0,\infty)}\text{-f.s.}$$

$$\Rightarrow P_\vartheta(\Psi(T) = 0) = 1$$

### 7.12 Beispiel

In einer strikt k-parametrischen Exponentialfamilie

$$f(x, \vartheta) = C(\vartheta) \cdot e^{\vartheta^T T(x)} h(x)$$

(mit natürlichem Parameterraum) ist die Statistik  $T$  vollständig.

(Beweis z.B. Shao, S.110 oder Pruscha, S.82)

Beispiel:

Sei  $X_1, X_2 \stackrel{iid}{\sim} \mathcal{N}(\vartheta, 1), \vartheta \in \mathbb{R}$ .

$T(X_1, X_2) = X_1 + X_2$  ist vollständig nach 7.12.

$S(X_1, X_2) = X_1 - X_2$  dagegen nicht!

$$T \sim \mathcal{N}(2\vartheta, 2) = P_\vartheta^T$$

$$E_\vartheta \Psi(T) = \int_{\mathbb{R}} \Psi(t) \cdot \underbrace{\varphi_{2\vartheta, 2}(t)}_{\text{Dichte NV}} dt$$

$S \sim \mathcal{N}(0, 2) = P_\vartheta^S$ ,  $\Psi(S) = S$ :

$$E_\vartheta \Psi(S) = \vartheta - \vartheta = 0 \quad \forall \vartheta \in \Theta$$

$$\not\Rightarrow \Psi(S) = X_1 - X_2 = 0 \quad P_\vartheta\text{-f.s.}$$

$\{P_\vartheta^T : \vartheta \in \mathbb{R}\} = \{\mathcal{N}(2\vartheta, 2) : \vartheta \in \mathbb{R}\}$  ist viel „reichhaltiger“ als  $\{P_\vartheta^S : \vartheta \in \mathbb{R}\} = \{\mathcal{N}(0, 2)\}$ !

### 7.13 Satz von Lehmann-Scheffé

In der Situation von 7.8 ( $U_g \neq \emptyset$ ) sei die suffiziente Statistik  $T$  auch vollständig für  $\vartheta$ . Dann existiert ein eindeutig bestimmter erwartungstreuer Schätzer für  $g(\vartheta)$  der Gestalt

$$S^*(X) = h(T(X))$$

mit einer messbaren Funktion  $h : \mathbb{R}^k \rightarrow \mathbb{R}$ .

Dieser Schätzer ist UMVUE für  $g(\vartheta)$ .

Beweis:

Sei  $S \in U_g$  und  $\tilde{S}(X) := E[S(X)|T(X)]$ .

Faktorisierungssatz des bedingten Erwartungswerts  $\Rightarrow$  es existiert  $h$  messbar mit

$$\tilde{S}(X) = h(T(X))$$

Annahme:  $\exists S_* \in U_g$  mit  $S_*(X) = h_*(T(X))$  für ein  $h_*$

$$\Rightarrow E_\vartheta[\underbrace{(h - h_*)}_{=: \Psi}(T)] = g(\vartheta) - g(\vartheta) = 0 \quad \forall \vartheta \in \Theta$$

$$\stackrel{(+)}{\Rightarrow} h = h_* \quad P_\vartheta\text{-f.s.} \quad \forall \vartheta \in \Theta$$

(+): Vollständigkeit von  $T$  ( $\Psi = h - h_* = 0$ )

$\tilde{S}(X)$  ist UMVUE für  $g(\vartheta)$ !

[Annahme:  $S_2$  „besser“ als  $\tilde{S}$

$$\Rightarrow \tilde{S}_2(X) = E[S_2(X)|T(X)]$$

„mindestens so gut“ wie  $S_2$  (Rao-Blackwell);  $\tilde{S}_2 = \tilde{S}$  wegen Eindeutigkeit]

### 7.14 Beispiel (Fortsetzung von 7.11)

$\frac{n+1}{n} \max_j X_j$  ist UMVUE für  $\vartheta$ , falls  $X_1, \dots, X_n \stackrel{uiv}{\sim} U(0, \vartheta)$ ,  $\vartheta > 0$  unbekannt.

### 7.15 Beispiel (Anwendungen von Lehmann-Scheffé)

Sei  $T$  vollständig und suffizient für  $\vartheta$ ,  $\vartheta \in \Theta$ .

Finde  $h$ , so dass  $E_{\vartheta}[h(T)] = g(\vartheta) \forall \vartheta \in \Theta$ . (Lösen!, Raten!)

Falls  $\text{Var}_{\vartheta}[h(T)] < \infty \Rightarrow h(T)$  UMVUE.

Sei  $X_1, \dots, X_n \stackrel{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$ ,  $\vartheta = (\mu, \sigma^2)$ .

(i) Aufgabe 20:

$$\text{Var}_{\vartheta}((\bar{X}_n, S_n^2)^T) = \begin{bmatrix} \frac{\sigma^2}{n} & 0 \\ 0 & \frac{2\sigma^4}{n-1} \end{bmatrix}$$

$$\frac{2\sigma^4}{n-1} > \text{CR-Schranke } \frac{2\sigma^4}{n}$$

$\Rightarrow (\bar{X}_n, S_n^2)$  nicht CR-effizient für  $\vartheta$

Aber:

$$T(X) = \left( \sum_i X_i, \sum_i X_i^2 \right)$$

suffizient und vollständig für  $\bar{\vartheta} = \left( \frac{\mu}{\sigma^2}, -\frac{1}{2\sigma^2} \right)$  (nach 7.12).

$\Rightarrow T(X) = \left( \sum_i X_i, \sum_i X_i^2 \right)$  suffizient und vollständig für  $\vartheta = (\mu, \sigma^2)$ .

Sei  $h(T(X)) = (\bar{X}_n, S_n^2)$ .

$$\left. \begin{array}{l} E_{\vartheta}[h(T(X))] = \vartheta \forall \vartheta \\ \text{Var}_{\vartheta}[h(T(X))] \text{ existiert } \forall \vartheta \end{array} \right\} \Rightarrow (\bar{X}_n, S_n^2) \text{ ist UMVUE für } \vartheta = (\mu, \sigma^2)$$

Bemerkung:

Auch  $(\bar{X}_n, S_n^2)$  suffizient und vollständig für  $\vartheta$  nach Bemerkung 7.3(ii) und analoge Aussage für Vollständigkeit.

(ii) Analog:

Der Schätzer aus Aufgabe 9 der Form  $\sqrt{c_n S_n^2}$  ist UMVUE für  $\sigma$ .

(iii) Gesucht: UMVUE für  $\frac{\mu}{\sigma}$

$$(T_1(X), T_2(X)) := \left( \sum_i X_i, \sum_i (X_i - \bar{X}_n)^2 \right)$$

$T_1(X), T_2(X)$  unabhängig,  $\frac{T_1(X)}{\sigma^2} \sim \chi_{n-1}^2 = \Gamma\left(\frac{n-1}{2}, \frac{1}{2}\right)$

$$\Rightarrow E_{\vartheta} T_2^{-\frac{1}{2}} = \frac{1}{\sigma} \cdot \frac{\Gamma\left(\frac{n}{2} - 1\right)}{\sqrt{2}\Gamma\left(\frac{n-1}{2}\right)} \quad (n \geq 3)$$

$$\text{Var}_{\vartheta} T_2^{-\frac{1}{2}} < \infty \text{ für } n \geq 4$$

$$\begin{aligned}\Rightarrow E_{\vartheta}\left(\frac{T_1}{\sqrt{T_2}}\right) &= E_{\vartheta}T_1 \cdot E_{\vartheta}T_2^{-\frac{1}{2}} \\ &= \frac{\mu}{\sigma} \cdot \frac{n\Gamma(\frac{n}{2} - 1)}{\sqrt{2}\Gamma(\frac{n-1}{2})} \\ &=: \frac{\mu}{\sigma}K_n\end{aligned}$$

( $n \geq 3$ )

$$\Rightarrow K_n^{-1} \cdot \frac{T_1}{\sqrt{T_2}}$$

ist UMVUE für  $\frac{\mu}{\sigma}$  für  $n \geq 4$ .

## 8 Asymptotik von Schätzfehlern

### 8.1 Problemstellung

Seien  $X_1, X_2, \dots, X_n \stackrel{uiv}{\sim} P_\vartheta$ , mit  $\vartheta \in \Theta \subset \mathbb{R}^k$ .

Die Schätzfolge  $\hat{\vartheta}_n = \hat{\vartheta}_n(X_1, \dots, X_n)$  sei konsistent, es gilt also

$$\hat{\vartheta}_n \xrightarrow{P_\vartheta} \vartheta \text{ für } n \rightarrow \infty \quad \forall \vartheta \in \Theta$$

Sei  $(a_n)$  eine reelle Folge mit  $a_n > 0 \quad \forall n$  und  $a_n \rightarrow \infty$  für  $n \rightarrow \infty$ .

Die Folge  $(\hat{\vartheta}_n)_{n \geq 1}$  heißt  $a_n$ -**konsistent**, wenn

$$a_n(\hat{\vartheta}_n - \vartheta) = O_{P_\vartheta}(1) \quad \forall \vartheta \in \Theta.$$

Hierbei bedeutet  $Y_n = O_P(1)$  für eine Folge  $(Y_n)$ , dass für jedes  $\varepsilon > 0$  eine kompakte Menge  $K \subset \mathbb{R}^d$  existiert, so dass  $P(Y_n \in K) \geq 1 - \varepsilon$  für alle  $n \in \mathbb{N}$ .<sup>22</sup>

Typischerweise liegt  $\sqrt{n}$ -Konsistenz vor, d.h. es gilt

$$\sqrt{n}(\hat{\vartheta}_n - \vartheta) = O_{P_\vartheta}(1) \quad \forall \vartheta \in \Theta.$$

Zusätzlich kann man oftmals Aussagen über Konvergenz in Verteilung machen, insbesondere

$$\sqrt{n}(\hat{\vartheta}_n - \vartheta) \xrightarrow{D_\vartheta} \mathcal{N}_k(0, \Sigma(\vartheta)), \quad \vartheta \in \Theta.$$

### 8.2 Multivariater Zentraler-Grenzwert-Satz (ZGWS)

Seien  $Y_1, Y_2, \dots \stackrel{uiv}{\sim} Y$  mit einer  $\mathbb{R}^d$ -wertigen Zufallsvariablen  $Y$  mit  $E\|Y\|^2 < \infty$ . Mit  $a := EY$  und  $\Sigma := E(Y - a)(Y - a)^T$ , gilt

$$\frac{1}{\sqrt{n}} \left( \sum_{j=1}^n Y_j - na \right) \xrightarrow{D} \mathcal{N}_d(0, \Sigma).$$

<sup>22</sup>vergleiche Stochastik II: Straffheit

### 8.3 $\delta$ -Methode

Seien  $Z_1, Z_2, \dots$  d-dimensionale Zufallsvariablen mit

$$\sqrt{n}(Z_n - a) \xrightarrow{D} \mathcal{N}_d(0, \Sigma),$$

mit  $a := (a_1, \dots, a_d) \in \mathbb{R}^d$  und  $\Sigma \in \mathbb{R}^{d \times d}$ .

Sei  $g := (g_1, \dots, g_s)^T : \mathbb{R}^d \rightarrow \mathbb{R}^s$  differenzierbar und

$$\frac{dg}{da} := \left( \frac{\partial g_j}{\partial a_k} \right)_{\substack{1 \leq j \leq s, \\ 1 \leq k \leq d}}$$

dann gilt

$$\sqrt{n}(g(Z_n) - g(a)) \xrightarrow{D} \mathcal{N}_s \left( 0, \frac{dg}{da} \Sigma \left( \frac{dg}{da} \right)^T \right).$$

Beweis:

Nach der Definition der Differenzierbarkeit gilt

$$\sqrt{n}(g(Z_n) - g(a)) = \underbrace{\frac{dg}{da} \sqrt{n}(Z_n - a)}_{=: U_n} + \underbrace{\|\sqrt{n}(Z_n - a)\| \cdot r(Z_n - a)}_{=: V_n},$$

mit  $r(Z_n - a) \rightarrow 0$  für  $Z_n \rightarrow a$ .

Beachte, dass  $\|\sqrt{n}(Z_n - a)\| \in O_p(1)$ .

Aus  $Z_n \xrightarrow{P} a$  folgt, dass  $r(Z_n - a) \xrightarrow{P} 0$ , und somit

$$V_n \xrightarrow{P} 0.$$

Aus der Voraussetzung folgt mit dem Abbildungssatz weiter

$$U_n \xrightarrow{D} \frac{dg}{da} \cdot T$$

mit  $T \sim \mathcal{N}_d(0, \Sigma)$  und somit

$$U_n \xrightarrow{D} \mathcal{N}_s \left( 0, \frac{dg}{da} \Sigma \left( \frac{dg}{da} \right)^T \right).$$

Die Behauptung folgt schließlich aus dem Lemma von Slutsky.

### 8.4 Asymptotik des Momentenschätzers (vgl. 4.8)

$X_1, \dots, X_n \stackrel{i.i.d.}{\sim} X$ ,  $X$   $\mathbb{R}^1$ -wertig,  $P^X \in \{P_\vartheta : \vartheta \in \Theta\}$ ,  $\Theta \subset \mathbb{R}^k$ ,  
 $\vartheta = (\vartheta_1, \dots, \vartheta_n)$

Sei  $m_l := EX^l$ ,  $\hat{m}_l = \frac{1}{n} \sum_{i=1}^n X_i^l$  ( $\hat{=} \bar{X}_n^l$  aus 4.8)

Voraussetzung:

$\vartheta = g(m_1, \dots, m_k)$  mit  $g : \mathbb{R}^k \rightarrow \mathbb{R}^k$

Momentenschätzer:  $\hat{\vartheta} = g(\hat{\vartheta}_1, \dots, \hat{\vartheta}_k)$

Sei

$$Y_j := \begin{pmatrix} X_j \\ \vdots \\ X_j^k \end{pmatrix} \quad Y := \begin{pmatrix} X \\ \vdots \\ X^k \end{pmatrix}, \quad a := EY = \begin{pmatrix} m_1 \\ \vdots \\ m_k \end{pmatrix}$$

$$E \|Y\|^2 < \infty \Leftrightarrow EX^{2k} < \infty$$

$$\begin{aligned} \Sigma &:= E[(Y - a)(Y - a)^T] = (E[(X^i - m_i)(X^j - m_j)^T])_{i,j=(1,\dots,k)} \\ &= (EX^{i+j} - m_i m_j)_{i,j} \\ &= (m_{i+j} - m_i m_j)_{i,j} \end{aligned}$$

8.2  $\Rightarrow$

$$\begin{aligned} \frac{1}{\sqrt{n}} \left( \sum_{j=1}^n Y_j - na \right) &= \frac{1}{\sqrt{n}} \left( \begin{pmatrix} \sum_j X_j \\ \vdots \\ \sum_j X_j^k \end{pmatrix} - n \cdot \begin{pmatrix} m_1 \\ \vdots \\ m_k \end{pmatrix} \right) \\ &= \sqrt{n} \cdot \begin{pmatrix} \hat{m}_1 - m_1 \\ \vdots \\ \hat{m}_k - m_k \end{pmatrix} \xrightarrow{D_\vartheta} \mathcal{N}(0, \Sigma(\vartheta)) \end{aligned}$$

Aus 8.3 folgt: Falls  $EX^{2k} < \infty$  und  $g$  differenzierbar, so gilt:

$$\sqrt{n}(\hat{\vartheta}_n - \vartheta) \xrightarrow{D_\vartheta} \mathcal{N}_k(0, \frac{dg}{da} \Sigma(\vartheta) (\frac{dg}{da})^T)$$

Achtung:  $\Sigma$  hängt von  $m_1, \dots, m_k$  und somit von unbekanntem  $\vartheta$  ab.

(Schreibweise „asymptotisch normalverteilt“:)

$$\sqrt{n}(\hat{\vartheta}_n - \vartheta) \approx \mathcal{N}_k(0, T) \Leftrightarrow \hat{\vartheta}_n \approx \mathcal{N}_k(\vartheta, \frac{T}{n}), \quad \hat{\vartheta}_n \sim AN(\vartheta, \frac{T}{n}), \quad \hat{T} = T(\hat{\vartheta}_n)$$

### 8.5 Asymptotik des ML-Schätzers

$X_1, \dots, X_n \stackrel{uiv}{\sim} f_1(\xi, \vartheta)$  (Dichte bezüglich dominierendem Maß  $\mu$ )  
 $\vartheta \in \Theta \subset \mathbb{R}^k$ ,  $\Theta$  offen

Regularitätsvoraussetzungen: (a)-(e) aus 5.7- 5.9 seien erfüllt.

Zusätzlich gelte:

$\{\xi : f_1(\xi, \vartheta) > 0\}$  ist unabhängig von  $\vartheta!$

$\forall i, j, l \in \{1, \dots, k\}$  existiert  $\frac{\partial^3 \log f_1(\xi, \vartheta)}{\partial \vartheta_i \partial \vartheta_j \partial \vartheta_l} = L_{ijl}(\xi, \vartheta)$

$\forall \vartheta \in \Theta \forall \delta > 0 \forall i, j, l \in \{1, \dots, k\}$  existiert eine Funktion  $M_{i,j,l}(\xi) \geq 0$  mit

$$|L_{i,j,l}(\xi, \eta)| \leq M_{i,j,l}(\xi), \quad \|\eta - \vartheta\| \leq \delta$$

und  $E_{\vartheta} M_{i,j,l}(X_1) < \infty$

Sei

$$\mathcal{U}_n(\vartheta) := \sum_{j=1}^n \frac{d}{d\vartheta} \log f_1(X_j, \vartheta), \quad E_{\vartheta} \mathcal{U}_n(\vartheta) = 0$$

$$I_n(\vartheta) = E[\mathcal{U}_n(\vartheta) \mathcal{U}_n(\vartheta)^T] = nI_1(\vartheta)$$

$$W_n(\vartheta) = \frac{d}{d\vartheta^T} \mathcal{U}_n(\vartheta), \quad E_{\vartheta}[W_n(\vartheta)] = -I_n(\vartheta)$$

#### 8.5.1 Satz

Es gelte  $\mathcal{U}_n(\hat{\vartheta}_n) = 0$  (d.h.  $\hat{\vartheta}_n$  ist Lösung der Likelihood-Gleichung  $\mathcal{U}_n(\vartheta) = 0$ ) und  $\hat{\vartheta}_n \xrightarrow{P_{\vartheta}} \vartheta$ ,  $\vartheta \in \Theta$  (d.h.  $(\hat{\vartheta}_n)_n$  ist konsistent). Dann folgt:

$$\sqrt{n}(\hat{\vartheta}_n - \vartheta) \xrightarrow{D_{\vartheta}} \mathcal{N}_k(0, I_1(\vartheta)^{-1})$$

$$\Leftrightarrow \hat{\vartheta}_n \sim AN\left(\vartheta, \frac{I_n^{-1}}{n}\right)$$

Beweisskizze:

$$\begin{aligned} 0 &= \frac{1}{\sqrt{n}} \mathcal{U}_n(\hat{\vartheta}_n) \\ &\stackrel{\text{Taylor}}{=} \frac{1}{\sqrt{n}} \mathcal{U}_n(\vartheta) + \frac{1}{\sqrt{n}} W_n(\vartheta)(\hat{\vartheta}_n - \vartheta) + \underbrace{\frac{1}{\sqrt{n}} R_n(\vartheta, \hat{\vartheta}_n - \vartheta)}_{z.z.: = o_{P_{\vartheta}}(1)} \end{aligned}$$

$$\begin{aligned} \Rightarrow \underbrace{\frac{1}{n} W_n(\vartheta)}_{\xrightarrow{P_{\vartheta}-f.s.} -I_1(\vartheta)(SGGZ)} \sqrt{n}(\hat{\vartheta}_n - \vartheta) &= \underbrace{-\frac{1}{\sqrt{n}} \mathcal{U}_n(\vartheta)}_{\xrightarrow{D_{\vartheta}} \mathcal{N}_k(0, I_1(\vartheta)) \text{ (ZGWS 8.2)}} + o_{P_{\vartheta}}(1) \quad (*) \\ &\underbrace{\hspace{10em}}_{\xrightarrow{D_{\vartheta}} \mathcal{N}_k(0, I_1(\vartheta))} \end{aligned}$$

$$\Rightarrow \sqrt{n}(\hat{\vartheta}_n - \vartheta) \xrightarrow{D_{\vartheta}} \mathcal{N}_k(0, -I_1(\vartheta)^{-1}I_1(\vartheta)(-I_1(\vartheta)^{-1}))$$

(z.B. Knight, 249 oder Lehmann/Casella, 443-468)<sup>23</sup>

Bemerkung: (asymptotische Linearisierbarkeit des Schätzfehlers)

$$\begin{aligned} (*) \Rightarrow \sqrt{n}(\hat{\vartheta}_n - \vartheta) &= I_1(\vartheta)^{-1} \frac{1}{\sqrt{n}} \mathcal{U}_n(\vartheta) + o_{P_{\vartheta}}(1) \\ &= \frac{1}{\sqrt{n}} \sum_{j=1}^n \underbrace{I_1(\vartheta)^{-1} \frac{d}{d\vartheta} \log f_1(X_j, \vartheta)}_{=: \tilde{l}(X_j, \vartheta)} + o_{P_{\vartheta}}(1) \end{aligned}$$

mit  $E_{\vartheta} \tilde{l}(X_1, \vartheta) = 0$

### 8.5.2 Satz

Unter den obigen Voraussetzungen existiert eine Folge  $\hat{\vartheta}_n = \hat{\vartheta}_n(X_1, \dots, X_n)$  mit:

Ist  $\vartheta_0$  der wahre Parameter, so gilt:

$$\lim P_{\vartheta_0}(\mathcal{U}_n(\hat{\vartheta}_n) = 0, |\hat{\vartheta}_n - \vartheta_0| \leq \varepsilon) = 1 \quad \forall \varepsilon > 0$$

#### Korollar

Besitzt die Likelihood-Gleichung  $\mathcal{U}_n(\vartheta) = 0$  für jedes  $n$  eine eindeutige Lösung  $\hat{\vartheta}_n$ , so gilt:

$$\hat{\vartheta}_n \xrightarrow{P_{\vartheta_0}} \vartheta, \quad \vartheta \in \Theta$$

Anmerkung:

(1)  $\{P_{\vartheta} : \vartheta \in \Theta\}$  mit Dichte  $f(x, \vartheta)$  bzgl. dem Maß  $\mu$ . Dann  $\forall \vartheta \neq \vartheta_0$ :

$$\begin{aligned} E_{\vartheta_0} \left[ \log \frac{f(X, \vartheta)}{f(X, \vartheta_0)} \right] &\stackrel{\text{Jensensche Ungl.}}{<} \log E_{\vartheta_0} \left[ \frac{f(X, \vartheta)}{f(X, \vartheta_0)} \right] \\ &= \underbrace{\log \int f(x, \vartheta) dx}_{=1} < 0 \end{aligned}$$

$$\Leftrightarrow E_{\vartheta_0}[\log f(X, \vartheta_0)] > E_{\vartheta_0}[\log f(X, \vartheta)] \quad \forall \vartheta \neq \vartheta_0$$

d.h.  $\vartheta_0$  maximiert  $E_{\vartheta_0}[\log f(X, \vartheta)]$  bezüglich  $\vartheta$ !

<sup>23</sup> $o_{P_{\vartheta}}(1)$  bedeutet stochastische Konvergenz gegen 0

- (2) Funktional in (\*) ist nicht auswertbar, da  $\vartheta_0$  unbekannt!

Aber:

$$\frac{1}{n} \underbrace{\sum_{i=1}^n \log f(X_i, \vartheta)}_{l(X, \vartheta), \text{ „Log-Likelihood Funktion“}} \xrightarrow{P_\vartheta\text{-f.s.}} E_{\vartheta_0}[\log f(X, \vartheta)] \forall \vartheta \in \Theta$$

Maximierung von  $l(X, \vartheta)$  als „Ersatz“ für (\*).

- (3) f, g  $\mu$ -Dichten:

$$E_f \left[ \log \frac{f(X)}{g(X)} \right] \geq 0$$

“=”  $\Leftrightarrow f = g$

„Entropie“ zwischen f und g, Kullbach-Leibler-Information von g bezüglich f, Kullbach-Leibler-Abstand zwischen f und g

- (4) Was tun, falls Lösung von  $\mathcal{U}_n(\vartheta) = 0$  nicht eindeutig?

- (i) Oft ist die Folge von globalen Maxima konsistent.  
(Theorie von Wald 1949, Le Cam 1953)
- (ii) Sei  $(\delta_n)$  konsistent. Wähle Folge  $\vartheta_n^*$ , die am nächsten zu  $\delta_n$  liegt  $\Rightarrow (\vartheta_n^*)$  konsistent, 8.5.1 anwendbar.
- (iii) 1-Schritt-MLE verwenden:  
 $(\vartheta_n^{(0)})$  sei  $\sqrt{n}$ -konsistent. Mache einen Newton-Schritt zur Lösung von  $\mathcal{U}_n(\vartheta) = 0$ :

$$\vartheta_n^{(1)} = \vartheta_n^{(0)} - \frac{\mathcal{U}'_n(\vartheta_n^{(0)})}{\mathcal{U}''_n(\vartheta_n^{(0)})}$$

Dann hat  $(\vartheta_n^{(1)})_{n \geq 1}$  dasselbe asymptotische Verhalten wie in 8.5.1.

## 9 Robuste Schätzer

Seien  $X_1, \dots, X_n, X_{n+1} \stackrel{uiv}{\sim} F$ ,  $F \in \mathfrak{F}$ : Verteilungsannahme,  $x_1, \dots, x_n, x$  Realisierungen von  $X_1, \dots, X_n, X_{n+1}$ ,  $X_i$  reellwertig.

Sei  $\vartheta : \mathfrak{F} \rightarrow \mathbb{R}$ ,  $\hat{\vartheta}_n = \vartheta(\hat{F}_n)$  Plug-In-Schätzer für  $\vartheta(F)$ .  
 $(\hat{F}_n(t) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}\{X_i \leq t\})$

### 9.1 Definition (Sensitivitätskurve)

$$S(x, \hat{\vartheta}) = \frac{\hat{\vartheta}_{n+1} - \hat{\vartheta}_n}{\frac{1}{n+1}}$$

Dabei:  $\hat{\vartheta}_{n+1} = \vartheta(\hat{F}_{n+1})$  basierend auf  $X_1, \dots, X_n$  und einer zusätzlichen Beobachtung  $x$ .

$S(x, \hat{\vartheta})$  ist die Änderung von  $\hat{\vartheta}$  bei einer zusätzlichen Beobachtung  $x$  relativ gesehen zur Masse  $\frac{1}{n+1}$  von  $x$ .

Beispiele:

a)  $\vartheta(F) = \int x dF(x)$ ,  $\vartheta(\hat{F}_n) = \bar{x}_n$

$$S(x, \hat{\vartheta}) = \frac{\bar{x}_{n+1} - \bar{x}_n}{\frac{1}{n+1}} = \sum_{i=1}^n x_i + x - \frac{n+1}{n} \sum_{i=1}^n x_i = x - \bar{x}_n$$

linear in  $x \Rightarrow$  unbeschränkt in  $x$

Große Änderung von  $S$ , falls  $|x|$  groß!

b) Sei  $\mathfrak{F} = \{F : F \text{ streng monoton wachsend auf } \{x : 0 < F(x) < 1\}\}$ ,  
 $\vartheta(F) = F^{-1}(\frac{1}{2})$ .  
 Sei  $n = 2r - 1$  ungerade.

$$\Rightarrow \vartheta(\hat{F}_n) = x_{(r)} =: x_{r:n}$$

(„das  $r$  kleinste unter  $n$ “)

$$n + 1 = 2r:$$

$$\hat{F}_{n+1}^{-1}(\frac{1}{2}) = x_{r:n+1}$$

$$\hat{\vartheta}_{n+1} = \vartheta(\hat{F}_{n+1}) \in [x_{(r-1)}, x_{(r)}]$$

$\Rightarrow S$  beschränkt in  $x$ !

Nachteil der Sensitivitätskurve:

Hängt von Stichprobe ab.

Wünschenswert wäre Abhängigkeit nur von  $x$  und  $F$ .

## 9.2 Definition

a) Sei  $\Delta_x$  die zum Dirac-Maß in  $x$  gehörende Verteilungsfunktion, also

$$\Delta_x(y) = \begin{cases} 0, & y < x \\ 1, & y \geq x \end{cases}$$

Die Einflusskurve (*influence curve*) von  $\vartheta(F)$  ist

$$\begin{aligned} \varphi(x, F) &= \lim_{t \rightarrow 0} \frac{\vartheta((1-t)F + t\Delta_x) - \vartheta(F)}{t} \\ &= \frac{d}{dt} \vartheta((1-t)F + t\Delta_x)|_{t=0} \end{aligned}$$

wobei die Existenz der Ableitung vorausgesetzt wird.

b)  $\hat{\vartheta} = \vartheta(\hat{F}_n)$  heißt **robust**, falls  $\varphi(x, F)$  beschränkt ist in  $x$ .

Bemerkung:

Gegeben:

Stichprobe  $x_1, \dots, x_n$ : Schätze  $\vartheta(F)$  durch  $\hat{\vartheta}_n = \vartheta(\hat{F}_n)$ .

Weiterer Wert  $x$ : Schätze  $\vartheta(F)$  durch  $\hat{\vartheta}_{n+1} = \vartheta(\hat{F}_{n+1})$ , wobei

$$\hat{F}_{n+1}(y) = \frac{n}{n+1} \hat{F}_n(y) + \frac{1}{n+1} \Delta_x(y)$$

Sei nun  $t = \frac{1}{n+1}$ , also  $1-t = \frac{n}{n+1}$ . Damit gilt:

$$\begin{aligned} \hat{\vartheta}_{n+1} = \vartheta(\hat{F}_{n+1}) &= \vartheta((1-t)\hat{F}_n + t\Delta_x) \\ &= \frac{\vartheta((1-t)\hat{F}_n + t\Delta_x) - \vartheta(\hat{F}_n)}{t} t + \underbrace{\vartheta(\hat{F}_n)}_{=\hat{\vartheta}_n} \\ &\approx \hat{\vartheta}_n + \frac{1}{n+1} \varphi(x, \hat{F}_n) \end{aligned}$$

(Diese Approximation setzt voraus, dass  $\varphi(x, \hat{F}_n)$  existiert.)

In diesem Fall gilt:

$$\varphi(x, \hat{F}_n) \approx \frac{\hat{\vartheta}_{n+1} - \hat{\vartheta}_n}{\frac{1}{n+1}} = S(x, \hat{\vartheta})$$

### 9.3 Beispiel

Sei  $\vartheta(F) = \int y dF(y)$ .

$$\begin{aligned} \Rightarrow \varphi(x, F) &= \lim_{t \rightarrow 0} \frac{1}{t} [(1-t) \int y dF(y) + t \cdot \int y d\Delta_x(y) - \int y dF(y)] \\ &= - \int y dF(y) + x \\ &= x - \vartheta(F) \end{aligned}$$

Hier gilt sogar<sup>24</sup>:  $\varphi(x, \hat{F}_n) = x - \bar{x}_n = S(x, \hat{\vartheta})$ .

### 9.4 Satz (Eigenschaften von $\varphi(x, F)$ )

Sei  $\varphi(x, F)$  Einflusskurve von  $\vartheta(F)$ .

- Sei  $\vartheta(F) = \int h dF = Eh(X)$ , wobei  $X \sim F$  und  $E|h(X)| < \infty$ .  
Dann gilt:  $\varphi(x, F) = h(x) - \vartheta(F)$
- Sei  $\vartheta(F) = \vartheta_1(F) + \vartheta_2(F)$  mit Einflusskurven  $\varphi_1(x, F), \varphi_2(x, F)$ .  
Dann:  $\varphi(x, F) = \varphi_1(x, F) + \varphi_2(x, F)$
- Sei  $I \subset \mathbb{R}$ ,  $\vartheta(F) = \int_I g(s) \varphi_s(x, F) ds$ .  
Ist  $\varphi_s(x, F)$  die Einflusskurve von  $\vartheta_s(F)$  ( $s \in I$ ), so gilt (unter Regularität<sup>25</sup>):

$$\varphi(x, F) = \int_I g(s) \varphi_s(x, F) ds$$

- (Kettenregel)  
Ist  $g$  differenzierbar, so ist die Einflusskurve von  $g(\vartheta(F))$  gegeben durch

$$g'(\vartheta(F)) \cdot \varphi(x, F)$$

- (implizit definierter Parameter)  
 $\vartheta(F)$  sei Lösung der Gleichung  $h(F, \vartheta(F)) = 0$ , wobei für festes  $u$   $\lambda(x, F, u)$  die Einflusskurve von  $h(F, u)$  sei und die Ableitung  $h'(F, u)$  nach  $u$  existiere. Dann gilt:

$$\varphi(x, F) = - \frac{\lambda(x, F, \vartheta(F))}{h'(F, \vartheta(F))}$$

<sup>24</sup>vergleiche 9.1, Beispiel (a)

<sup>25</sup>siehe Beweis

Beweis:

Sei  $F_{t,x} = (1-t)F + t\Delta_x$ .

a) Aus

$$\vartheta(F_{t,x}) = (1-t) \int h(y) dF(y) + t \cdot h(x)$$

(vergleiche 9.3) folgt:

$$\frac{1}{t}(\vartheta(F_{t,x}) - \vartheta(F)) = h(x) - \vartheta(F)$$

b) Klar.

c)

$$\begin{aligned} \varphi(x, F) &= \frac{d}{dt} \vartheta(F_{t,x})|_{t=0} \\ &= \frac{d}{dt} \int_I g(s) \vartheta_s(F_{t,x}) ds |_{t=0} \\ &\stackrel{(*)}{=} \int_I g(s) \frac{d}{dt} \vartheta_s(F_{t,x})|_{t=0} ds \\ &= \int_I g(s) \varphi_s(x, F) ds \end{aligned}$$

(\*): Vertauschbarkeit vorausgesetzt! (Regularität)

d)

$$\begin{aligned} \frac{1}{t}(g(\vartheta(F_{t,x})) - g(\vartheta(F))) &= \frac{g(\vartheta(F_{t,x})) - g(\vartheta(F))}{\underbrace{\vartheta(F_{t,x}) - \vartheta(F)}} \\ &\stackrel{t \rightarrow 0}{\rightarrow} g'(\vartheta(F)), \text{ da } \vartheta(F_{t,x}) \stackrel{t \rightarrow 0}{\rightarrow} \vartheta(F) \\ &\quad \cdot \frac{\vartheta(F_{t,x}) - \vartheta(F)}{\underbrace{t}} \\ &\quad \stackrel{t \rightarrow 0}{\rightarrow} \varphi(x, F) \\ &\stackrel{t \rightarrow 0}{\rightarrow} g'(\vartheta(F)) \varphi(x, F) \end{aligned}$$

e)

$$\begin{aligned}
0 &= \frac{1}{t} \underbrace{[h(F_{t,x}, \vartheta(F_{t,x})) - h(F, \vartheta(F))]}_{=0} \\
&= \frac{1}{t} [h(F_{t,x}, \vartheta(F_{t,x})) - h(F_{t,x}, \vartheta(F))] \\
&\quad + \frac{1}{t} \underbrace{[h(F_{t,x}, \vartheta(F)) - h(F, \vartheta(F))]}_{\xrightarrow{t \rightarrow 0} \lambda(x, F, \vartheta(F))} \\
&= \underbrace{\frac{h(F_{t,x}, \vartheta(F_{t,x})) - h(F_{t,x}, \vartheta(F))}{h(F, \vartheta(F_{t,x})) - h(F, \vartheta(F))}}_{\xrightarrow{t \rightarrow 0} 1 \text{ (Forderung)}} \cdot \underbrace{\frac{h(F, \vartheta(F_{t,x})) - h(F, \vartheta(F))}{\vartheta(F_{t,x}) - \vartheta(F)}}_{\xrightarrow{t \rightarrow 0} h'(F, \vartheta(F))} \\
&\quad \cdot \underbrace{\frac{\vartheta(F_{t,x}) - \vartheta(F)}{t}}_{\xrightarrow{t \rightarrow 0} \varphi(x, F)} + \frac{1}{t} [h(F_{t,x}, \vartheta(F)) - h(F, \vartheta(F))] \\
&\xrightarrow{t \rightarrow 0} h'(F, \vartheta(F)) \cdot \varphi(x, F) + \lambda(x, F, \vartheta(F))
\end{aligned}$$

Also:

$$h'(F, \vartheta(F)) \cdot \varphi(x, F) + \lambda(x, F, \vartheta(F)) = 0$$

$$h' \neq 0 \xrightarrow{\text{(Forderung)}} \text{Behauptung.}$$

## 9.5 Bemerkung (Einflusskurven-Heuristik)

Sei  $\varphi(x, F)$  Einflusskurve von  $\vartheta(F)$ ,  $X \sim F$ 

Oft gilt:

$$(i) \quad E[\varphi(X, F)] = \int \varphi(x, F) dF(x) = 0$$

$$(ii) \quad \vartheta(\hat{F}_n) - \vartheta(F) = \int \varphi(x, F) d(\hat{F}_n(x) - F(x)) + R_n, \text{ wobei } \sqrt{n}R_n \xrightarrow{n \rightarrow \infty} 0$$

*[wird oft als erfüllt angenommen]*

$$(iii) \quad 0 < \tau^2(F) = E[\varphi^2(X, F)] < \infty$$

Dann gilt:

$$\sqrt{n}(\hat{\vartheta}_n - \vartheta) = \sqrt{n}(\vartheta(\hat{F}_n) - \vartheta(F)) \xrightarrow{D} \mathcal{N}(0, \tau^2(F))$$

Beweis:

Mit (i) und (ii) gilt:

$$\begin{aligned}\vartheta(\hat{F}_n) - \vartheta(F) &= \frac{1}{n} \sum_{i=1}^n \varphi(X_i, F) + R_n \\ \Rightarrow \sqrt{n}(\hat{\vartheta}_n - \vartheta) &= \underbrace{\frac{1}{\sqrt{n}} \sum_{i=1}^n \varphi(X_i, F)}_{\stackrel{D}{\rightarrow} \mathcal{N}(0, \tau^2(F))} + \underbrace{\sqrt{n}R_n}_{\xrightarrow{P} 0}\end{aligned}$$

Lemma von Slutsky  $\Rightarrow$  Behauptung

## 9.6 Beispiel (Median)

Sei  $F$  stetig mit Dichte  $f = F'$ .  $f(x) > 0$  für  $\{x : 0 < F(x) < 1\}$ ,  $X \sim F$   
Median

$$\vartheta(F) = F^{-1}\left(\frac{1}{2}\right)$$

bzw.  $F(\vartheta(F)) - \frac{1}{2} = 0 \Leftrightarrow h(F, \vartheta(F)) = 0$  mit

$$\begin{aligned}h(F, u) &= F(u) - \frac{1}{2} \\ &= \int \underbrace{(\mathbf{1}\{x \leq u\} - \frac{1}{2})}_{=: \tilde{h}_u(x)} dF(x) \\ &= \int \tilde{h}_u(x) dF(x)\end{aligned}$$

$$\stackrel{9.4(a)}{\Rightarrow} \lambda(x, F, u) = \tilde{h}_u(x) - h(F, u) = \mathbf{1}\{x \leq u\} - F(u)$$

$$\begin{aligned}\stackrel{9.4(c)}{\Rightarrow} \varphi(x, F) &= -\frac{\lambda(x, F, \vartheta(F))}{h'(F, \vartheta(F))} \\ &= -\frac{\mathbf{1}\{x \leq \vartheta(F)\} - F(\vartheta(F))}{f(\vartheta(F))} \\ &= \frac{\frac{1}{2} - \mathbf{1}\{x \leq \vartheta(F)\}}{f(\vartheta(F))} \\ &= \begin{cases} -\frac{1}{2f(\vartheta(F))}, & x \leq \vartheta(F) \\ +\frac{1}{2f(\vartheta(F))}, & x > \vartheta(F) \end{cases}\end{aligned}$$

Bemerkungen:

- (i)  $\hat{\vartheta}$  ist robust
- (ii)  $\hat{F}_n$  ist Treppenfunktion  $\Rightarrow \varphi(x, \hat{F}_n)$  existiert nicht  
 $\Rightarrow$  Bemerkung nach 9.2 ist hier nicht zutreffend
- (iii)

$$E[\varphi(X, F)] = \frac{\frac{1}{2} - P(X \leq \vartheta(F))}{f(\vartheta(F))} = 0$$

$$\tau^2(F) = E[\varphi^2(X, F)] = \frac{1}{4f^2(\vartheta(F))}$$

$$\stackrel{9.5}{\Rightarrow} \sqrt{n}(\hat{\vartheta}_n - \vartheta) \xrightarrow{D} \mathcal{N}\left(0, \frac{1}{4f^2(\vartheta(F))}\right)$$

Konkret:

$X_1, \dots, X_n \stackrel{uiv}{\sim} \mathcal{N}(\mu, \sigma^2)$ ,  $\vartheta(F) = \mu$

$$\Rightarrow \sqrt{n}(\hat{\vartheta}_n - \mu) \xrightarrow{D} \mathcal{N}\left(0, \frac{\pi\sigma^2}{2}\right)$$

$$f(\mu) = \frac{1}{\sqrt{2\pi}\sigma} \cdot 1$$

$$\hat{\vartheta}_n \sim AN\left(\mu, \underbrace{\frac{\pi\sigma^2}{2n}}_{\tau_1^2}\right)$$

$$\bar{X} \sim \mathcal{N}\left(0, \underbrace{\frac{\sigma^2}{n}}_{\tau_2^2}\right)$$

( $\bar{X}$  UMVUE)

$$\frac{\tau_1^2}{\tau_2^2} = \frac{\pi}{2} \approx 1,57$$

Einflusskurve des Medians  $\vartheta(F) = F^{-1}(\frac{1}{2})$  ist also

$$\varphi_{\frac{1}{2}}(x, F) = \frac{\frac{1}{2} - \mathbf{1}\{x \leq \vartheta(F)\}}{f(\vartheta(F))}$$

Ganz analog: Einflusskurve von  $F^{-1}(p)$  ist

$$(*) \quad \varphi_p(x, F) = \frac{p - \mathbf{1}\{x \leq F^{-1}(p)\}}{f(F^{-1}(p))}, \quad 0 < p < 1$$

### 9.7 Beispiel ( $\alpha$ -getrimmtes Mittel)

Sei  $F$  stetig,  $F' = f$ ,  $f(x) > 0$  für  $\{x : 0 < F(x) < 1\}$ .

$f$  symmetrisch mit Zentrum  $\mu = EX$ .

Für  $0 < \alpha < \frac{1}{2}$  heißt

$$\mu_\alpha(F) = \frac{1}{1-2\alpha} \int_{F^{-1}(\alpha)}^{F^{-1}(1-\alpha)} x dF(x) = \frac{1}{1-2\alpha} \int_\alpha^{1-\alpha} F^{-1}(p) dp$$

#### $\alpha$ -getrimmtes Mittel.

Für symmetrische Verteilungen gilt:

$$\mu_\alpha(F) = \mu$$

(Denn:)

$$\begin{aligned} \frac{1}{1-2\alpha} \int_{F^{-1}(\alpha)}^{F^{-1}(1-\alpha)} \mu dF(x) &= \mu \\ \Rightarrow \mu_\alpha(F) - \mu &= \frac{1}{1-2\alpha} \int_{F^{-1}(\alpha)}^{F^{-1}(1-\alpha)} (x - \mu) dF(x) = 0 \end{aligned}$$

Der Plug-In Schätzer für  $\mu_\alpha(F)$  ist

$$\mu_\alpha(\hat{F}_n) = \frac{1}{1-2\alpha} \int_\alpha^{1-\alpha} \hat{F}_n^{-1}(p) dp$$

wobei  $\hat{F}_n^{-1}(t) = X_{(i)}$ , falls  $\frac{i-1}{n} < t \leq \frac{i}{n}$ . (Aufgabe 16)

In der Praxis wird der (asymptotisch gleichwertige) Schätzer

$$\bar{X}_{n,\alpha} = \frac{1}{n-2[\alpha n]} \sum_{k=[\alpha n]+1}^{n-[\alpha n]} X_{(k)}$$

verwendet.

Einflusskurve von  $\mu_\alpha(F)$ :

9.4(c)  $\Rightarrow$

$$\begin{aligned} (**) \quad \varphi^\alpha(x, F) &= \frac{1}{1-2\alpha} \int_\alpha^{1-\alpha} \varphi_p(x, F) dp \\ &\stackrel{(*)}{=} \frac{1}{1-2\alpha} \int_\alpha^{1-\alpha} \frac{p - \mathbf{1}\{x \leq F^{-1}(p)\}}{f(F^{-1}(p))} dp \end{aligned}$$

Nun sei  $F(x) < \alpha$ . Dann:

$$\begin{aligned}
 (**) &= \frac{1}{1-2\alpha} \int_{\alpha}^{1-\alpha} (p-1) \underbrace{\frac{1}{f(F^{-1}(p))}}_{\text{Dichte von } F^{-1}} dp \\
 &= \frac{1}{1-2\alpha} \int_{\alpha}^{1-\alpha} \underbrace{(p-1)}_G dF^{-1}(p) \\
 (+) &\stackrel{=}{=} \frac{1}{1-2\alpha} \left[ \underbrace{((1-\alpha)-1)}_{=G(b)} \cdot F^{-1}(1-\alpha) - \underbrace{(\alpha-1)}_{=G(a)} \cdot F^{-1}(\alpha) \right. \\
 &\quad \left. - \underbrace{\int_{\alpha}^{1-\alpha} F^{-1}(p) dp}_{=(1-2\alpha)\cdot\mu} \right] \\
 &= \frac{1}{1-2\alpha} \left[ (-\alpha) \underbrace{(F^{-1}(1-\alpha) + F^{-1}(\alpha))}_{=2\mu} + F^{-1}(\alpha) - (1-2\alpha)\mu \right] \\
 &= \frac{F^{-1}(\alpha) - \mu}{1-2\alpha}
 \end{aligned}$$

(+): partielle Integration (Stochastik II),  $F$  weiterhin symmetrisch

Ähnliche Überlegungen für  $F(x) > 1 - \alpha$  bzw.  $\alpha \leq F(x) \leq 1 - \alpha$  ergeben:

$$\varphi^{\alpha}(x, F) = \begin{cases} \frac{F^{-1}(\alpha) - \mu}{1-2\alpha}, & x < F^{-1}(\alpha) \\ \frac{x - \mu}{1-2\alpha}, & F^{-1}(\alpha) \leq x \leq F^{-1}(1-\alpha) \\ \frac{F^{-1}(1-\alpha) - \mu}{1-2\alpha}, & x > F^{-1}(1-\alpha) \end{cases}$$

Insbesondere ist  $\varphi^{\alpha}(x, F)$  beschränkt in  $x$ .

$$\Rightarrow \bar{X}_{n,\alpha} = \frac{1}{n - 2[\alpha n]} \sum_{k=[\alpha n]+1}^{n-[\alpha n]} X_{(k)}$$

ist robust.

Einflusskurven-Heuristik ergibt:

$$\sqrt{n}(\bar{X}_{n,\alpha} - \mu_{\alpha}) \xrightarrow{D} \mathcal{N} \left( 0, \frac{1}{(1-2\alpha)^2} [2\alpha(F^{-1}(\alpha) - \mu)^2 + \int_{F^{-1}(\alpha)}^{F^{-1}(1-\alpha)} (x - \mu)^2 dF] \right)$$



## 10 Grundbegriffe der Testtheorie

Sei  $(\Omega, \mathcal{A}, \mathbb{P})$  Wahrscheinlichkeitsraum,  $(\mathfrak{X}, \mathcal{B}, \{P_\vartheta : \vartheta \in \Theta\})$  statistischer Raum,  $X : \Omega \rightarrow \mathfrak{X}$  Zufallsvariable,  $\Theta = \Theta_0 + \Theta_1$  mit  $\Theta_0, \Theta_1 \neq \emptyset$ .  
 $(\Theta_0 \cap \Theta_1 = \emptyset)$

### 10.1 Definition

Die Aussage  $H_0 : \vartheta \in \Theta_0$  heißt (Null-)Hypothese,  $H_1 : \vartheta \in \Theta_1$  heißt Alternativhypothese oder Alternative.

$|\Theta_j| = 1 \Rightarrow \Theta_j$  heißt einfach, sonst zusammengesetzt

### 10.2 Definition

Ein **randomisierter Test** zur Prüfung von  $H_0$  gegen  $H_1$  ist eine messbare Abbildung  $\varphi : \mathfrak{X} \rightarrow [0, 1]$  mit der Interpretation

$$\varphi(x) = P(H_0 \text{ ablehnen} \mid X = x)$$

Gilt  $\varphi(\mathfrak{X}) = \{0, 1\}$ , so heißt  $\varphi$  **nicht randomisiert**. Mit  $\mathcal{K} := \{x \in \mathfrak{X} : \varphi(x) = 1\}$  gilt dann  $\varphi = \mathbf{1}_{\mathcal{K}}$  und die Testvorschrift lautet:

$$\begin{aligned} x \in \mathcal{K} &\Rightarrow H_0 \text{ ablehnen} \\ x \in \mathfrak{X} \setminus \mathcal{K} &\Rightarrow H_0 \text{ nicht ablehnen} \end{aligned}$$

$\mathcal{K}$  heißt kritischer Bereich (Ablehnbereich),  $\mathfrak{X} \setminus \mathcal{K}$  heißt Annahmebereich.

### 10.3 Bemerkung

Falls  $0 < \varphi(x) < 1$ , so muss „externes“ Bernoulli-Experiment durchgeführt werden; man erhält also Realisierung  $y$  einer Zufallsvariablen  $Y$  mit  $Y \sim \text{Bin}(1, \varphi(x))$ .

In praktischen Anwendungen ist „Randomisierung“ unerwünscht.

### 10.4 Definition

Es sei  $T : \mathfrak{X} \rightarrow \mathbb{R}$  eine messbare Abbildung. Häufig besitzt ein nicht randomisierter Test die Gestalt

$$(*) \quad \begin{aligned} T(x) \geq c &\Rightarrow H_0 \text{ ablehnen} \\ T(x) < c &\Rightarrow \text{kein Widerspruch zu } H_0 \end{aligned}$$

(d.h.  $\mathcal{K} = \{x \in \mathfrak{X} : T(x) \geq c\} = T^{-1}([c, \infty))$ )

Dann heißt  $T$  Testgröße (Prüfgröße) und  $c \in \mathbb{R}$  heißt kritischer Wert.

(\*) liefert Test mit **oberem Ablehnbereich**.

In (\*)  $\geq$  durch  $\leq$  und  $<$  durch  $>$  ersetzen  $\leftrightarrow$  Test mit unterem Ablehnbereich

### 10.5 Beispiel

$$(\mathfrak{X}, \mathcal{B}) = (\mathbb{R}^{m+n}, \mathcal{B}^{n+m}), X = (\underbrace{X_1, \dots, X_m}_{\overset{uiv}{F}}, \underbrace{Y_1, \dots, Y_n}_{\overset{uiv}{G}}), X_1, \dots, Y_n$$

unabhängig,  $\vartheta = (F, G)$ ,  $\Theta = \{(F, G) : F, G \text{ stetig}\}$ ,  $\Theta_0 = \{(F, G) \in \Theta : F = G\}$

$$H_0 : F = G$$

$$H_1 : F \neq G$$

(nichtparametrisches 2-Stichproben-Problem mit allgemeiner Alternative)

Sei

$$\hat{F}_m(x) = \frac{1}{m} \sum_{i=1}^m \mathbf{1}\{X_i \leq x\}, \hat{G}_n(x) = \frac{1}{n} \sum_{j=1}^n \mathbf{1}\{Y_j \leq x\}$$

Mögliche Prüfgröße (mit oberem Ablehnbereich):

$$T(X_1, \dots, X_m, Y_1, \dots, Y_n) = \sup_{x \in \mathbb{R}} |\hat{F}_m(x) - \hat{G}_n(x)|$$

(Kolmogorov-Smirnov-Testgröße)

### 10.6 Definition und Bemerkung

Ein Fehler 1. Art ist das Verwerfen von  $H_0$ , obwohl  $H_0$  richtig ist.

Ein Fehler 2. Art ist das Nichtverwerfen von  $H_0$ , obwohl  $H_0$  falsch ist.

Entscheidung	$H_0$ richtig	$H_0$ falsch
$H_0$ nicht verwerfen	richtige Entscheidung	Fehler 2. Art
$H_0$ verwerfen	Fehler 1. Art	richtige Entscheidung

Die Funktion

$$G_\varphi: \Theta \rightarrow [0, 1] \\ \vartheta \mapsto G_\varphi(\vartheta) := E_\vartheta[\varphi] = \int_{\mathfrak{X}} \varphi(x) P_\vartheta(dx)$$

heißt **Gütefunktion** des Tests  $\varphi$ .

$$(\varphi = \mathbf{1}_{\mathcal{K}} \Rightarrow G_\varphi(\vartheta) = P_\vartheta(\mathcal{K}), \varphi = \mathbf{1}\{T(x) \geq c\} \Rightarrow G_\varphi(\vartheta) = P_\vartheta(T \geq c))$$

Ideale Gütefunktion wäre

$$G_\varphi(\vartheta) = \begin{cases} 1, & \vartheta \in \Theta_1 \\ 0, & \vartheta \in \Theta_0 \end{cases}$$

Sei  $\alpha \in (0, 1)$ .  $\varphi$  heißt Test zum **Niveau**  $\alpha$   $:\Leftrightarrow G_\varphi(\vartheta) \leq \alpha \forall \vartheta \in \Theta_0$ <sup>26</sup>

In Praxis übliche Werte:  $\alpha = 0,05; 0,01; 0,001$

Kleines  $\alpha$  dient „Sicherheit von  $H_1$ “.<sup>27</sup>

Die Zahl  $\sup_{\vartheta \in \Theta_0} G_\varphi(\vartheta)$  heißt **Umfang** (size) von  $\varphi$ .

## 10.7 Definition

Sei

$$\Phi_\alpha = \{\varphi: \mathfrak{X} \rightarrow [0, 1] \mid \sup_{\vartheta \in \Theta_0} G_\varphi(\vartheta) \leq \alpha\}$$

die Menge aller Niveau  $\alpha$ -Tests.

$\Phi_\alpha \neq \emptyset$ , da  $\varphi \equiv \alpha \in \Phi_\alpha$ .

Sei  $\tilde{\Phi}_\alpha \subset \Phi_\alpha$

$\varphi_1 \in \tilde{\Phi}_\alpha$  heißt **gleichmäßig besser** als  $\varphi_2 \in \tilde{\Phi}_\alpha$   $:\Leftrightarrow$

$$G_{\varphi_1}(\vartheta) \geq G_{\varphi_2}(\vartheta) \forall \vartheta \in \Theta_1$$

$\varphi^* \in \tilde{\Phi}_\alpha$  heißt (gleichmäßig) **bester Test** in  $\tilde{\Phi}_\alpha$   $:\Leftrightarrow$

$$G_{\varphi^*}(\vartheta) \geq G_\varphi(\vartheta) \forall \vartheta \in \Theta_1 \forall \varphi \in \tilde{\Phi}_\alpha$$

Bezeichnung: UMP-Test („uniformly most powerfully“)

<sup>26</sup>Wahrscheinlichkeit für einen Fehler 1. Art ist  $\leq \alpha$

<sup>27</sup>vgl. „Wahl der Nullhypothese“; das Verwerfen von  $H_0$  ist „fast nie“ falsch, also in diesem Fall umgekehrt  $H_1$  auch „fast immer“ richtig (...)

## 11 Neyman-Pearson-Tests (NP-Tests)

Es sei  $\Theta_0 = \{\vartheta_0\}$   $\Theta_1 = \{\vartheta_1\}$ ,  $f_j$  sei die Dichte von  $P_{\vartheta_j}$  bezüglich dem Maß  $\mu$  auf  $\mathfrak{X}$ .

### 11.1 Definition

$\varphi$  heißt NP-Test für  $H_0 : \vartheta = \vartheta_0$  gegen  $H_1 : \vartheta = \vartheta_1$   
 $:\Leftrightarrow \exists c \geq 0 \exists \gamma \in [0, 1]$  mit

$$(1) \quad \varphi(x) = \begin{cases} 1, & \text{falls } f_1(x) > cf_0(x) \\ \gamma, & \text{falls } f_1(x) = cf_0(x) \\ 0, & \text{falls } f_1(x) < cf_0(x) \end{cases}$$

Beachte<sup>28</sup>:  $E_{\vartheta_0}(\varphi) = P_{\vartheta_0}(f_1 > cf_0) + \gamma P_{\vartheta_0}(f_1 = cf_0)$

$$(2) \quad Q(x) := \begin{cases} \frac{f_1(x)}{f_0(x)}, & \text{falls } f_0(x) > 0 \\ \infty, & \text{falls } f_0(x) = 0 \end{cases}$$

$$\tilde{\varphi}(x) = \begin{cases} 1, & \text{falls } Q(x) > c \\ \gamma, & \text{falls } Q(x) = c \\ 0, & \text{falls } Q(x) < c \end{cases}$$

$$\begin{aligned} \text{falls } f_0(x) > 0 & \Rightarrow \varphi(x) = \tilde{\varphi}(x) \\ \text{falls } f_0(x) = 0, f_1(x) > 0 & \Rightarrow \varphi(x) = \tilde{\varphi}(x) \\ \text{falls } f_0(x) = 0, f_1(x) = 0 & \Rightarrow \varphi(x) \neq \tilde{\varphi}(x) \end{aligned}$$

Es gilt:  $\{f_0 > 0\} \cup \{f_1 > 0\} \subset \{\varphi = \tilde{\varphi}\}$

$$\Rightarrow P_{\vartheta_1}(\varphi = \tilde{\varphi}^*) = P_{\vartheta_1}(\varphi = \tilde{\varphi}) = 1$$

Beachte:  $E_{\vartheta_0}(\tilde{\varphi}) = P_{\vartheta_0}(Q > c) + \gamma P_{\vartheta_0}(Q = c)$

### 11.2 Satz

Der Test aus 11.1(1) ist bester Test zum Niveau  $\alpha := E_{\vartheta_0}(\varphi)$ .

Beweis:

Sei  $\Psi$  beliebiger Test mit  $E_{\vartheta_0}(\Psi) \leq \alpha$ .

Zu zeigen:

$$E_{\vartheta_1}(\varphi) \geq E_{\vartheta_1}(\Psi)$$

---

<sup>28</sup>Niveau

Sei  $M^{(+)} := \{x : \varphi(x) > \Psi(x)\}$ ,  $M^{(-)} := \{x : \varphi(x) < \Psi(x)\}$ ,  
 $M^{(=)} := \{x : \varphi(x) = \Psi(x)\}$

$$x \in M^{(+)} \Rightarrow \varphi(x) > 0 \Rightarrow f_1(x) \geq cf_0(x)$$

$$x \in M^{(-)} \Rightarrow \varphi(x) < 1 \Rightarrow f_1(x) \leq cf_0(x)$$

$$\begin{aligned} \Rightarrow E_{\vartheta_1}(\varphi - \Psi) &= \int_{\mathfrak{X}} (\varphi(x) - \Psi(x)) f_1(x) \mu(dx) \\ &= \int_{M^{(+)}} \underbrace{(\varphi(x) - \Psi(x))}_{>0} \underbrace{f_1(x)}_{\geq cf_0} d\mu(x) \\ &\quad + \int_{M^{(-)}} \underbrace{(\varphi - \Psi)}_{<0} \underbrace{f_1}_{\leq cf_0} d\mu + \underbrace{\int_{M^{(=)}} (\varphi - \Psi) f_1 d\mu}_{=0} \\ &\geq \int_{M^{(+)}} (\varphi - \Psi) cf_0 d\mu + \int_{M^{(-)}} (\varphi - \Psi) cf_0 d\mu \\ &\quad + \int_{M^{(=)}} (\varphi - \Psi) cf_0 d\mu \\ &= c \int_{\mathfrak{X}} (\varphi - \Psi) f_0 d\mu \\ &= \underbrace{c}_{\geq 0} \underbrace{[E_{\vartheta_0}(\varphi) - E_{\vartheta_0}(\Psi)]}_{\geq 0} \\ &\geq 0 \end{aligned}$$

### 11.3 Bemerkung

Beweis deckt auch den Fall  $\varphi(x) = \gamma(x)$ , falls  $f_1(x) = cf_0(x)$  ab

### 11.4 Lemma von Neyman-Pearson

- Zu jedem  $\alpha \in (0, 1)$  existiert ein NP-Test  $\varphi$  der Form 11.1(1).
- Ist  $\Psi$  ebenfalls bester Test zum Niveau  $\alpha$ ,  
so gilt mit  $\varphi$  aus (a) und  $D = \{x : f_1(x) \neq cf_0(x)\}$

$$\varphi(x) = \Psi(x) \text{ f\"ur } \mu\text{- fast alle } x \in D$$

Beweis:

a) Sei  $Q$  wie in 11.1(2). Zu zeigen:

$$\exists c \geq 0 \exists \gamma \in [0, 1] \text{ mit } P_{\vartheta_0}(Q > c) + \gamma P_{\vartheta_0}(Q = c) = \alpha \quad (*)$$

Sei  $F_0(t) := P_{\vartheta_0}(Q \leq t)$  die Verteilungsfunktion von  $Q$  unter  $\vartheta_0$ .

Dann wird (\*) zu  $1 - F_0(c) + \gamma(F_0(c) - F_0(c-0)) \stackrel{!}{=} \alpha$ .

Setze  $c := F_0^{-1}(1 - \alpha)$  und

$$\gamma := \begin{cases} 0, & \text{falls } F_0(c) = F_0(c-0) \\ \frac{F_0(c) - (1 - \alpha)}{F_0(c) - F_0(c-0)}, & \text{sonst} \end{cases}$$

b) siehe Pruscha, Vorlesungen über Mathematische Statistik, S. 225

Beispiel: (Poissonverteilung)

$X \sim Po(\lambda)$ , ( $\lambda > 0$ ),  $0 < \lambda_0 < \lambda_1$

$$H_0 : \lambda = \lambda_0, \quad H_1 : \lambda = \lambda_1$$

$$f(x, \lambda) = e^{-\lambda} \frac{\lambda^x}{x!} \quad x = 1, 2, \dots$$

$\Rightarrow$  Dichtequotient ist

$$T(x) = \frac{f(x, \lambda_1)}{f(x, \lambda_0)} = \underbrace{\left(\frac{\lambda_1}{\lambda_0}\right)^x}_{>1} e^{-(\lambda_1 - \lambda_0)}$$

streng monoton wachsend in  $x$ .

$\Rightarrow$  Bereich  $\{T(x) > c\}$  bzw.  $\{T(x) = c\}$  kann umgeschrieben werden in  $\{x > k\}$  bzw.  $\{x = k\}$ .

NP-Test

$$\varphi(x) = \begin{cases} 1 & , x > k \\ \gamma & , x = k \\ 0 & , x < k \end{cases}$$

für  $\alpha \in (0, 1)$  wähle  $k \in \mathbb{N}_0$ ,  $\gamma \in [0, 1]$  so, dass

$$P_{\lambda_0}(X > k) + \gamma P_{\lambda_0}(X = k) \stackrel{!}{=} \alpha$$

zum Beispiel  $\alpha = 0,05$ ,  $\lambda_0 = 1$  :

$$P_{\lambda_0}(X = 3) = 0,0613, \quad P_{\lambda_0}(X > 3) = 0,0190$$

$$\Rightarrow P_{\lambda_0}(X \geq 3) > 0,05$$

$$P_{\lambda_0}(X > 3) + \gamma P_{\lambda_0}(X = 3) \stackrel{!}{=} 0,05$$

$$\Rightarrow \gamma = \frac{\alpha - P_{\lambda_0}(X > 3)}{P_{\lambda_0}(X = 3)} = 0,5057$$

Bemerkung:

Wird bei der konkreten Testdurchführung z.B. der Wert  $x = 3$  beobachtet, so wird in der Praxis der sogenannte p-Wert

$$\begin{aligned} p^*(x) &= p^*(3) \\ &= P_{\lambda_0}(\text{„mindestens so extremes Ergebnis wie das beobachtete“}) \\ &= P_{\lambda_0}(X \geq 3) \\ &= 0,0803 \end{aligned}$$

$[p^*(2) > 0,1, p^*(4) = 0,019, \text{ usw}]$  angegeben.

Bei diesem Vorgehen wird das Problem der Randomisierung umgangen:

Ist zum Beispiel  $\alpha = 0,05$  gewählt, so entscheidet man bei  $p^*(x) \leq 0,05$  gegen die Hypothese.

Bei  $p^*(x) > 0,05$  wird die Hypothese nicht verworfen.

Im Folgenden: „Loslösen“ vom Fall  $|\Theta_0| = 1 = |\Theta_1|$

Sei  $\{P_{\vartheta} : \vartheta \in \Theta\}$  dominiert durch  $\sigma$ -endliches Maß  $\mu$  auf  $\mathcal{B}$ .

$$f(x, \vartheta) = \frac{dP_{\vartheta}}{d\mu}(x)$$

$\Theta \subset \mathbb{R}^1$ ,  $\Theta$  offen

**11.5 Definition**

Es sei  $T : \mathfrak{X} \rightarrow \mathbb{R}$  messbar mit  $\forall \vartheta, \vartheta' \in \Theta$  mit  $\vartheta < \vartheta'$  existiert eine monoton wachsende Funktion  $g(\cdot, \vartheta, \vartheta') : \mathbb{R} \rightarrow [0, \infty]$  mit

$$\frac{f(x, \vartheta')}{f(x, \vartheta)} = g(T(x), \vartheta, \vartheta'), \quad x \in \mathfrak{X}$$

Dann heißt  $\{P_{\vartheta} : \vartheta \in \Theta\}$  **Klasse mit monotonem Dichtequotienten (DQ) in T.**

Falls  $f(x, \vartheta') > f(x, \vartheta) = 0$ , so  $\frac{f(x, \vartheta')}{f(x, \vartheta)} := \infty$ .

## 11.6 Beispiel

Sei

$$f(x, \vartheta) = c(\vartheta) \cdot e^{q(\vartheta)T(x)} \cdot h(x), \quad x \in \mathfrak{X}$$

(einparametrische Exponentialfamilie)

Ist  $q : \Theta \rightarrow \mathbb{R}$  streng monoton wachsend und gilt  $\text{Var}_{\vartheta}(T) > 0 \forall \vartheta \in \Theta$  (\*),  
so ist  $\{P_{\vartheta} : \vartheta \in \Theta\}$  Klasse mit monotonem DQ in T.

Beweis:

(i) Aus (\*) folgt Injektivität von  $\Theta \ni \vartheta \rightarrow P_{\vartheta}$ :

Annahme:  $\vartheta \neq \vartheta'$  und  $P_{\vartheta} = P_{\vartheta'}$

$$\Rightarrow f(\cdot, \vartheta) = f(\cdot, \vartheta') \quad \mu\text{-f.ü.}$$

$$\Rightarrow \log c(\vartheta) + q(\vartheta) \cdot T(x) = \log c(\vartheta') + q(\vartheta') \cdot T(x) \quad \mu\text{-f.ü.}$$

$$\Rightarrow T(x) = \frac{\log c(\vartheta') - \log c(\vartheta)}{q(\vartheta) - q(\vartheta')} \quad \mu\text{-f.ü.}$$

$$\Rightarrow \text{Var}(T) = 0$$

Widerspruch zu (\*)!

(ii)  $\vartheta < \vartheta'$

$$\Rightarrow \frac{f(x, \vartheta')}{f(x, \vartheta)} = \frac{c(\vartheta')}{c(\vartheta)} \exp(\underbrace{(q(\vartheta') - q(\vartheta)) \cdot T(x)}_{>0}) =: g(T(x), \vartheta, \vartheta')$$

Spezialfall:  $\text{Bin}(n, \vartheta)$ ,  $0 < \vartheta < 1$

$$f(x, \vartheta) = \binom{n}{x} \vartheta^x (1 - \vartheta)^{n-x} = (1 - \vartheta)^n e^{xq(\vartheta)} \binom{n}{x}$$

wobei  $q(\vartheta) = \log \frac{\vartheta}{1-\vartheta}$  streng monoton wachsend in  $\vartheta$  ist.

$\Rightarrow$  monotoner DQ in  $T(x) = x$ ,  $x \in \{0, \dots, n\}$

In der Situation von 11.5 sei  $H_0 : \vartheta \leq \vartheta_0$  gegen  $H_1 : \vartheta > \vartheta_0$  zu testen.  
( $\vartheta_0 \in \Theta$  vorgegeben)

Für  $c^* \in \mathbb{R}$  und  $\gamma^* \in [0, 1]$  sei

$$(*) \quad \varphi^*(x) = \begin{cases} 1, & T(x) > c^* \\ \gamma^*, & T(x) = c^* \\ 0, & T(x) < c^* \end{cases}$$

$$\Rightarrow E_{\vartheta_0}(\varphi^*) = P_{\vartheta_0}(T > c^*) + \gamma^* P_{\vartheta_0}(T = c^*)$$

### 11.7 Satz

Die Klasse  $\{P_\vartheta : \vartheta \in \Theta\}$ ,  $\Theta \subset \mathbb{R}^1$ , besitze monotonen DQ in  $T$ . Dann gilt:

- Ist  $\varphi^*$  von der Form (\*) mit  $\alpha := E_{\vartheta_0}(\varphi^*) > 0$ , so ist  $\varphi^*$  UMP-Test für  $H_0$  gegen  $H_1$ .
- Zu vorgegebenem  $\vartheta_0 \in \Theta$  und  $\alpha \in (0, 1)$  existieren  $c^* \in \mathbb{R}, \gamma^* \in [0, 1]$ , so dass  $\varphi^*$  aus (\*) ein Test zum Umfang  $\alpha$  ist.
- Die Gütefunktion  $E_\vartheta \varphi^*$  ist monoton wachsend und auf  $\{\vartheta : 0 < E_\vartheta \varphi^* < 1\}$  streng monoton.

Beweis:

- Sei  $\vartheta_1 \in \Theta$  mit  $\vartheta_1 > \vartheta_0$  beliebig.

$$H'_0 : \vartheta = \vartheta_0 \text{ gegen } H'_1 : \vartheta = \vartheta_1$$

Sei  $f_j(x) := f(x, \vartheta_j)$ . Wegen

$$\frac{f_1(x)}{f_0(x)} = g(T(x), \vartheta_0, \vartheta_1)$$

existiert zu  $c^*$  ein  $c := g(c^*, \vartheta_0, \vartheta_1)$  mit

$$\{x : \frac{f_1(x)}{f_0(x)} > c\} \subset \{x : T(x) > c^*\}$$

$$\{x : \frac{f_1(x)}{f_0(x)} < c\} \subset \{x : T(x) < c^*\}$$

[Echte Teilmengen, denn aus  $T(x) > c^*$  folgt  $g(T(x), \vartheta_0, \vartheta_1) \geq c$ .]

Aus

$$\begin{aligned} 0 < \alpha &= E_{\vartheta_0} \varphi^* \\ &= P_{\vartheta_0}(T > c^*) + \gamma^* P_{\vartheta_0}(T = c^*) \\ &\leq P_{\vartheta_0}(T \geq c^*) \\ &= P_{\vartheta_0}\left(\frac{f_1(x)}{f_0(x)} \geq c\right) \end{aligned}$$

folgt  $c < \infty$ . [Denn:  $P_{\vartheta_0}(\frac{f_1(x)}{f_0(x)} = \infty) = 0$ ]

Für  $\varphi^*$  aus (\*) gilt

$$(**) \quad \varphi^*(x) = \begin{cases} 1, & \frac{f_1(x)}{f_0(x)} > c \\ \gamma(x), & \frac{f_1(x)}{f_0(x)} = c \\ 0, & \frac{f_1(x)}{f_0(x)} < c \end{cases}$$

mit  $\gamma(x) \in \{0, 1, \gamma^*\}$ .

Nach 11.2 und 11.3 ist  $\varphi^*$  bester Test für  $H'_0$  gegen  $H'_1$  zum Niveau  $\alpha = E_{\vartheta_0}(\varphi^*)$ .

Da  $\varphi^*$  in (\*) nicht von  $\vartheta_1$  abhängt, ist (a) für  $H'_0$  gegen  $H_1 : \vartheta > \vartheta_0$  bewiesen.

Teil (c)  $\Rightarrow E_{\vartheta}\varphi^* \leq \alpha \forall \vartheta \leq \vartheta_0$ , d.h. Test  $\varphi^*$  ist UMP-Test für  $H_0$  gegen  $H_1$  zu  $\alpha := E_{\vartheta_0}\varphi^*$ .

b) Analog zu 11.4(a).

Nach (c) gilt  $\sup_{\vartheta \in \Theta_0} E_{\vartheta}\varphi^* = E_{\vartheta_0}\varphi^* = \alpha$ , d.h. der Test hat Umfang  $\alpha$ .

c) Sei  $\vartheta_1 < \vartheta_2$  beliebig,  $\alpha_1 := E_{\vartheta_1}\varphi^*$ .

Analog zu 11.7 (\*\*) ist  $\varphi^*$  NP-Test für  $H_0^* : \vartheta = \vartheta_1$  gegen  $H_1^* : \vartheta = \vartheta_2$ .  
Da  $\varphi^*$  besser als  $\varphi_1 := \alpha_1$  folgt

$$\alpha_1 = E_{\vartheta_2}(\varphi_1) \leq E_{\vartheta_2}(\varphi^*)$$

d.h.  $E_{\vartheta}(\varphi^*)$  monoton wachsend.

(Für strenge Monotonie siehe Pruscha, S. 230)

#### Anmerkung:

Die Tests in (\*) und (\*\*) sind äquivalent.  $\varphi^*$  in (\*) hängt nicht von  $\vartheta_1$  ab, also hängt auch der Test in (\*\*) nicht von  $\vartheta_1$  ab. Dies ist jedoch nicht beweisbar, da  $\vartheta_1$  sowohl in  $f_1(x)$  als auch in  $c = c(\vartheta_0, \vartheta_1)$  eingeht.

Beide Tests haben gleichen Ablehnbereich!

### 11.8 Bemerkung

a) Testproblem  $H_0 : \vartheta \geq \vartheta_0$  gegen  $H_1 : \vartheta < \vartheta_0$  analog.

[ $\vartheta$  durch  $-\vartheta$  und T durch -T ersetzen  $\Rightarrow$  in (\*) werden  $<$  und  $>$  vertauscht]

b) Für **zweiseitiges Testproblem**  $H_0 : \vartheta = \vartheta_0$  gegen  $H_1 : \vartheta \neq \vartheta_0$  existiert i.A. kein UMP-Test zum Niveau  $\alpha$ .

Ein solcher Test  $\varphi^*$  wäre

(i) UMP-Test für  $H_0 : \vartheta = \vartheta_0$  gegen  $H_1^> : \vartheta > \vartheta_0$

$$\Rightarrow E_{\vartheta} \varphi^* < \alpha \quad \forall \vartheta < \vartheta_0$$

( $\hat{=} H_0$ )

(ii) UMP-Test für  $H_0 : \vartheta = \vartheta_0$  gegen  $H_1^< : \vartheta < \vartheta_0$

$$\Rightarrow E_{\vartheta} \varphi^* > \alpha \quad \forall \vartheta < \vartheta_0$$

( $\hat{=} H_1$ )

Widerspruch!

## 11.9 Beispiel (Der einseitige Gauss-Test)

Sei  $X = (X_1, \dots, X_n)$ ,  $X_1, \dots, X_n \stackrel{uiv}{\sim} \mathcal{N}(\mu, \sigma_0^2)$ ,  $\sigma_0^2$  bekannt. Da

$$\begin{aligned} \frac{f(x, \mu_1, \sigma_0^2)}{f(x, \mu_0, \sigma_0^2)} &= \frac{\exp(-\frac{1}{2\sigma_0^2} \sum_{j=1}^n (x_j - \mu_1)^2)}{\exp(-\frac{1}{2\sigma_0^2} \sum_{j=1}^n (x_j - \mu_0)^2)} \\ &= \exp\left(\frac{\mu_1 - \mu_0}{\sigma_0^2} \underbrace{\sum_j x_j}_{=T(x)} - \frac{n(\mu_1^2 - \mu_0^2)}{2\sigma_0^2}\right) \end{aligned}$$

streng monoton wachsend in  $T(x) = \sum_j x_j$  ist für  $\mu_1 > \mu_0$ , besitzt

$\{\otimes \mathcal{N}(\mu, \sigma_0^2) : \mu \in \mathbb{R}\}$  monotonen DQ in  $T(x) = \sum_{j=1}^n x_j$

Als UMP-Test zum Niveau  $\alpha$  für  $H_0 : \mu \leq \mu_0$  gegen  $H_1 : \mu > \mu_0$  ergibt sich

$$\varphi^*(x) = \begin{cases} 1, & \sum_j x_j > c^* \\ \gamma^*, & \sum_j x_j = c^* \\ 0, & \sum_j x_j < c^* \end{cases}$$

Da  $P_{\mu_0}(\sum_j X_j = c^*) = 0$  kann  $\gamma^* \in [0, 1]$  beliebig gewählt werden, z.B.  $\gamma^* = 0$ . Außerdem:

$$\begin{aligned} E_{\mu_0} \varphi^* &= P_{\mu_0}\left(\sum_{j=1}^n X_j > c^*\right) = P_{\mu_0}\left(\underbrace{\sqrt{n} \frac{\bar{X}_n - \mu_0}{\sigma_0}}_{\sim \mathcal{N}(0,1)} > \sqrt{n} \frac{c^* - \mu_0}{\sigma_0}\right) \stackrel{!}{=} \alpha \\ &\Rightarrow \sqrt{n} \frac{c^* - \mu_0}{\sigma_0} \stackrel{!}{=} z_{1-\alpha} := \Phi^{-1}(1 - \alpha) \end{aligned}$$

Ergebnis:

$$\varphi^*(x) = \begin{cases} 1, & \sqrt{n} \frac{\bar{x}_n - \mu_0}{\sigma_0} > z_{1-\alpha} \\ 0, & \sqrt{n} \frac{\bar{x}_n - \mu_0}{\sigma_0} \leq z_{1-\alpha} \end{cases}$$

ist UMP-Test zum Niveau  $\alpha$  für  $H_0 : \mu \leq \mu_0$  gegen  $H_1 : \mu > \mu_0$ .

### 11.10 Beispiel (UMP-Tests in einparametrischen Exponentialfamilien)

Sei  $f_1(x_1, \vartheta) = c(\vartheta)e^{\vartheta T(x)}h(x)$ ,  $X_1, \dots, X_n \stackrel{uiv}{\sim} f_1$ .

$$\Rightarrow f(x, \vartheta) = c(\vartheta)^n \exp(\vartheta \sum_i T(x_i)) \prod_i h(x_i)$$

und  $f$  hat monotonen DQ in  $\tilde{T}(x) = \sum_{j=1}^n T(x_j)$  (vgl. 11.6).

$\Rightarrow$  UMP-Test zum Niveau  $\alpha$  für  $H_0 : \vartheta \leq \vartheta_0$  gegen  $H_1 : \vartheta > \vartheta_0$  ist

$$\varphi^*(x) = \begin{cases} 1, & \tilde{T}(x) > c^* \\ \gamma^*, & \tilde{T}(x) = c^* \\ 0, & \tilde{T}(x) < c^* \end{cases}$$

wobei  $P_{\vartheta_0}(\tilde{T} > c^*) + \gamma^* P_{\vartheta_0}(\tilde{T} = c^*) \stackrel{!}{=} \alpha$ .

### 11.11 Korollar

Sei  $h = h(t)$  streng monoton wachsend,  $\tilde{T}(x) = h(T(x))$ .

In der Situation von 11.7 ist dann auch

$$\tilde{\varphi}^*(x) = \begin{cases} 1, & \tilde{T}(x) > \tilde{c}^* \\ \tilde{\gamma}^*, & \tilde{T}(x) = \tilde{c}^* \\ 0, & \tilde{T}(x) < \tilde{c}^* \end{cases}$$

mit  $\tilde{c}^*, \underbrace{\tilde{\gamma}^*}_{\in[0,1]}$  gemäß  $E_{\vartheta_0} \tilde{\varphi}^* \stackrel{!}{=} \alpha$  UMP-Test zum Niveau  $\alpha$  für  $H_0$  gegen  $H_1$ .

## 12 UMPU Tests („UMP unbiased“)

Nach Bemerkung 11.8(b) existiert im Allgemeinen kein zweiseitiger UMP-Test zu einem Niveau  $\alpha$ . Deshalb Einschränkung auf unverfälschte Tests:  $\varphi \in \Phi_\alpha$  heißt **unverfälscht** (unbiased) zum Niveau  $\alpha$  für  $H_0 : \vartheta \in \Theta_0$  gegen  $H_1 : \vartheta \in \Theta_1$ , falls

$$(1) E_\vartheta \varphi \leq \alpha \quad \forall \vartheta \in \Theta_0, \quad E_\vartheta \varphi \geq \alpha \quad \forall \vartheta \in \Theta_1$$

Im Folgenden liegen einparametrische Exponentialfamilien mit Dichte

$$(*) f(x, \vartheta) = c(\vartheta) \cdot \exp(\vartheta T(x)) \cdot h(x), \quad x \in \mathfrak{X}$$

und natürlichem Parameterbereich  $\Theta$  vor.

Zu testen sei  $H_0 : \vartheta = \vartheta_0$  gegen  $H_1 : \vartheta \neq \vartheta_0$ .

Nach Lemma 6.12 ist die Gütefunktion  $\beta(\vartheta) = E_\vartheta \varphi(X)$  beliebig oft differenzierbar. Aus Forderung (1) folgt:

$$(2) E_{\vartheta_0} \varphi(X) = \alpha, \quad \frac{d}{d\vartheta} E_\vartheta \varphi(X) \Big|_{\vartheta=\vartheta_0} = 0$$

Mit

$$c(\vartheta) = \left[ \int e^{\vartheta T(x)} h(x) \mu(dx) \right]^{-1}$$

$$c'(\vartheta) = - \int T(x) e^{\vartheta T(x)} h(x) \mu(dx) \cdot c(\vartheta)^2$$

folgt weiter

$$\begin{aligned} \beta'(x) &= \left[ \int \varphi(x) c(\vartheta) e^{\vartheta T(x)} h(x) \mu(dx) \right]' \\ &= c'(\vartheta) \int \varphi(x) e^{\vartheta T(x)} h(x) \mu(dx) + c(\vartheta) \int \varphi(x) T(x) e^{\vartheta T(x)} h(x) \mu(dx) \\ &= -\bar{c}(\vartheta)^2 \int T(x) e^{\vartheta T(x)} h(x) \mu(dx) \int \varphi(x) e^{\vartheta T(x)} h(x) \mu(dx) \\ &\quad + E_\vartheta[\varphi(x) T(x)] \\ &= E_\vartheta[\varphi(x) T(x)] - E_\vartheta T(x) E_\vartheta \varphi(x) \end{aligned}$$

Damit ist (2) äquivalent zu

$$(3) E_{\vartheta_0} \varphi(x) = \alpha, \quad E_{\vartheta_0}[\varphi(x) T(x)] = \alpha E_{\vartheta_0} T(x)$$

### 12.1 Satz (UMPU-Tests in einparametrischen Exponentialfamilien)

Exponentialfamilie wie in (\*). Weiter sei

$$\varphi^*(x) = \begin{cases} 1, & T(x) < c_1^* \text{ oder } T(x) > c_2^* \\ \gamma_i^*, & T(x) = c_i^* \quad (i = 1, 2) \\ 0, & c_1^* < T(x) < c_2^* \end{cases}$$

wobei  $c_1^*, c_2^*, 0 \leq \gamma_1^*, \gamma_2^* \leq 1$  so, dass  $\varphi^*$  (3) erfüllt. Dann:

- Unter allen Niveau  $\alpha$  Tests für  $H_0 : \vartheta = \vartheta_0$  gegen  $H_1 : \vartheta \neq \vartheta_0$  die (3) erfüllen ist  $\varphi^*$  gleichmäßig bester Test.
- $\varphi^*$  ist UMPU-Test zum Niveau  $\alpha$  für  $H_0$  gegen  $H_1$ .

Anmerkung:

UMP-Tests sind eventuell auf einer Seite besser, versagen dafür aber auf der anderen Seite. Sie sind hier aber sowieso unzulässig, da sie nicht unverfälscht sind!

### 12.2 Bemerkungen

- Aus (3) folgt

$$E_{\vartheta_0}[\varphi(X) \cdot (aT(X) + b)] = a \underbrace{E_{\vartheta_0}[\varphi(X)T(X)]}_{= \alpha E_{\vartheta_0}T} + \alpha \cdot b = \alpha E_{\vartheta_0}[aT(X) + b]$$

d.h. Bedingung (3) und auch die Form des Tests  $\varphi^*$  ändern sich nicht unter linear affinen Transformationen  $\tilde{T}(x) = a \cdot T(x) + b$  ( $a \neq 0$ ). Also ist

$$\tilde{\varphi}^*(x) = \begin{cases} 1, & \tilde{T}(x) < \tilde{c}_1^* \text{ oder } \tilde{T}(x) > \tilde{c}_2^* \\ \tilde{\gamma}_i^*, & \tilde{T}(x) = \tilde{c}_i^* \quad (i = 1, 2) \\ 0, & \tilde{c}_1^* < \tilde{T}(x) < \tilde{c}_2^* \end{cases}$$

mit  $E_{\vartheta_0}\tilde{\varphi}^* \stackrel{!}{=} \alpha$ ,  $E_{\vartheta_0}[\tilde{\varphi}^*\tilde{T}] = \alpha \cdot E_{\vartheta_0}\tilde{T}$  ebenfalls UMPU-Test zum Niveau  $\alpha$  für  $H_0$  gegen  $H_1$ .

- Sei  $P_{\vartheta_0}^T$  symmetrisch bezüglich  $t_0$ , d.h.

$$P_{\vartheta_0}(T - t_0 \leq -t) = P_{\vartheta_0}(T - t_0 \geq t) \quad \forall t \in \mathbb{R}$$

Sei

$$\varphi^*(x) = \begin{cases} 1, & |T(x) - t_0| > c^* \\ \gamma^*, & |T(x) - t_0| = c^* \\ 0, & |T(x) - t_0| < c^* \end{cases}$$

mit  $P_{\vartheta_0}(T(X) - t_0 > \underbrace{c^*}_{>0}) + \gamma^* P_{\vartheta_0}(T(X) - t_0 = c^*) \stackrel{!}{=} \frac{\alpha}{2}$ .

$\Rightarrow P_{\vartheta_0}(|T(X) - t_0| > c^*) + \gamma^* P_{\vartheta_0}(|T(X) - t_0| = c^*) = \alpha$ , d.h.  
 $E_{\vartheta_0} \varphi^* = \alpha$  (\*).

Weiter gilt:  $E_{\vartheta_0} T(X) = t_0$ ,  $\varphi^*$  symmetrisch bezüglich  $t_0$

$$\Rightarrow E_{\vartheta_0}[\varphi^* \cdot T] = \underbrace{E_{\vartheta_0}[(T - t_0) \cdot \varphi^*]}_{=0 \text{ s.u.}} + t_0 E_{\vartheta_0} \varphi^* \stackrel{(*)}{=} t_0 \cdot \alpha = \alpha \cdot E_{\vartheta_0} T$$

[Betrachte  $g(t) = (t - t_0) \cdot \varphi^*(t)$   
 $\Rightarrow E_{\vartheta_0}[(T - t_0) \cdot \varphi^*(T)] = \int g(t) P_{\vartheta_0}^T(dt) = 0.]$

D.h. auch die zweite Bedingung in (3) ist erfüllt.

$\varphi^*$  ist also UMPU-Test zum Niveau  $\alpha$  für  $H_0$  gegen  $H_1$ .

Bestimmung von  $c^*, \gamma^*$  also wie beim einseitigen UMP-Test zum Niveau  $\frac{\alpha}{2}$ .

Bemerkung:

Form des Tests bleibt unverändert unter streng monotonen Transformationen  $\tilde{T}(x) = h(|T(x) - t_0|)$ .

### 12.3 Beispiel (Zweiseitiger Gauss-Test)

$X_1, \dots, X_n \stackrel{i.i.d.}{\sim} \mathcal{N}(\mu, \sigma_0^2), \sigma_0^2 > 0$  bekannt.

$$H_0 : \mu = \mu_0 \text{ gegen } H_1 : \mu \neq \mu_0$$

Verteilung von  $X = (X_1, \dots, X_n)$  ist einparametrische Exponentialfamilie mit  $\vartheta = \frac{\mu}{\sigma_0^2}$ ,  $T(x) = \sum_{i=1}^n x_i$ ,  $\sum_{i=1}^n X_i \sim \mathcal{N}(n\mu_0, n\sigma_0^2)$  unter  $H_0$ .

Linear affine Transformation

$$\tilde{T}(x) = \frac{T(x) - n\mu_0}{\sqrt{n\sigma_0^2}} = \sqrt{n} \frac{\bar{x}_n - \mu_0}{\sigma_0}$$

liefert  $P_{\mu_0}^{\tilde{T}} = \mathcal{N}(0, 1)$ , also symmetrisch bezüglich 0.

Verteilungsfunktion ist stetig

$$\Rightarrow \varphi^* = \begin{cases} 1, & \sqrt{n} \left| \frac{\bar{x}_n - \mu_0}{\sigma_0} \right| > z_{1-\frac{\alpha}{2}} \\ 0, & \sqrt{n} \left| \frac{\bar{x}_n - \mu_0}{\sigma_0} \right| \leq z_{1-\frac{\alpha}{2}} \end{cases}$$

ist UMPU-Test für  $H_0$  gegen  $H_1$ .

## 12.4 Beispiel

$X = (X_1, \dots, X_n)$ ,  $X_i \overset{uiv}{\sim} \text{Bin}(1, p)$ ,  $0 < p < 1$

$$H_0 : p = p_0 \text{ gegen } H_1 : p \neq p_0$$

Einparametrische Exponentialfamilie mit  $\vartheta = \log \frac{p}{1-p}$ ,  $T(x) = \sum_{i=1}^n x_i$ ,  
 $\sum_{i=1}^n X_i \sim \text{Bin}(n, p_0)$  unter  $H_0$ .

Im Allgemeinen nicht symmetrisch! UMPU-Test:

$$\Rightarrow \varphi^*(x) = \begin{cases} 1, & \sum x_i < c_1^* \text{ oder } \sum x_i > c_2^* \\ \gamma_i^*, & \sum x_i = c_i^* \\ 0, & c_1^* < \sum x_i < c_2^* \end{cases}$$

mit (komplizierten) Bedingungen für  $c_1^*$ ,  $c_2^*$ ,  $\gamma_1^*$ ,  $\gamma_2^*$ .

In der Praxis oft:

Konstruktion des Tests aus zwei einseitigen UMP-Tests zum Niveau  $\frac{\alpha}{2}$ ,  
 ist aber nicht UMPU.

Im Folgenden Exponentialfamilie mit

$$(4) \quad f(x, \vartheta, \xi) = c(\vartheta, \xi) \cdot \exp(\vartheta \cdot U(x) + \sum_{i=1}^k \xi_i T_i(x)) \cdot h(x)$$

$(\vartheta, \xi) \in \Theta \subset \mathbb{R} \times \mathbb{R}^k$ ,  $\Theta$  konvex,  $\dot{\Theta} \neq \emptyset$ .

Zu testen:

$$H_0 : \vartheta \leq \vartheta_0 \text{ gegen } H_1 : \vartheta > \vartheta_0$$

bzw.

$$\tilde{H}_0 : \vartheta = \vartheta_0 \text{ gegen } \tilde{H}_1 : \vartheta \neq \vartheta_0$$

$\xi = (\xi_1, \dots, \xi_k)$  ist Störparameter,  $T(x) = (T_1(x), \dots, T_k(x))$

Für festes  $t$  ist Dichte in (4) einparametrische Exponentialfamilie.

[Genauer: Man kann zeigen, dass die bedingte Verteilung  $P_{\vartheta, \xi}^{U|T=t}$  eine einparametrische Exponentialfamilie mit Dichte

$$c_t(\vartheta) \cdot e^{\vartheta \cdot U} h(x)$$

(unabhängig von  $\xi$ ) ist.]

$\Rightarrow$  (bedingte) UMP- bzw. UMPU-Tests für  $H_0$  bzw.  $\tilde{H}_0$  existieren.

Es lässt sich zeigen, dass diese bedingten Tests auch für zufälliges  $T = T(X)$  optimal sind:

**12.5 Satz**

a) Der Test  $\varphi_1$ , definiert durch

$$\varphi_1(x) = \begin{cases} 1, & U > c(t) \\ \gamma(t), & U = c(t) \\ 0, & U < c(t) \end{cases}$$

wobei  $E_{\vartheta_0}[\varphi_1(U, T)|T = t] \stackrel{!}{=} \alpha$ , ist UMPU-Test<sup>29</sup> zum Niveau  $\alpha$  für  $H_0$  gegen  $H_1$ .

b) Der Test  $\varphi_2$ , definiert durch<sup>30</sup>

$$\varphi_2(x) = \begin{cases} 1, & U < c_1(t) \text{ oder } U > c_2(t) \\ \gamma_i^*, & U = c_i(t) \\ 0, & c_1(t) < U < c_2(t) \end{cases}$$

wobei  $E_{\vartheta_0}[\varphi_2(U, T)|T = t] \stackrel{!}{=} \alpha$ ,

$$E_{\vartheta_0}[\varphi_2(U, T) \cdot U|T = t] \stackrel{!}{=} \alpha \cdot E_{\vartheta_0}[U|T = t]$$

ist UMPU-Test zum Niveau  $\alpha$  für  $\tilde{H}_0$  gegen  $\tilde{H}_1$ .

Die Tests aus 12.5 können manchmal so transformiert werden, dass  $c(t), \gamma(t)$  beziehungsweise  $c_1(t), c_2(t), \gamma_i(t)$  nicht von  $t$  abhängen.

**12.6 Satz**

Unter der Verteilungsannahme (4) sei  $V = h(U, T)$  eine unter  $\vartheta = \vartheta_0$  von  $T$  unabhängige reellwertige Statistik. Dann gilt:

a) Ist  $h(u, t)$  streng monoton wachsend in  $u$  bei festem  $t$ , so ist

$$\tilde{\varphi}_1(v) = \begin{cases} 1, & v > \tilde{c} \\ \tilde{\gamma}, & v = \tilde{c} \\ 0, & v < \tilde{c} \end{cases}$$

wobei  $E_{\vartheta_0}\tilde{\varphi}_1(V) = \alpha$ , UMPU-Test zum Niveau  $\alpha$  für  $H_0$  gegen  $H_1$ .

<sup>29</sup>Kein Schreibfehler! Test ist kein UMP-Test sondern nur UMPU!

<sup>30</sup>besser:  $\gamma_i(t)$

b) Gilt  $h(u, t) = a(t)u + b(t)$ ,  $a(t) > 0$  so ist

$$\tilde{\varphi}_2(v) = \begin{cases} 1, & v < \tilde{c}_1 \text{ oder } v > \tilde{c}_2 \\ \tilde{\gamma}_i, & v = \tilde{c}_i \\ 0, & \tilde{c}_1 < v < \tilde{c}_2 \end{cases}$$

wobei  $E_{\vartheta_0} \tilde{\varphi}_2(V) = \alpha$ ,  $E_{\vartheta_0}[\tilde{\varphi}_2(V)V] = \alpha E_{\vartheta_0}(V)$  UMPU-Test zum Niveau  $\alpha$  für  $\tilde{H}_0$  gegen  $\tilde{H}_1$ .

Beweis:

a) Nach Korollar 11.11 bleibt die Form des Tests unter streng monotoner Transformation unverändert, man erhält also einen Test der Form  $\tilde{\varphi}_1$  mit  $\tilde{c} = \tilde{c}(t)$ ,  $\tilde{\gamma} = \tilde{\gamma}(t)$ . Nach Voraussetzung ist  $V$  aber unabhängig von  $T$  unter  $\vartheta = \vartheta_0$ , deshalb hängen  $\tilde{c}, \tilde{\gamma}$  nicht von  $t$  ab.

b) folgt analog mit Bemerkung 12.2(a)

Nachweis der Unabhängigkeit von  $V$  und  $T$ ?

Übliche Methoden der Wahrscheinlichkeitstheorie, oder

## 12.7 Satz (Basu's Theorem)

Sei  $\varphi = \{P_\vartheta : \vartheta \in \Theta\}$ . Statistik  $T$  sei suffizient und vollständig für  $\vartheta$ . Ist  $V$  eine Statistik deren Verteilung nicht von  $\vartheta$  abhängt, so sind  $V$  und  $T$  stochastisch unabhängig.<sup>31</sup>

Beispiel:

$X_1, \dots, X_n \stackrel{iid}{\sim} \mathcal{N}(\mu, \sigma_0^2)$ ,  $\sigma_0^2 > 0$  bekannt,  $\Theta = \{\mu : \mu \in \mathbb{R}\}$ ,  $T = \sum_{i=1}^n X_i$  suffizient und vollständig für  $\mu$ .

$$V = \underbrace{\sum_{i=1}^n (X_i - \bar{X}_n)^2}_{(*)}$$

$(*) = \sum_i ((X_i - \mu)(\bar{X}_n - \mu))^2 = \sum_{i=1}^n (Y_i - \bar{Y}_n)^2$  wobei  $Y_i \sim \mathcal{N}(0, \sigma_0^2)$   
Verteilung von  $V$  unabhängig von  $\mu$  ( $V \sim \sigma_0^2 \chi_{n-1}^2$ ).

$\stackrel{12.7}{\Rightarrow}$   $V$  und  $T$  sind unabhängig.

<sup>31</sup> $V$  „ancillary“

Beweis:

Sei  $g$  beliebige beschränkte Funktion,  $m = E_{\vartheta}g(V)$  (unabhängig von  $\vartheta$  nach Voraussetzung).

$$h(T(x)) := E_{\vartheta}[g(V) - m | T = T(x)]$$

unabhängig von  $\vartheta$ , da  $T$  suffizient. Wegen

$$E_{\vartheta}h(T) = E_{\vartheta}[E_{\vartheta}[g(V) - m | T]] = 0 \forall \vartheta \in \Theta$$

und der Vollständigkeit von  $T$  folgt  $h(T) = 0$   $P_{\vartheta}$  - *f.s.*, also

$$E_{\vartheta}[g(V) | T] = m = E_{\vartheta}g(V) \quad P_{\vartheta} - \text{f.s.}$$

und somit die Unabhängigkeit von  $V$  und  $T$ .

## 12.8 Korollar

Sei  $\varphi$  Exponentialfamilie wie in (4), wobei  $\vartheta (= \vartheta_0)$  fest gewählt ist. Hängt die Verteilung einer Statistik  $V$  nicht von  $\xi$  ab, so sind  $V$  und  $T$  unabhängig.

Beweis:

Nach Beispiel 7.7 und 7.12 ist  $T$  vollständig und suffizient für  $\xi$ .  
12.7  $\Rightarrow$  Behauptung.

## 12.9 Beispiel (1-Stichproben-t-Test)

$X_1, \dots, X_n \stackrel{uiv}{\sim} \mathcal{N}(\mu, \sigma^2)$ ,  $\vartheta = (\mu, \sigma^2) \in \Theta = \mathbb{R} \times \mathbb{R}_{>0}$ ,  $X = (X_1, \dots, X_n)$

a)  $H_0 : \mu \leq \mu_0$  gegen  $H_1 : \mu > \mu_0$

2-parametrische Exponentialfamilie nach Beispiel 6.3, hat die Form in (4) mit  $\vartheta = \frac{\mu}{\sigma^2}$ ,  $\xi = -\frac{1}{2\sigma^2}$ ,  $U(x) = \sum_{i=1}^n x_i$ ,  $T(x) = \sum_{i=1}^n x_i^2$ .

Ohne Einschränkung sei  $\mu_0 = 0$ , andernfalls betrachte man  $x_i - \mu_0$  anstelle der  $x_i$ .

$H_0$ ,  $H_1$  sind dann äquivalent zu  $H_0 : \vartheta \leq 0$ ,  $H_1 : \vartheta > 0$ .

Betrachte:

$$v = \frac{\sqrt{n}\bar{x}_n}{\sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x}_n)^2}} = \frac{1}{\sqrt{n}} \frac{u}{\sqrt{\frac{t - \frac{u^2}{n}}{n-1}}} =: h(u, t)$$

$\frac{\partial h(u, t)}{\partial u} > 0 \Rightarrow h(u, t)$  streng monoton wachsend in  $u$  bei festem  $t$ .  
(Beachte:  $t > \frac{u^2}{n} > 0$ .)

Weiter gilt: Unter  $\vartheta = \vartheta_0$  gilt  $V \sim t_{n-1}$ , also unabhängig von  $\xi$ .

$\stackrel{12.8}{\Rightarrow}$   $V$  und  $T$  sind stochastisch unabhängig (unter  $\vartheta = \vartheta_0$ ).

$\stackrel{12.6(a)}{\Rightarrow}$  Der UMPU-Test für  $H_0 : \mu \leq \mu_0$  gegen  $\mu > \mu_0$  zum Niveau  $\alpha$  ist

$$\tilde{\varphi}_1(v) = \begin{cases} 1, & \sqrt{n} \frac{\bar{x}_n - \mu_0}{s} \geq t_{n-1; 1-\alpha} \\ 0, & \sqrt{n} \frac{\bar{x}_n - \mu_0}{s} < t_{n-1; 1-\alpha} \end{cases}$$

b)  $\tilde{H}_0 : \mu = \mu_0$  gegen  $\tilde{H}_1 : \mu \neq \mu_0$

Ohne Einschränkung  $\mu_0 = 0$ , dann  $\tilde{H}_0 : \vartheta = \vartheta_0 = 0$ ,  $\tilde{H}_1 : \vartheta \neq \vartheta_0$

$$h(u, t) = \frac{1}{\sqrt{n}} \frac{u}{\sqrt{\frac{t-u^2/n}{n-1}}}$$

nicht linear in  $u$ .

Betrachte

$$\tilde{v} = \tilde{h}(u, t) = \frac{u}{\sqrt{t}} = \frac{\sum x_i}{\sqrt{\sum x_i^2}}$$

Unter  $\vartheta = 0$  gilt  $\tilde{V} \sim \frac{\sum Y_i}{\sqrt{\sum Y_i^2}}$ , wobei  $Y_i \sim \mathcal{N}(0, 1)$ .<sup>32</sup>

$\Rightarrow$  Verteilung von  $\tilde{V}$  ist unabhängig von  $\xi$  und symmetrisch um 0.

Nach 12.6(b) existiert ein UMPU-Test  $\tilde{\varphi}_2(\tilde{v})$ , der wegen der Symmetrie der Verteilung von  $\tilde{V}$  nach 12.2(b) einen Ablehnbereich der Form  $|\tilde{v}| > \tilde{c}$  hat.

Nun gilt

$$v = h(u, t) = g(\tilde{v}) = \sqrt{\frac{n-1}{n}} \frac{\tilde{v}}{\sqrt{1 - \tilde{v}^2/n}}$$

bzw.  $|v| = g(|\tilde{v}|)$ .

$g(|\tilde{v}|)$  ist streng monoton wachsend auf  $[0, \sqrt{n}]$ <sup>33</sup>, so dass nach Bemerkung in 12.2(b) der UMPU-Test auch auf einem Ablehnbereich der Form  $|v| \geq c$  basieren kann. Somit ist

$$\tilde{\varphi}_2(x) = \begin{cases} 1, & \sqrt{n} \frac{|\bar{x}_n - \mu_0|}{s} \geq t_{n-1; 1-\frac{\alpha}{2}} \\ 0, & \sqrt{n} \frac{|\bar{x}_n - \mu_0|}{s} < t_{n-1; 1-\frac{\alpha}{2}} \end{cases}$$

UMPU-Test für  $\tilde{H}_0$  gegen  $\tilde{H}_1$ .

<sup>32</sup>Erweitere  $\tilde{v}$  mit  $\frac{1}{\sigma}$  um dies zu erkennen!

<sup>33</sup>Beachte:  $\tilde{v} \in (-\sqrt{n}, \sqrt{n})$  (nachrechenbar)

### 12.10 Bemerkung

Ähnliche Überlegungen zeigen, dass auch der ein- bzw. zweiseitige 2-Stichproben-t-Test UMPU-Test ist.  
(z.B. Lehmann/Romano, S. 157-161, 3. ed.)

### 12.11 Beispiel (Unabhängigkeitstest unter NV-Annahme)

$(X_1, Y_1), \dots, (X_n, Y_n) \stackrel{uiv}{\sim} \mathcal{N}_2(\mu, \nu, \sigma^2, \tau^2, \varrho)$ , also Dichte<sup>34</sup>

$$f((x_1, y_1), \dots, (x_n, y_n), \mu, \nu, \sigma^2, \tau^2, \varrho) = (2\pi\sigma\tau\sqrt{1-\varrho^2})^{-n}.$$

$$\exp\left(-\frac{1}{2(1-\varrho^2)}\left(\frac{1}{\sigma^2}\sum_i(x_i-\mu)^2 - \frac{2\varrho}{\sigma\tau}\sum_i(x_i-\mu)(y_i-\nu) + \frac{1}{\tau^2}\sum_i(y_i-\nu)^2\right)\right) \quad (*)$$

Zu testen:  $\tilde{H}_0$ :  $X_1, Y_1$  unabhängig;  $\tilde{H}_1$ :  $X_1, Y_1$  nicht unabhängig

Äquivalent:  $\tilde{H}_0$ :  $\varrho = 0$ ;  $\tilde{H}_1$ :  $\varrho \neq 0$

Bzw. die einseitige Hypothese  $H_0$ :  $\varrho \leq 0$  gegen  $H_1$ :  $\varrho > 0$ .

(\*) ist Exponentialfamilie wie in (4) mit

$$U = \sum_i x_i y_i, T_1 = \sum_i x_i^2, T_2 = \sum_i y_i^2, T_3 = \sum_i x_i, T_4 = \sum_i y_i$$

$$\vartheta = \frac{\varrho}{\sigma\tau(1-\varrho^2)}$$

$$\xi_1 = -\frac{1}{2\sigma^2(1-\varrho^2)}, \quad \xi_2 = -\frac{1}{2\tau^2(1-\varrho^2)},$$

$$\xi_3 = \frac{1}{1-\varrho^2}\left(\frac{\mu}{\sigma^2} - \frac{\nu\varrho}{\sigma\tau}\right), \quad \xi_4 = \frac{1}{1-\varrho^2}\left(\frac{\nu}{\tau^2} - \frac{\mu\varrho}{\sigma\tau}\right)$$

a)  $H_0$ :  $\vartheta \leq 0$  gegen  $H_1$ :  $\vartheta > 0$

Sei

$$R = \frac{\sum_i (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_i (X_i - \bar{X})^2 \cdot \sum_i (Y_i - \bar{Y})^2}}$$

empirischer Korrelationskoeffizient nach Pearson.

Transformation  $X_i \rightarrow \frac{X_i - \mu}{\sigma}$ ,  $Y_j \rightarrow \frac{Y_j - \nu}{\tau}$  ändert R nicht, deshalb hängt die Verteilung von R nicht von  $\mu, \nu, \sigma^2, \tau^2$  ab, sondern nur von  $\varrho$ .

Für  $\vartheta = 0$  ist die Verteilung von R also unabhängig von  $\xi_1, \xi_2, \xi_3, \xi_4$ .

<sup>34</sup> $\varrho$  ist Korrelationskoeffizient (s. Stochastik 1)

Korollar 12.8  $\Rightarrow$   $R$  ist unabhängig von  $(T_1, \dots, T_4)$  unter  $\vartheta = 0$ .

$\stackrel{12.6}{\Rightarrow}$  UMPU-Test hat Ablehnbereich der Form  $R \geq c$  oder äquivalent

$$w := \frac{R}{\sqrt{\frac{1-R^2}{n-2}}} \geq \tilde{c}$$

$[R = \frac{U-T_3T_4/n}{\sqrt{(T_1-T_3^2/n)(T_2-T_4^2/n)}}$  ist streng monoton wachsend in  $U$

$\Rightarrow w$  ist streng monoton wachsend<sup>35</sup> in  $U$ ]

Nach Aufgabe 36 gilt:  $w \sim t_{n-2}$  falls  $\varrho = 0$  (bzw.  $\vartheta = 0$ ).

Deshalb:

$$\varphi_1(w) = \begin{cases} 1, & w \geq t_{n-2, 1-\alpha} \\ 0, & w < t_{n-2, 1-\alpha} \end{cases}$$

UMPU-Test zum Niveau  $\alpha$  für  $H_0$  gegen  $H_1$ .

b) Test von  $\tilde{H}_0 : \vartheta = 0, \tilde{H}_1 : \vartheta \neq 0$

$R$  ist linear in  $U$  mit um 0 symmetrischer Verteilung für  $\vartheta = 0$

$\Rightarrow$  UMPU-Test hat Ablehnbereich der Form  $|R| \geq \tilde{c}$ .

Die Funktion  $x \rightarrow \frac{x}{\sqrt{1-x^2}}$  ist streng monoton wachsend für  $0 \leq x \leq 1$ ,  
woraus wie in 12.9(b) folgt:

$$\varphi_2(w) = \begin{cases} 1, & |w| \geq t_{n-2, 1-\frac{\alpha}{2}} \\ 0, & |w| < t_{n-2, 1-\frac{\alpha}{2}} \end{cases}$$

ist UMPU-Test zum Niveau  $\alpha$  für  $\tilde{H}_0 : \varrho = 0$  gegen  $\tilde{H}_1 : \varrho \neq 0$ .

---

<sup>35</sup> $w$  ist streng monoton wachsend in  $R$  (Beachte:  $R \in [-1, 1]$  und  $w'(R) > 0 \forall R \in (-1, 1)$ )

## 13 Konfidenzbereiche

Sei  $(\mathfrak{X}, \mathcal{B}, \{P_\vartheta : \vartheta \in \Theta\})$  statistisches Modell,  $g : \Theta \rightarrow \mathbb{R}^s$ .

### 13.1 Definition

Sei  $\alpha \in (0, 1)$ . Eine Abbildung  $C : \mathfrak{X} \rightarrow \mathcal{P}(\mathbb{R}^s)$  heißt **Konfidenzbereich** für  $g(\vartheta)$  zum Niveau  $1 - \alpha$  genau dann, wenn

- (1)  $\{x \in \mathfrak{X} : C(x) \ni g(\vartheta)\} \in \mathcal{B} \quad \forall \vartheta \in \Theta$
- (2)  $P_\vartheta(\{x \in \mathfrak{X} : C(x) \ni g(\vartheta)\}) \geq 1 - \alpha \quad \forall \vartheta \in \Theta$ .

Falls  $X : \Omega \rightarrow \mathfrak{X}$  eine Zufallsvariable mit Verteilung  $P_\vartheta$  ist, so die zweite Bedingung gleichbedeutend mit

$$P_\vartheta(C(X) \ni g(\vartheta)) \geq 1 - \alpha \quad \forall \vartheta \in \Theta.$$

Falls  $s = 1$  und  $C(x)$  für alle  $x \in \mathfrak{X}$  ein Intervall ist, so heißt  $C(\cdot)$  ein **Konfidenzintervall**.<sup>36</sup>

Beispiel:

$X = (X_1, \dots, X_n), X_1, \dots, X_n \stackrel{i.i.d.}{\sim} \mathcal{N}(\mu, \sigma^2), \vartheta = (\mu, \sigma^2), g(\vartheta) = \mu$

$$C(X) = \left[ \bar{X}_n - \frac{S_n}{\sqrt{n}} \cdot t_{n-1; 1-\frac{\alpha}{2}}, \bar{X}_n + \frac{S_n}{\sqrt{n}} \cdot t_{n-1; 1-\frac{\alpha}{2}} \right]$$

ist Konfidenzintervall zum Niveau  $1 - \alpha$  nach 2.4.

### 13.2 Bemerkung (Pivot-Methode)

Praktische Berechnung von Konfidenzintervallen:

Finde Funktion  $k$  so, dass die Verteilung von  $k(X, \vartheta)$  unabhängig von  $\vartheta$  ist, d.h., dass  $H(x) := P_\vartheta(k(X, \vartheta) \leq x)$  unabhängig von  $\vartheta$  ist.

Dann existieren Konstanten  $a, b$ :

$$P_\vartheta(a \leq k(X, \vartheta) \leq b) \geq 1 - \alpha \quad \forall \vartheta \in \Theta$$

<sup>36</sup>Anmerkung: Ermitteln wir z.B. das 95%-Konfidenzintervall für den wahren Erwartungswert einer Population, dann bedeutet dies, dass wir bei durchschnittlich 5 von 100 gleichgroßen Zufallsstichproben ein Konfidenzintervall ermitteln, das den Erwartungswert nicht enthält.

Falls man das Ereignis  $\{a \leq k(X, \vartheta) \leq b\}$  umschreiben kann als  $\{U(X) \leq g(\vartheta) \leq O(X)\}$ , so ist  $[U(X), O(X)]$  Konfidenzintervall für  $g(\vartheta)$  zum Niveau  $1 - \alpha$ .

Im Beispiel oben:  
Verteilung von

$$k(X, \vartheta) = \frac{\sqrt{n}(\bar{X}_n - \mu)}{S_n}$$

unabhängig von  $\vartheta = (\mu, \sigma^2)$ .

$\frac{\sqrt{n}(\bar{X}_n - \mu)}{S_n}$  ist Pivot für  $g(\vartheta) = \mu$ .

$[-t_{n-1; 1-\frac{\alpha}{2}} \leq k(X, \vartheta) \leq t_{n-1; 1-\frac{\alpha}{2}}] \rightarrow C(X)$  im Beispiel oben]

Weiteres Beispiel:

$X_1, \dots, X_n \stackrel{iid}{\sim} U(0, \vartheta)$ ,  $\vartheta > 0$ ,  $g(\vartheta) = \vartheta$

MLE<sup>37</sup> von  $\vartheta$ :  $X_{(n)} = \max_{1 \leq i \leq n} X_i$

Verteilungsfunktion von  $X_{(n)}$  ist  $(\frac{x}{\vartheta})^n$ ,  $0 \leq x \leq \vartheta$

$\Rightarrow$  Verteilungsfunktion von  $\frac{X_{(n)}}{\vartheta}$  ist  $x^n$ ,  $0 \leq x \leq 1$ , also ist  $\frac{X_{(n)}}{\vartheta}$  Pivot für  $\vartheta$ .

Wähle a,b so, dass

$$P_{\vartheta}(a \leq \frac{X_{(n)}}{\vartheta} \leq b) = b^n - a^n \stackrel{!}{=} 1 - \alpha \quad (\forall \vartheta \in \Theta)$$

Dann ist  $[\frac{X_{(n)}}{b}, \frac{X_{(n)}}{a}]$   $(1 - \alpha)$ -Konfidenzintervall für  $\vartheta$ .

Wie a und b wählen?

- Intervall  $[a, b]$  „kleinstmöglich“ wählen
- andere Optimalitätsbegriffe

### 13.3 Zusammenhang zwischen Konfidenzintervallen und (nichtrandomisierten) Tests

1.  $C(x)$  sei Konfidenzintervall zum Niveau  $1 - \alpha$  für  $\vartheta$  (d.h.

$$P_{\vartheta}(C(X) \ni \vartheta) \geq 1 - \alpha \quad \forall \vartheta \in \Theta).$$

Zu testen ist  $H_0 : \vartheta = \vartheta_0$  gegen  $H_1 : \vartheta \neq \vartheta_0$ .

Definiere Test  $\varphi$ :

$$\varphi(x) = \begin{cases} 1 & , \vartheta_0 \notin C(x) \\ 0 & , \vartheta_0 \in C(x) \end{cases}$$

---

<sup>37</sup>ML-Schätzer (*Estimator*)

Umfang von  $\varphi$ :

$$E_{\vartheta_0} \varphi(x) = 1 - \underbrace{P_{\vartheta_0}(\vartheta_0 \in C(x))}_{\geq 1-\alpha} \leq \alpha$$

d.h.  $\varphi$  ist Niveau  $\alpha$ -Test.

2. Umgekehrt sei für jedes  $\vartheta_0 \in \Theta$  ein Niveau  $\alpha$ -Test  $\varphi_{\vartheta_0}(x)$  für obige Situation gegeben (d.h.  $P_{\vartheta_0}(\varphi_{\vartheta_0}(X) = 0) \geq 1 - \alpha$ ,  $\vartheta_0 \in \Theta$ ).  
Definiere  $C^*(x) = \{\vartheta_0 : \varphi_{\vartheta_0}(x) = 0\}$

$$\Rightarrow P_{\vartheta}(C^*(X) \ni \vartheta) = P_{\vartheta}(\varphi_{\vartheta}(x) = 0) \geq 1 - \alpha \quad \forall \vartheta \in \Theta$$

d.h.  $C^*(X)$  ist  $(1 - \alpha)$ -Konfidenzbereich für  $\vartheta$ .

Beispiel (1 Stichproben-t-Test):

1.  $(1 - \alpha)$ -Konfidenzintervall für  $\mu$ :  $[\bar{x}_n - \frac{s_n}{\sqrt{n}} t_{n-1, 1-\frac{\alpha}{2}}, \bar{x}_n + \frac{s_n}{\sqrt{n}} t_{n-1, 1-\frac{\alpha}{2}}]$ .  
Lehne  $H_0 : \mu = \mu_0$  ab, falls  $\mu_0 \notin$  Konfidenzintervall.

$$\begin{aligned} &\hat{=} |\mu_0 - \bar{x}_n| > \frac{s_n}{\sqrt{n}} t_{n-1, 1-\frac{\alpha}{2}} \\ &\hat{=} \frac{\sqrt{n} |\bar{x}_n - \mu_0|}{s_n} > t_{n-1, 1-\frac{\alpha}{2}} \end{aligned}$$

2. Umgekehrt:

$$\frac{\sqrt{n} |\bar{x}_n - \mu_0|}{s_n} > t_{n-1, 1-\frac{\alpha}{2}}$$

Ablehnbereich für Test  $\varphi_{\mu_0}$  von  $H_0 : \mu = \mu_0$  gegen  $H_1 : \mu \neq \mu_0$  für jedes  $\mu_0 \in \mathbb{R}$ .

$$\begin{aligned} C^*(x) &= \{\mu : \varphi_{\mu}(x) = 0\} \\ &= \left\{ \mu : \frac{\sqrt{n} |\bar{x}_n - \mu|}{s_n} \leq t_{n-1, 1-\frac{\alpha}{2}} \right\} \\ &= \left\{ \bar{x}_n - \frac{s_n}{\sqrt{n}} t_{n-1, 1-\frac{\alpha}{2}} \leq \mu \leq \bar{x}_n + \frac{s_n}{\sqrt{n}} t_{n-1, 1-\frac{\alpha}{2}} \right\} \end{aligned}$$

$(1 - \alpha)$ -Konfidenzintervall für  $\mu$ .

Bemerkungen:

- (i) Es besteht also eine Dualität zwischen Signifikanztests und Konfidenzbereichen, allerdings nur, wenn eine ganze Schar von Hypothesen

$H_{\vartheta_0} : \vartheta = \vartheta_0$  getestet wird.

Bei Beschränkung auf einen Test (was bei praktischer Testdurchführung immer der Fall ist) ist der Test „weniger“ informativ.

[Allerdings: Bei Tests wird in der Praxis p-Wert (siehe Beispiel nach 11.4) angegeben  $\Rightarrow$  andere Information als Konfidenzintervall].

- (ii) UMP(U)-Tests führen auf Konfidenzbereiche, die gewisse (komplizierte) Optimalitätseigenschaften haben.  
(Im Allgemeinen aber nicht kürzeste Konfidenzintervalle.)

### 13.4 Definition

Ist für jedes  $n$  die Abbildung  $C_n : \mathfrak{X}_n \rightarrow \mathbb{R}^s$  ein Konfidenzbereich für  $g(\vartheta)$ , basierend auf  $(X_1, \dots, X_n)$ , und gilt

$$\lim_{n \rightarrow \infty} P_{\vartheta}(\{(x_1, \dots, x_n) \in \mathfrak{X}_n : C_n(x_1, \dots, x_n) \ni g(\vartheta)\}) = 1 - \alpha$$

für alle  $\vartheta \in \Theta$ , so heißt die Folge  $(C_n)$  ein **asymptotischer Konfidenzbereich** für  $g(\vartheta)$  zum Niveau  $1 - \alpha$ .

### 13.5 Beispiel

$X_1, \dots, X_n \stackrel{i.i.d.}{\sim} X$ ,  $EX^2 < \infty$ ,  $F(x) = P(X \leq x)$ ,  $\vartheta := F$ ,  
 $g(\vartheta) = \int x dF(x) = EX =: \mu$

$$S_n^2 := \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2 \xrightarrow{P} \sigma^2 := \text{Var}(X)$$

ZGWS:  $\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \xrightarrow{D} \mathcal{N}(0, 1)$

$$\Rightarrow \lim_{n \rightarrow \infty} P_{\vartheta} \left( \underbrace{\bar{X}_n - \frac{S_n}{\sqrt{n}} \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \leq \mu \leq \bar{X}_n + \frac{S_n}{\sqrt{n}} \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)}_{\text{asymptotisches Konfidenzintervall zum Niveau } 1 - \alpha} \right) = 1 - \alpha$$

### 13.6 Hilfssatz

$$Y \sim \mathcal{N}_k(0, \Sigma), \Sigma > 0 \implies Y^T \Sigma^{-1} Y \sim \chi_k^2.$$

Beweis:

$$\Sigma^{-1/2} Y \sim \mathcal{N}_k(0, I_k) \implies \|\Sigma^{-1/2} Y\|^2 = Y^T \Sigma^{-1} Y \sim \chi_k^2.$$

### 13.7 Asymptotische Konfidenzbereiche in parametrischen Modellen

Seien  $X_1, \dots, X_n \stackrel{i.i.d.}{\sim} f(\xi; \vartheta)$ ,  $\vartheta \in \Theta$ ,  $\Theta \subset \mathbb{R}^k$  offen und  $f$  eine reguläre Dichte im  $\mathbb{R}^s$  bezüglich  $\mu$  ( $= \lambda^s$  oder Zählmaß).

Sei  $\hat{\vartheta}_n$  eine Schätzfolge für  $\vartheta$  mit der Eigenschaft

$$\sqrt{n}(\hat{\vartheta}_n - \vartheta) \xrightarrow{D_\vartheta} \mathcal{N}_k(0, \Sigma(\vartheta)), \quad \vartheta \in \Theta \quad (1)$$

wobei  $\Sigma(\vartheta) > 0$  und  $\Sigma(\cdot)$  stetig.

Aus (1) und Hilfssatz 13.6 folgt, dass

$$n(\hat{\vartheta}_n - \vartheta)^T \Sigma(\hat{\vartheta}_n)^{-1} (\hat{\vartheta}_n - \vartheta) \xrightarrow{D_\vartheta} \chi_k^2, \quad \vartheta \in \Theta$$

das heißt

$$\lim_{n \rightarrow \infty} P_\vartheta \left( n(\hat{\vartheta}_n - \vartheta)^T \Sigma(\hat{\vartheta}_n)^{-1} (\hat{\vartheta}_n - \vartheta) \leq \chi_{k;1-\alpha}^2 \right) = 1 - \alpha \quad \forall \vartheta \in \Theta.$$

Da die Menge

$$\left\{ \vartheta \in \mathbb{R}^k : (\hat{\vartheta}_n - \vartheta)^T \Sigma(\hat{\vartheta}_n)^{-1} (\hat{\vartheta}_n - \vartheta) \leq \frac{\chi_{k;1-\alpha}^2}{n} \right\}$$

ein Ellipsoid in  $\mathbb{R}^k$  mit Zentrum  $\hat{\vartheta}_n$  ist, handelt es sich hier um einen elliptischen Konfidenzbereich für  $\vartheta$ .

Falls  $g : \mathbb{R}^k \rightarrow \mathbb{R}$  differenzierbar ist, so folgt aus (1), dass

$$\sqrt{n}(g(\hat{\vartheta}_n) - g(\vartheta)) \xrightarrow{D_\vartheta} \mathcal{N}(0, \sigma^2(\vartheta)),$$

wobei

$$\sigma^2(\vartheta) = g'(\vartheta)^T \Sigma(\vartheta) g(\vartheta).$$

Somit gilt

$$\frac{\sqrt{n}(g(\hat{\vartheta}_n) - g(\vartheta))}{\sigma(\hat{\vartheta}_n)} \xrightarrow{D_\vartheta} \mathcal{N}(0, 1).$$

Mit  $r_n = \sigma(\hat{\vartheta}_n) \cdot \Phi^{-1} \left( 1 - \frac{\alpha}{2} \right) / \sqrt{n}$  folgt

$$\lim_{n \rightarrow \infty} P_\vartheta \left( g(\hat{\vartheta}_n) - r_n \leq g(\vartheta) \leq g(\hat{\vartheta}_n) + r_n \right) = 1 - \alpha.$$

Man hat also einen asymptotischen Konfidenzbereich für  $g(\vartheta)$  konstruiert.

### 13.8 Beispiele

- a)  $X_1, \dots, X_n \stackrel{uiv}{\sim} \text{Bin}(1, p)$ ,  $0 < p < 1$ ,  $\vartheta = p$ ,  $\hat{p}_n = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X}_n$   
 ZGWS:

$$\sqrt{n}(\hat{p}_n - p) \xrightarrow{D} \mathcal{N}(0, \underbrace{p(1-p)}_{=\Sigma(\vartheta)})$$

$g: \mathbb{R} \rightarrow \mathbb{R}$ ,  $g(p) = \log \frac{p}{1-p}$  „logit“-Funktion

$$g'(p) = \frac{1}{p(1-p)}$$

$$\Rightarrow \sigma^2(p) = g'(p)^2 \Sigma(p) = \frac{1}{p(1-p)^2} = \frac{1}{p} + \frac{1}{1-p}$$

$$\Rightarrow \sqrt{n} \left( \log \frac{\hat{p}_n}{1-\hat{p}_n} - \log \frac{p}{1-p} \right) \xrightarrow{D} \mathcal{N}\left(0, \frac{1}{p(1-p)}\right)$$

und

$$\left[ \log \frac{\hat{p}_n}{1-\hat{p}_n} - \frac{\Phi^{-1}(1-\frac{\alpha}{2})}{\sqrt{n\hat{p}_n(1-\hat{p}_n)}}, \log \frac{\hat{p}_n}{1-\hat{p}_n} + \frac{\Phi^{-1}(1-\frac{\alpha}{2})}{\sqrt{n\hat{p}_n(1-\hat{p}_n)}} \right]$$

ist asymptotisches  $(1-\alpha)$ -Konfidenzintervall für  $\log \frac{p}{1-p}$ .

- b) Konfidenzintervall für „log odds ratio“  
 $X_1, \dots, X_n \sim \text{Bin}(1, p)$ ,  $Y_1, \dots, Y_n \sim \text{Bin}(1, q)$

$$\Theta = \log \frac{\frac{p}{1-p}}{\frac{q}{1-q}}, \quad \Theta = 0 \Leftrightarrow p = q$$

siehe Übung

### 13.9 Beispiel

Sei  $X_1, \dots, X_n \stackrel{uiv}{\sim} \mathcal{N}_2(\mu, \Sigma)$ ,  $X_i = \begin{pmatrix} X_i^{(1)} \\ X_i^{(2)} \end{pmatrix}$ ,  $\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$

$\Sigma$  regulär,  $\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix}$ ,  $\bar{X}_n = \begin{pmatrix} \bar{X}_n^{(1)} \\ \bar{X}_n^{(2)} \end{pmatrix}$  mit  $\bar{X}_n^{(k)} = \frac{1}{n} \sum_{i=1}^n X_i^{(k)}$ ,

$k = 1, 2$

$\Sigma$  bekannt:  $\sqrt{n}(\bar{X}_n - \mu) \sim \mathcal{N}_2(0, \Sigma)$

$$\stackrel{13.6}{\Rightarrow} n(\bar{X}_n - \mu)^T \Sigma^{-1} (\bar{X}_n - \mu) \sim \chi_2^2$$

$$\Rightarrow P_\mu \left( \underbrace{n(\bar{X}_n - \mu)^T \Sigma^{-1} (\bar{X}_n - \mu)}_{\text{elliptischer } (1-\alpha)\text{-Konfidenzbereich für } \mu} \leq \chi_{2;1-\alpha}^2 \right) = 1 - \alpha$$

$\Sigma$  unbekannt: Konsistenter Schätzer für  $\Sigma$  ist

$$\hat{\Sigma}_n = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)(X_i - \bar{X}_n)^T$$

$$\vartheta = (\mu, \Sigma), \hat{\vartheta}_n = (\bar{X}_n, \hat{\Sigma}_n)$$

Für  $n > d(= 2)$ <sup>38</sup> ist  $\hat{\Sigma}_n$  nicht singulär mit Wahrscheinlichkeit 1.

$$\Rightarrow n(\bar{X}_n - \mu)^T \hat{\Sigma}_n^{-1} (\bar{X}_n - \mu) \xrightarrow{D} \chi_2^2$$

Betrachte  $g(\vartheta) = \mu_1 - \mu_2$ .

$$g'(\vartheta) = \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \sigma^2(\vartheta) = (1, -1)\Sigma \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \sigma_{11} - 2\sigma_{12} + \sigma_{22}$$

$$\Rightarrow \frac{\sqrt{n}((\bar{X}_n^{(1)} - \bar{X}_n^{(2)}) - (\mu_1 - \mu_2))}{\sigma(\hat{\vartheta}_n)} \xrightarrow{D} \mathcal{N}(0, 1)$$

---

<sup>38</sup>d ist Dimension



## 14 Lineare statistische Modelle

### 14.1 Definition

Es seien  $X = (X_1, \dots, X_n)^T$  ein (beobachtbarer) Zufallsvektor,  
 $C = (c_{ij})_{\substack{i=1, \dots, n \\ j=1, \dots, s}}$  eine bekannte  $n \times s$ -Matrix mit  $\text{Rang}(C) = s$   
 (insbesondere  $n \geq s$ ),  $\vartheta = (\vartheta_1, \dots, \vartheta_s)^T$  unbekannter Parametervektor,  
 $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)^T$  ein (nicht beobachtbarer) Zufallsvektor mit

$$E(\varepsilon) = 0, \quad \text{Var}(\varepsilon) = E(\varepsilon \cdot \varepsilon^T) = \sigma^2 \cdot I_n$$

$\sigma^2$  unbekannt.

Ein **lineares Modell (LM)** wird beschrieben durch die Gleichung

$$X = C\vartheta + \varepsilon \quad (1)$$

also

$$\begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} = \begin{pmatrix} c_{11} & \cdots & c_{1s} \\ \vdots & & \vdots \\ c_{n1} & \cdots & c_{ns} \end{pmatrix} \begin{pmatrix} \vartheta_1 \\ \vdots \\ \vartheta_s \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

$C$  heißt „Designmatrix“.

(1) heißt klassisch, falls  $\varepsilon \sim \mathcal{N}_n(0, \sigma^2 I_n)$ .

Bemerkungen:

- a) Im klassischen LM gilt:  $X \sim \mathcal{N}_n(C\vartheta, \sigma^2 I_n)$ .  
 Die Beobachtungen  $X_1, \dots, X_n$  sind also unabhängig, aber nicht identisch verteilt.
- b)  $\text{Rang}(C) = s \Leftrightarrow C^T C$  nicht singulär  
 Denn<sup>39</sup>:

$$\begin{aligned} C^T C \text{ singulär} &\Leftrightarrow \exists u \in \mathbb{R}^s, u \neq 0 : C^T C u = 0 \\ &\Leftrightarrow \exists u \in \mathbb{R}^s, u \neq 0 : u^T C^T C u = (Cu)^T C u = 0 \\ &\Leftrightarrow \exists u \in \mathbb{R}^s, u \neq 0 : C u = 0 \\ &\Leftrightarrow \text{Rang}(C) < s \end{aligned}$$

<sup>39</sup>In der Hinrichtung multipliziere  $C^T C u = 0$  mit  $u^T$ , in der Rückrichtung multipliziere  $C u = 0$  mit  $C^T$ .

## 14.2 Beispiele

a)  $X_i = \vartheta + \varepsilon_i, i = 1, \dots, n$

$$(s = 1, C = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix})$$

(wiederholte Messung)

b)  $X_i = a + bt_i + \varepsilon_i, i = 1, \dots, n$

$$(s = 2, a = \vartheta_1, b = \vartheta_2, C = \begin{pmatrix} 1 & t_1 \\ \vdots & \vdots \\ 1 & t_n \end{pmatrix})$$

(einfache lineare Regression)

c)  $X_i = a + bt_i + ct_i^2 + \varepsilon_i, i = 1, \dots, n$

$$(s = 3, \vartheta = (a, b, c)^T, C = \begin{pmatrix} 1 & t_1 & t_1^2 \\ \vdots & \vdots & \vdots \\ 1 & t_n & t_n^2 \end{pmatrix})$$

(einfache quadratische Regression)

d)  $X_i = \sum_{j=1}^s \vartheta_j \cdot f_j(t_i) + \varepsilon_i, i = 1, \dots, n$   
 $f_1, \dots, f_s$  beliebige gegebene Funktionen! (allgemeine (lineare) Regression)

z.B.  $f_j(t) = \sin(\omega_j \cdot t)$  (trigonometrische Regression)

e)  $X_i = a + bu_i + cv_i + \dots + gz_i + \varepsilon_i, i = 1, \dots, n$

$$\vartheta = \begin{pmatrix} a \\ \vdots \\ g \end{pmatrix}, C = \begin{pmatrix} 1 & u_1 & v_1 & \cdots & z_1 \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & u_n & v_n & \cdots & z_n \end{pmatrix}$$

(multiple lineare Regression)

f)  $X_{1,i} = \vartheta_1 + \varepsilon_{1,i}, i = 1, \dots, n_1$   
 $X_{2,i} = \vartheta_2 + \varepsilon_{2,i}, i = 1, \dots, n_2$

$$\begin{pmatrix} X_{1,1} \\ \vdots \\ X_{1,n_1} \\ X_{2,1} \\ \vdots \\ X_{2,n_2} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \\ 0 & 1 \\ \vdots & \vdots \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \vartheta_1 \\ \vartheta_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_{1,1} \\ \vdots \\ \varepsilon_{1,n_1} \\ \varepsilon_{2,1} \\ \vdots \\ \varepsilon_{2,n_2} \end{pmatrix}$$

(2-Stichproben-Modell)

- g)  $X_{i,j} = \vartheta_i + \varepsilon_{i,j}$ ,  $i = 1, \dots, k$ ,  $j = 1, \dots, n_i$   
 (Modell der einfachen Varianzanalyse, 1-faktorielle ANOVA)  
 z.B. Effekt  $X_{i,j}$  bei  $k$  unterschiedlichen Behandlungen

### 14.3 Schätzung von $\vartheta$

Sei  $R(C) := \{C\vartheta : \vartheta \in \mathbb{R}^s\}$   $s$ -dimensionaler Unterraum des  $\mathbb{R}^n$ .

14.1(1) besagt  $EX \in R(C)$ .

Forderung:  $\|X - C\vartheta\|^2 = \min_{\vartheta}$ ! (kleinste-Quadrate-Methode; vgl. 4.6)

Lösung:

$$\hat{\vartheta} = \hat{\vartheta}(X) = (C^T C)^{-1} \cdot C^T X$$

Beweis:

Wegen  $\mu(\vartheta) = C\vartheta$  folgt  $M(\vartheta) = \left( \frac{\partial \mu_i}{\partial \vartheta_j} \right)_{i,j} = C$  in 4.6 und somit die Normalgleichung  $C^T C \vartheta = C^T X$ .

Da  $C^T C$  nach Bemerkung 14.1(b) invertierbar ist, ist

$$\hat{\vartheta} = \hat{\vartheta}(X) = (C^T C)^{-1} \cdot C^T X$$

die (einzige) Lösung.

Bemerkung:

Es gilt<sup>40</sup>:

$$E_{\vartheta, \sigma^2}(\hat{\vartheta}) = (C^T C)^{-1} C^T \underbrace{E_{\vartheta, \sigma^2}(X)}_{=C\vartheta} = \vartheta$$

d.h.  $\hat{\vartheta}$  ist erwartungstreu für  $\vartheta$ .

$$\text{Var}_{\vartheta, \sigma^2}(\hat{\vartheta}) = (C^T C)^{-1} C^T \underbrace{\text{Var}_{\vartheta, \sigma^2}(X)}_{=\text{Var}_{\vartheta, \sigma^2}(\varepsilon) = \sigma^2 \cdot I_n} \cdot C (C^T C)^{-1} = \sigma^2 (C^T C)^{-1}$$

Beispiele:

- a) In 14.2(b) (einfache lineare Regression) ist (vgl. 4.7)

$$\hat{\vartheta}_1 = \bar{X} - \hat{\vartheta}_2 \bar{t}, \quad \hat{\vartheta}_2 = \frac{\sum_{i=1}^n t_i X_i - n \cdot \bar{t} \cdot \bar{X}}{\sum_{i=1}^n (t_i - \bar{t})^2}$$

<sup>40</sup>Beachte:  $(A^T)^{-1} = (A^{-1})^T$



$$\Rightarrow \text{Var}_{\vartheta, \sigma^2}(S) - \text{Var}_{\vartheta, \sigma^2}(T) = \sigma^2 b^T \underbrace{(I_n - \underbrace{C(C^T C)^{-1} C^T}_{=: P})}_{=: Q} b$$

Wegen  $P = P^T = P^2$  folgt  $Q = Q^T = Q^2$  (vgl. Aufgabe 44) folgt

$$b^T Q b = b^T Q^2 b = b^T Q^T Q b = \|Q b\|^2 \geq 0$$

$\Rightarrow$  Behauptung

Beispiele:

a) 1-faktorielle ANOVA (14.2(g), Beispiel 14.3(b))

$$a^T = (0, \dots, 0, \underbrace{1}_{a_i}, 0, \dots, 0, \underbrace{-1}_{a_j}, 0, \dots, 0), \quad a^T \vartheta = \vartheta_i - \vartheta_j$$

Differenz der Erwartungswerte der i-ten und j-ten Gruppe.

$T = a^T \hat{\vartheta} = \bar{X}_{i+} - \bar{X}_{j+}$  ist BLUE für  $a^T \vartheta$ .

b) einfache lineare Regression

$$a = \begin{pmatrix} 1 \\ t^* \end{pmatrix}, \quad a^T \vartheta = \vartheta_1 + \vartheta_2 t^*$$

$T = a^T \hat{\vartheta} = \hat{\vartheta}_1 + \hat{\vartheta}_2 t^*$  ist BLUE.

Hier:

$$C = \begin{pmatrix} 1 & t_1 \\ \vdots & \vdots \\ 1 & t_n \end{pmatrix}$$

$$C^T C = \begin{pmatrix} n & n\bar{t} \\ n\bar{t} & \sum t_i^2 \end{pmatrix}, \quad (C^T C)^{-1} = \frac{1}{\sum (t_i - \bar{t})^2} \begin{pmatrix} \frac{1}{n} \sum t_i^2 & -\bar{t} \\ -\bar{t} & 1 \end{pmatrix}$$

$$\Rightarrow \text{Var}_{\vartheta, \sigma^2}(\hat{\vartheta}_1) = \sigma^2 \frac{\frac{1}{n} \sum t_i^2}{\sum (t_i - \bar{t})^2}$$

$$\text{Var}_{\vartheta, \sigma^2}(\hat{\vartheta}_2) = \sigma^2 \frac{1}{\sum (t_i - \bar{t})^2} \quad (\text{vgl. 4.7})$$

$$\text{Cov}_{\vartheta, \sigma^2}(\hat{\vartheta}_1, \hat{\vartheta}_2) = \frac{-\sigma^2 \bar{t}}{\sum (t_i - \bar{t})^2} \quad (= 0, \text{ falls } \bar{t} = 0)$$

$$\begin{aligned} \text{Var}_{\vartheta, \sigma^2}(T) &= \sigma^2 a^T (C^T C)^{-1} a \\ &= \frac{\sigma^2}{\sum (t_i - \bar{t})^2} \left( \frac{1}{n} \sum t_i^2 - 2t^* \bar{t} + (t^*)^2 \right) \\ &= \sigma^2 \left( \frac{1}{n} + \frac{(t^* - \bar{t})^2}{\sum (t_i - \bar{t})^2} \right) \end{aligned}$$

### 14.5 Schätzung von $\sigma^2$

$$\hat{\sigma}^2 = \hat{\sigma}^2(X) = \frac{1}{n} \underbrace{\|X - C\hat{\vartheta}\|^2}_{=:\hat{\varepsilon}} = \frac{1}{n} \|\hat{\varepsilon}\|^2 = \frac{\hat{\varepsilon}^T \hat{\varepsilon}}{n}$$

( $\hat{\varepsilon}$  Residuenvektor)

Bemerkung:

$\hat{\sigma}^2$  ist asymptotisch erwartungstreu, aber nicht erwartungstreu für  $\sigma^2$ , da nach Aufgabe 44

$$\hat{S}^2 = \frac{n}{n-s} \hat{\sigma}^2 = \frac{1}{n-s} \|\hat{\varepsilon}\|^2$$

erwartungstreu für  $\sigma^2$  ist.

Ab jetzt stets klassisches lineares Modell ( $\varepsilon \sim \mathcal{N}_n(0, \sigma^2 I_n)$ )!

### 14.6 Satz

Im (klassischen) linearen Modell gilt:

- $(\hat{\vartheta}, \hat{\sigma}^2)$  ist ML-Schätzer für  $(\vartheta, \sigma^2)$
- $\hat{\vartheta} \sim \mathcal{N}_s(\vartheta, \sigma^2(C^T C)^{-1})$
- $\frac{n}{\sigma^2} \hat{\sigma}^2 \sim \chi_{n-s}^2$
- $\hat{\vartheta}$  und  $\hat{\sigma}^2$  sind stochastisch unabhängig

Beweis:

- $X \sim \mathcal{N}_n(C\vartheta, \sigma^2 I_n)$

$$\begin{aligned} \Rightarrow f(x, \vartheta, \sigma^2) &= \frac{1}{(\sigma\sqrt{2\pi})^n} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - (C\vartheta)_i)^2\right\} \\ &= \frac{1}{(\sigma^2 2\pi)^{\frac{n}{2}}} \exp\left\{-\frac{\|x - C\vartheta\|^2}{2\sigma^2}\right\} \\ &=: L_x(\vartheta, \sigma^2) \end{aligned}$$

Maximieren von  $L_x$  bezüglich  $\vartheta$  bei festem  $\sigma^2$  führt auf Minimierung von  $\|x - C\vartheta\|^2$ , Lösung ist  $\hat{\vartheta}$ .

$$\frac{\partial \log L_x(\hat{\vartheta}, \sigma^2)}{\partial \sigma^2} \stackrel{!}{=} 0$$

$$\hat{\sigma}^2 = \frac{1}{n} \|x - C\hat{\vartheta}\|^2$$

b) folgt aus Bemerkung 14.3 und Normalverteilungs-Annahme

c)

$$\begin{aligned} \varepsilon^T \varepsilon &= (X - C\vartheta)^T (X - C\vartheta) \\ &= (X - C\hat{\vartheta} + C(\hat{\vartheta} - \vartheta))^T (X - C\hat{\vartheta} + C(\hat{\vartheta} - \vartheta)) \\ &= (\hat{\varepsilon} + C(\hat{\vartheta} - \vartheta))^T (\hat{\varepsilon} + C(\hat{\vartheta} - \vartheta)) \\ \Rightarrow \frac{\varepsilon^T \varepsilon}{\sigma^2} &= \underbrace{\frac{\hat{\varepsilon}^T \hat{\varepsilon}}{\sigma^2}}_{\sim \chi_n^2} + \underbrace{(\hat{\vartheta} - \vartheta)^T \frac{C^T C}{\sigma^2} (\hat{\vartheta} - \vartheta)}_{\sim \chi_s^2 (1)} + 2 \underbrace{\hat{\varepsilon}^T C}_{(2)} \frac{(\hat{\vartheta} - \vartheta)}{\sigma^2} \end{aligned}$$

(1) nach Hilfssatz 13.6 und (b)

$$(2) = \varepsilon^T (I_n - P)^T C = \varepsilon^T (I_n - P) C = 0$$

Zu zeigen:  $\frac{\hat{\varepsilon}^T \hat{\varepsilon}}{\sigma^2} \sim \chi_{n-s}^2$

Die charakteristische Funktion von  $\chi_k^2$  ist

$$\varphi_{\chi_k^2}(t) = \int_{-\infty}^{\infty} e^{itx} f_k(x) dx = (1 - 2it)^{-\frac{k}{2}}$$

Unabhängigkeit von  $\hat{\vartheta}$  und  $\hat{\varepsilon}$  nach (d)

$$\Rightarrow (1 - 2it)^{-\frac{n}{2}} = \varphi_{\frac{\hat{\varepsilon}^T \hat{\varepsilon}}{\sigma^2}}(t) \cdot (1 - 2it)^{-\frac{s}{2}}$$

$$\Rightarrow \varphi_{\frac{\hat{\varepsilon}^T \hat{\varepsilon}}{\sigma^2}}(t) = (1 - 2it)^{-\frac{n-s}{2}}$$

Eindeutigkeitssatz für charakteristische Funktionen

$$\Rightarrow \frac{\hat{\varepsilon}^T \hat{\varepsilon}}{\sigma^2} = \frac{n}{\sigma^2} \hat{\sigma}^2 \sim \chi_{n-s}^2$$

$$d) \hat{\vartheta} = (C^T C)^{-1} C^T X = (C^T C)^{-1} C^T (C\vartheta + \varepsilon) = \vartheta + (C^T C)^{-1} C^T \varepsilon$$

$$\begin{aligned} \hat{\varepsilon} &= X - C\hat{\vartheta} \\ &= (I_n - C(C^T C)^{-1} C^T) X \\ &= (I_n - P)(C\vartheta + \varepsilon) \\ &= \underbrace{(I_n - P)C}_{C-C=0} \vartheta + (I_n - P)\varepsilon \\ &= (I_n - P)\varepsilon \\ ( &= Q\varepsilon \end{aligned}$$

$$\begin{aligned}
\Rightarrow \underbrace{\text{Cov}(\hat{\vartheta}, \hat{\varepsilon})}_{s \times n \text{ Matrix}} &= \text{Cov}(\vartheta + (C^T C)^{-1} C^T \varepsilon, (I_n - P)\varepsilon) \\
&= \text{Cov}((C^T C)^{-1} C^T \varepsilon, (I_n - P)\varepsilon) \\
&= \underbrace{(C^T C)^{-1} C^T}_{s \times n} \cdot \underbrace{\text{Cov}(\varepsilon, \varepsilon)}_{=\text{Var}(\varepsilon)=\sigma^2 I_n} \cdot \underbrace{(I_n - P)^T}_{n \times n} \\
&= \sigma^2 (C^T C)^{-1} \underbrace{((I_n - P)C)^T}_{=0} \\
&= 0
\end{aligned}$$

$$\begin{aligned}
\hat{\varepsilon} &= (I_n - P)\varepsilon \sim \mathcal{N}_n(0, (I_n - P)\sigma^2(I_n - P)^T) = \mathcal{N}_n(0, \sigma^2(I_n - P)) \\
\hat{\varepsilon}, \hat{\vartheta} &\text{ normalverteilt und unkorreliert} \Rightarrow \hat{\vartheta}, \hat{\varepsilon} \text{ unabhängig} \\
&\Rightarrow \hat{\vartheta}, \hat{\sigma}^2 \text{ stochastisch unabhängig.}
\end{aligned}$$

Bemerkung:

$\hat{\varepsilon} \sim \mathcal{N}_n(0, (I_n - P)\sigma^2)$ , d.h. die  $\hat{\varepsilon}_i$  haben nicht die gleiche Varianz.

## 14.7 Konfidenzbereiche für $\vartheta$

a) elliptischer Konfidenzbereich für  $\vartheta$ :

$$\begin{aligned}
\hat{\vartheta} - \vartheta &\sim \mathcal{N}(0, \sigma^2(C^T C)^{-1}) \\
\Rightarrow (\hat{\vartheta} - \vartheta)^T \frac{C^T C}{\sigma^2} (\hat{\vartheta} - \vartheta) &\sim \chi_s^2; \quad \frac{n\hat{\sigma}^2}{\sigma^2} \sim \chi_{n-s}^2
\end{aligned}$$

Beide Größen sind stochastisch unabhängig.

$$\begin{aligned}
&\Rightarrow \frac{\frac{1}{s}(\hat{\vartheta} - \vartheta)^T C^T C (\hat{\vartheta} - \vartheta)}{\frac{n}{n-s} \hat{\sigma}^2} \sim F_{s, n-s} \\
&\Rightarrow C_E := \{y \in \mathbb{R}^s : \frac{\frac{1}{s}(\hat{\vartheta} - y)^T C^T C (\hat{\vartheta} - y)}{\hat{s}^2} \leq F_{s, n-s, 1-\alpha}\}
\end{aligned}$$

erfüllt  $P_{\vartheta, \sigma^2}(C_E(X) \ni \vartheta) = 1 - \alpha \quad \forall \vartheta, \sigma^2$ , d.h.  $C_E$  ist ein (exakter)  $(1 - \alpha)$ -Konfidenzbereich für  $\vartheta$ .

b) Konfidenzintervall für  $\vartheta_j$ :

Sei  $(C^T C)^{-1} =: (b_{ij})_{s \times s}$ .  $\hat{\vartheta}_j \sim \mathcal{N}(\vartheta_j, b_{jj}\sigma^2)$

$$\stackrel{14.6(c),(d), 2.1}{\Rightarrow} \frac{\frac{\hat{\vartheta}_j - \vartheta_j}{\sigma \sqrt{b_{jj}}}}{\sqrt{\frac{n}{n-s} \frac{\hat{\sigma}^2}{\sigma^2}}} = \frac{\hat{\vartheta}_j - \vartheta_j}{\hat{s} \cdot \sqrt{b_{jj}}} \sim t_{n-s} (\sim \sqrt{F_{1, n-s}})$$

$$\Rightarrow P_{\vartheta, \sigma^2}(|\hat{\vartheta}_j - \vartheta_j| \leq t_{n-s, 1-\frac{\alpha}{2}} \cdot \hat{s} \sqrt{b_{jj}}) = 1 - \alpha$$

d.h.  $\hat{\vartheta}_j \pm t_{n-s, 1-\frac{\alpha}{2}} \cdot \hat{s} \sqrt{b_{jj}}$  ist zweiseitiges  $(1 - \alpha)$ -Konfidenzintervall für  $\vartheta_j$ .

- c) quaderförmiger Konfidenzbereich für  $\vartheta$  („Bonferroni-Methode“):  
Regel von den kleinen Ausnahmewahrscheinlichkeiten:

$$P(A_j) \geq 1 - \frac{\alpha}{s}, j = 1, \dots, s \Rightarrow P\left(\bigcap_{j=1}^s A_j\right) \geq 1 - \alpha$$

Denn:

$$P\left(\bigcap_{j=1}^s A_j\right) = 1 - P\left(\left(\bigcap_{j=1}^s A_j\right)^C\right) = 1 - P\left(\bigcup_{j=1}^s A_j^C\right) \geq 1 - \sum_{j=1}^s \underbrace{P(A_j^C)}_{\leq \frac{\alpha}{s}} \geq 1 - \alpha$$

Somit gilt für

$$C_Q(x) := \times_{j=1}^s [\hat{\vartheta}_j(x) - r(x), \hat{\vartheta}_j(x) + r(x)]$$

mit  $r(x) := t_{n-s, 1-\frac{\alpha}{2s}} \cdot \hat{s} \sqrt{b_{jj}}$ :

$$P_{\vartheta, \sigma^2}(C_Q(X) \ni \vartheta) \geq 1 - \alpha \quad \forall \vartheta, \sigma^2$$

d.h.  $C_Q$  ist quaderförmiger  $(1 - \alpha)$ -Konfidenzbereich für  $\vartheta$ .

Bemerkung:

$C_E$  hat kleineres Volumen wie  $C_Q$ , aber  $C_Q$  ist leichter zu interpretieren.

- d) Konfidenzintervall für  $a^T \vartheta$ :

$$\begin{aligned} a^T \hat{\vartheta} &\sim \mathcal{N}(a^T \vartheta, \sigma^2 \cdot a^T (C^T C)^{-1} a) \\ \Rightarrow \frac{a^T (\hat{\vartheta} - \vartheta)}{\hat{s} \sqrt{a^T (C^T C)^{-1} a}} &= \frac{\frac{a^T (\hat{\vartheta} - \vartheta)}{\sigma \sqrt{a^T (C^T C)^{-1} a}}}{\sqrt{\frac{\hat{s}^2}{\sigma^2}}} \sim t_{n-s} \end{aligned}$$

$\Rightarrow$  Mit  $r := t_{n-s, 1-\frac{\alpha}{2}} \cdot \hat{s} \sqrt{a^T (C^T C)^{-1} a}$  ist  $[a^T \hat{\vartheta} - r, a^T \hat{\vartheta} + r]$   $(1 - \alpha)$ -Konfidenzintervall für  $a^T \vartheta$ .

Beispiel:

einfache lineare Regression (vgl. Beispiel 14.4(b))

$$a = \begin{pmatrix} 1 \\ t^* \end{pmatrix}, \quad r = t_{n-2, 1-\frac{\alpha}{2}} \cdot \hat{s} \sqrt{\frac{1}{n} + \frac{(t^* - \bar{t})^2}{\sum_i (t_i - \bar{t})^2}}$$

$[\hat{\vartheta}_1 + \hat{\vartheta}_2 \cdot t^* - r, \hat{\vartheta}_1 + \hat{\vartheta}_2 \cdot t^* + r]$  ist  $(1 - \alpha)$ -Konfidenzintervall für  $a^T \vartheta = \vartheta_1 + \vartheta_2 \cdot t^*$ .

### 14.8 Tests von linearen Hypothesen im linearen Modell

$$X = C\vartheta + \varepsilon, \varepsilon \sim \mathcal{N}_n(0, \sigma^2 \cdot I_n)$$

Zu testen sei „lineare Hypothese“

$$H_0 : H\vartheta = h \quad \text{gegen} \quad H_1 : H\vartheta \neq h$$

Dabei:  $H$   $r \times s$ -Matrix,  $\text{Rang}(H) = r$  (insbesondere  $r \leq s$ ),  $h \in \mathbb{R}^r$  gegeben

$$H_0 \hat{=} \Theta_0 := \{(\vartheta, \sigma^2) \in \underbrace{\mathbb{R}^s \times \mathbb{R}_{>0}}_{=\Theta} : H\vartheta = h\}, \quad H_1 \hat{=} \Theta \setminus \Theta_0$$

### 14.9 Beispiele

a)  $X_j = \vartheta_1 + \vartheta_2 \cdot t_j + \varepsilon_j, j = 1, \dots, n$  (einfache lineare Regression)

$$H_0 : \vartheta_2 = 0 \quad \text{gegen} \quad H_1 : \vartheta_2 \neq 0$$

„Lineare Hypothese“:  $H = (0, 1), h = 0$  ( $s = 2, r = 1$ )

$$H_0 : H \cdot \begin{pmatrix} \vartheta_1 \\ \vartheta_2 \end{pmatrix} = 0$$

Möglicher Test: Verallgemeinerter Likelihood-Quotienten-Test  
Testgröße  $\Lambda_n$  bzw.  $\log \Lambda_n$ .

$$\Lambda_n := \frac{\sup_{(\vartheta, \sigma^2) \in \Theta_0} f(x, \vartheta, \sigma^2)}{\sup_{\Theta} f(x, \vartheta, \sigma^2)}$$

Unter  $H_0$ :  $X_j = \vartheta_1 + \varepsilon_j, X_j \sim \mathcal{N}(\vartheta_1, \sigma^2)$ , ML-Schätzer für  $\vartheta_1$ :  $\bar{X}_n$

Ohne Restriktion: ML-Schätzer = KQ-Schätzer<sup>41</sup> =  $\hat{\vartheta}$  (Satz 14.6(a))

Als Schätzer für  $\sigma^2$  wird aber üblicherweise in beiden Fällen der Schätzer  $\hat{\sigma}^2$  aus Obermodell verwendet!

Dann<sup>42</sup>:

$$\log \Lambda_n = -\frac{1}{2\hat{\sigma}^2} \left[ \underbrace{\sum_{i=1}^n (X_i - \bar{X}_n)^2}_{=: SS_0} - \underbrace{\sum_{i=1}^n (X_i - (\hat{\vartheta}_1 + \hat{\vartheta}_2 t_i))^2}_{=: SS_1 (=n\hat{\sigma}^2)} \right]$$

Als Testgröße wird

$$T := \frac{SS_0 - SS_1}{\frac{SS_1}{n-2}}$$

verwendet. Es gilt:

<sup>41</sup>Kleinste-Quadrate-Schätzer

<sup>42</sup>SS: sum of squares

- (i)  $\frac{SS_1}{\sigma^2} \sim \chi_{n-2}^2$  (nach 14.6(c))  
(ii)  $\frac{SS_0}{\sigma^2} \sim \chi_{n-1}^2$  unter  $H_0$  (nach 2.2)  
(iii)  $SS_0 - SS_1$  und  $SS_1$  stochastisch unabhängig (ohne Beweis)

$$\underbrace{\frac{SS_0}{\sigma^2}}_{\stackrel{H_0}{\sim} \chi_{n-1}^2} = \frac{SS_0 - SS_1}{\sigma^2} + \underbrace{\frac{SS_1}{\sigma^2}}_{\sim \chi_{n-2}^2}$$

$$\Rightarrow \frac{SS_0 - SS_1}{\sigma^2} \sim \chi_{n-1-(n-2)}^2 = \chi_1^2 \text{ unter } H_0 \text{ (vgl. Beweis von 14.6(c))}$$

Damit  $T \sim F_{1,n-2}$  unter  $H_0$ .

- b)  $X_{i,j} = \vartheta_j + \varepsilon_{i,j}$  ( $i = 1, \dots, k, j = 1, \dots, n_i$ ) (einfache Varianzanalyse<sup>43</sup>)

$$H_0 : \vartheta_1 = \dots = \vartheta_k$$

(„kein Effekt des zu untersuchenden Faktors“)

$$\underbrace{\begin{pmatrix} 1 & 0 & -1 \\ & \ddots & \vdots \\ 0 & 1 & -1 \end{pmatrix}}_{=: H \in \mathbb{R}^{k-1 \times k}} \cdot \begin{pmatrix} \vartheta_1 \\ \vdots \\ \vartheta_k \end{pmatrix} = \underbrace{\begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}}_{=: h \in \mathbb{R}^{k-1}}$$

$$\text{Rang}(H) = k - 1 (= r)$$

Testgröße: (vgl. Aufgabe 45)

$$\text{Sei } \bar{X}_{i+} = \frac{1}{n_i} \sum_{j=1}^{n_i} X_{i,j}, \bar{X}_{++} = \frac{1}{n} \sum_{i,j} X_{i,j}, n = \sum_{i=1}^k n_i,$$

$$\text{SQZ} = \sum_{i=1}^k n_i (\bar{X}_{i+} - \bar{X}_{++})^2, \text{SQI} = \sum_{i,j} (X_{i,j} - \bar{X}_{i+})^2$$

$$\sum_{i,j} (X_{i,j} - \bar{X}_{++})^2 = \text{SQI} + \text{SQZ}$$

$$T := \frac{\text{SQZ}}{\frac{\text{SQI}}{n-k}} \sim F_{k-1, n-k} \text{ unter } H_0$$

<sup>43</sup>  $k \hat{=} s$

### 14.10 Die Testgröße bei allgemeinen linearen Hypothesen

$$\hat{\vartheta} \sim \mathcal{N}_j(\vartheta, \sigma^2(C^T C)^{-1})$$

$$\Rightarrow H\hat{\vartheta} \sim \mathcal{N}_r(H\vartheta, \sigma^2 \underbrace{H(C^T C)^{-1}H^T}_{=:B})$$

$$\frac{\frac{1}{r} \cdot \frac{1}{\sigma^2} (H\hat{\vartheta} - H\vartheta)^T B^{-1} (H\hat{\vartheta} - H\vartheta)}{\frac{\hat{s}^2}{\sigma^2}} \sim \frac{\frac{\chi_r^2}{r}}{\frac{\chi_{n-s}^2}{n-s}} \sim F_{r, n-s}$$

(Zähler und Nenner sind stochastisch unabhängig.)

Sei

$$T := \frac{\frac{1}{r} (H\hat{\vartheta} - h)^T (H(C^T C)^{-1}H^T)^{-1} (H\hat{\vartheta} - h)}{\hat{s}^2} \sim F_{r, n-s} \text{ unter } H_0$$

Der sogenannte **F-Test** im linearen Modell besitzt die Gestalt:

$H_0$  ablehnen, falls  $T \geq F_{r, n-s, 1-\alpha}$ .

Kein Widerspruch zu  $H_0$ , falls  $T < F_{r, n-s, 1-\alpha}$ .

Bemerkung:

Für die Beispiele aus 14.9 stimmt die obige Testgröße mit den Testgrößen aus 14.9(a) bzw. (b) überein.