# 1 Statistische Modelle

Der Stichprobenraum $\mathcal{X}$: Die möglichen Beobachtungsergebnisse bilden eine Menge.

**Beispiel 1.** $\mathcal{X}=\{0,\ldots,N\}, \mathcal{X}=\mathbb{N}, \mathcal{X}=\mathbb{R}^d$

Wieso $\mathcal{X}$ und nicht $\Omega$ ? $\mathcal{X}$ ist das Bild eines Zufallsexperiments $\mathcal{X}\colon\Omega\to\mathcal{X}$.

Die Wahrscheinlichkeitsverteilung auf $\mathcal{X}$ ist unbekannt, daher betrachten wir eine Familie von W.-verteilungen.

**Definition 2.** *Ein statistisches Modell ist ein Tripel $\mathcal{M}=(\mathcal{X},\mathcal{F},(\mathbb{P}_\vartheta\colon\vartheta\in\theta))$. Wobei*

- *$\mathcal{X}$ : Stichprobenraum,*
- *$\mathcal{F}$ : $\sigma$-Algebra auf $\mathcal{X}$,*
- *$(\mathbb{P}_\vartheta\colon\vartheta\in\theta)$ : Familie von W-Maßen auf $\mathcal{X}$.*

**Bemerkung 3.** Wenn man $(\mathbb{P}_\vartheta\colon\vartheta\in\theta)$ schlecht wählt, wird das stat. Verfahren unsinnig!

Die Grundaufgabe des Statistikers besteht in der Wahl des geeigneten Modells !

**Definition 4.** *Ein statistisches Modell $\mathcal{M}$ heißt parametrisch falls $\theta\subseteq\mathbb{R}^d$ für ein $d\in\mathbb{N}$.*

*$\mathcal{M}$ heißt diskret, falls $\mathcal{X}$ diskret mit $\mathcal{F}=\mathcal{P}(\mathcal{X})$. Dann hat $\mathbb{P}_\vartheta$ eine Zähldichte: $\zeta_\vartheta\colon x\mapsto\mathbb{P}_\vartheta[\{x\}]$.*

*$\mathcal{M}$ heißt absolut-stetig, falls $\mathcal{X}\subset\mathbb{R}^d$ mit $\mathcal{F}=\mathcal{B}(\mathcal{X})$ und $\mathbb{P}_\vartheta$ eine Dichtefunktion $\zeta_\vartheta$ hat.*

*$\mathcal{M}$ heißt Standartmodell, falls es diskret oder absolut-stetig ist.*

*Sei $(E,\mathcal{E},\mathbb{Q}_\vartheta\colon\vartheta\in\theta)$ ein stat. Modell und $n\geq 2$.*

$$(\mathcal{X},\mathcal{F},(\mathbb{P}_\vartheta\colon\vartheta\in\theta)):=\left(E^n,\mathcal{E}^{\otimes n},{\mathbb{Q}_\vartheta}^{\otimes n}\colon\vartheta\in\theta\right)$$

*ist das zugehörige n-fache Produktmodell.*

*$X_k\colon\mathcal{X}\to E$ ist die k-te Koordinate und beschreit den Ausgang des k-ten Experiments. Insbesondere sind $X_1,\ldots,X_n$ u.i.v. (unabhängig und identisch verteilt) bzgl. $\mathbb{P}_\vartheta$ mit Verteilung $\mathbb{Q}_\vartheta$.*

# 2 Schätzer

Sei $(\mathcal{X},\mathcal{F},(\mathbb{P}_\vartheta\colon\vartheta\in\theta))$ stat. Modell und $(\Sigma,\zeta)$ ein Ereignisraum.

Eine bel. Zufallsvariable

$$\delta\colon(\mathcal{X},\mathcal{F})\to(\Sigma,\zeta)$$

heißt Statistik.

Sei $\tau\colon\theta\to\Sigma$ eine Abbildung. $\tau(\vartheta)\in\Sigma$ heißt Kenngröße. Eine Statistik $T\colon\mathcal{X}\to\Sigma$ heißt Schätzer für τ.

Ende Vorlesung 1

---

**Bemerkung 5.**

i. Statistik = ZV (im mathematischen Sinne), aber Zufallsvariable = unvorhersehbares Ereignis hervorgerufen durch Zufall. Eine Statistik = Vom Statistiker bestimmte Abbildung.

ii. Schätzer vs. Statistik: Ein Schätzer T ist eine Statistik, die speziell für die Schätzung von $\tau$ zugeschnitten ist.

iii. Was hat T mit $\tau$ zu tun ? Es gibt nicht nur einen Schätzer $T$ für $\tau(\vartheta)$. Daher ist es nicht formalisiert um nicht zu restrektiv zu sein.

iv. Man spricht auch von Punktschätzern um von Bereichsschätzern abzugrenzen. (Kapitel: Konfidenzbereiche)

**Beispiel 6.** $\mathcal{X}=\{0,1\}^n, \mathbb{P}_\vartheta = \mathrm{Ber}_\vartheta^{\otimes n}$ mit $\vartheta \in [0,1]$ unbekannt.

$\mathrm{Ber}_\vartheta(1) = \vartheta = 1 - \mathrm{Ber}_\vartheta(0)$.

Gesucht: $\tau(\vartheta) = \vartheta$.

Sei $X = \{X_1, \ldots, X_n\}$ die Stichprobe.

$$\Rightarrow T(X) = \frac{1}{n}\sum_{i=1}^{n} X_i$$

ist ein Schätzer für $\vartheta$. Ein anderer Schätzer ist $S(X) = \frac{1}{2}$.

Aus dem Gesetz der großen Zahlen

$$\lim_{n\to\infty} T = \vartheta, \mathbb{P} - f.s.$$

aber

$$\lim_{n\to\infty} S = \frac{1}{2}.$$

Außer im „Glücksfall" $\vartheta = \frac{1}{2}$ ist T der „bessere" Schätzer als S.

- Was sind Qualitätskriterien ?
- Wie Schätzer finden ?

## 2.1 Maximum-Likelihood

- Die Idee ist einen Schätzer T zu wählen, s.d. die Dichtefunktion so groß wie möglich ist(D.h. wir sind im Standardfall). Methode zur Bestimmung eines Schätzer: andere Methode: Momentenmethode

**Definition 7.** *Sei* $(\mathcal{X}, \mathcal{F}, (\mathbb{P}_\vartheta : \vartheta \in \theta))$ ein stat. Standartmodell. *Die Likelihoodfunktion ist*

$$\rho: \mathcal{X} \times \theta \to [0, \infty) \text{ mit}$$

$$\rho(x, \vartheta) = \rho_\vartheta(x),$$

*wobei* $\rho_\vartheta$ *die Dichtefunktion von* $\mathbb{P}_\vartheta$ *ist.*

*Die Likelihood-Funktion zum Beobachtungswert* $x \in X$ *ist*

$$\rho_x := \rho(x, \cdot): \theta \to [0, \infty]$$

$$\vartheta \mapsto \rho(x, \vartheta).$$

**Definition 8.** *Ein Schätzer $T: \mathcal{X} \to \theta$ für $\vartheta$ heißt Maximum-Likelihood-Schätzer (M-L-Schätzer) wenn $\rho(x, T(x)) = \max_{\vartheta \in \theta} \rho(x, \vartheta)$ für jedes $x \in \mathcal{X}$.*

*$\Rightarrow T(x)$ ist eine Maximalstelle der Funktion $\rho_x$ auf $\theta$.*

**Beispiel 9.** (Schätzung von Erfolgswahrscheinlichkeit)

Sei $\vartheta$ der Wirkungsgrad eines Medikaments.

$X_1, \ldots, X_n$ Stichprobe, $X_k \in \{0, 1\}$ ($1 \triangleq$gesund)

Sei $x \in \{0, \ldots, n\}$ Zahl der geheilten Personen.

Modell: Binomialmodell: $\mathcal{X} = \{0, \ldots, n\}$, $\mathcal{F} = \mathcal{P}(\{0, \ldots, n\}), \mathbb{P}_\vartheta = \mathrm{Bin}_{n,\vartheta}, \vartheta \in [0, 1]$

d.h. $\rho_\vartheta(x) = \binom{n}{x} \vartheta^x (1 - \vartheta)^{n-x}$

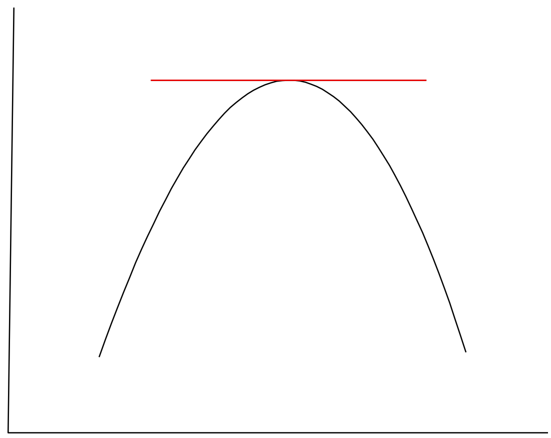
Was ist der M-L-Schätzer ?

Da $y \mapsto \ln y$ monoton wachsend, reicht es das Maximum von $\ln \rho_x(\vartheta)$ zu bestimmen.

$\Rightarrow: \frac{d}{d\vartheta} \ln \rho_x(\vartheta) = \frac{d}{d\vartheta} (\mathrm{x}\ln\vartheta + (n - x)\ln(1 - \vartheta)) = \frac{x}{\vartheta} - \frac{n - x}{1 - \vartheta} = \frac{x - \vartheta x - n\vartheta + \vartheta x}{\vartheta(1 - \vartheta)} = \frac{x - n\vartheta}{\vartheta(1 - \vartheta)} \overset{!}{=} 0$

Also $x = n\vartheta$.

Maximum ? Ja weil für $\vartheta \leqslant \frac{x}{n}$ ist $\rho_x$ wachsend, für $\vartheta \geqslant \frac{x}{n}$ ist $\rho_x$ fallend.



„unimodal“

$\Rightarrow T(x) = \frac{x}{n}$ ist (der) ML-Schätzer für $\vartheta$ im Binomialmodell

**Beispiel 10.** (Physikalische Messungen)

In jeder physikalischen Messung gibt es Messfehler.

**Annahme:**

Messungen sind u.i.v. ZV $X_1, \ldots, X_n$, n Zahl der Messungen mit $X_i \sim \mathcal{N}(\underbrace{m}, \underbrace{\sigma^2})$, wobei m,$\sigma$ unbekannt.

$\Rightarrow M = (\mathbb{R}^d, \mathcal{B}^{\otimes n}, \mathcal{N}(m, \sigma^2), m \in \mathbb{R}, \sigma > 0)$

d.h. $\rho_\vartheta(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x - m)^2}{2\sigma^2}\right)$

Der M-L-Schätzer für $(m, \sigma^2)$ ist

$$T(x) = \left(\frac{1}{n}\sum_{i=1}^{n} x_i, \frac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x}_n)^2\right)$$

Beweis: Übung

Weitere Beispiele: Blatt 1+ Präsenzblatt (kont. Version German Tank Problem)

## 2.2 Erwartungstreue und quadratische Fehler

Ein erstes elementares Qualitätskriterium.

**Definition 11.** *Sei $(\mathcal{X}, \mathcal{F}, (\mathbb{P}_\vartheta : \vartheta \in \theta))$ ein stat. Modell und $\tau : \theta \to \mathbb{R}$ eine relle Kenngröße.*

*Ein Schätzer $T : \mathcal{X} \to \mathbb{R}$ für $\tau$ heißt erwartungstreu*

*wenn*

$$\mathbb{E}_\vartheta[T] = \tau \forall \vartheta \in \theta$$

*sonst, ist*

$$\mathbb{B}_\vartheta[T] := \mathbb{E}_\vartheta[T] - \tau(\vartheta).$$

*der Bias oder systematischer Fehler von $T$.*

***M-L-Schätzer sind nicht unbedingt erwartungstreu!!!!***

Der M-L-Schätzer für die Varianz im Gauß Modell(Bsp. Physikalische Messungen) ist nicht erwartungstreu.

**Satz 12.** *(Schätzung von Erwartungswert und Varianz bei rellen Produktmodellen)*

*Sei $n \geqslant 2$ und $\mathcal{M} = (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ ...*

*Sei $m(\vartheta) := \mathbb{E}_\vartheta[X]$ und $\nu(\vartheta) := \mathrm{Var}_\vartheta[X]$ für jedes $\vartheta \in \theta$ definiert.*

*Der Stichprobenmittelwert*

$$M := \frac{1}{n}\sum_{k=1}^{n} X_k$$

*und die korrigierte Stichprobenvarianz*

$$V^* := \frac{1}{n-1}\sum_{k=1}^{n} (X_k - M)^2$$

*sind erwartungstreu für $(m, v)$.*

***Beweis.*** *Sei $\vartheta \in \theta$ fest.*

*1)* $\mathbb{E}_\vartheta[M] \underset{\text{lin}}{=\!=} \frac{1}{n}\sum_{k=1}^{n} \mathbb{E}_\vartheta[X_k] = \frac{1}{n}\mathrm{nm}(\vartheta) = m(\vartheta).$

*2) Sei $V = \frac{n-1}{n} V^*$ Stichprobenvarianz.*

$$\mathbb{E}_\vartheta[V] \xlongequal[\text{lin.}]{} \frac{1}{n}\sum_{k=1}^{n} \mathbb{E}[X_k - M]^2 \xlongequal[\mathbb{E}_\vartheta[X_k - M]=0]{} \frac{1}{n}\sum_{k=1}^{n} \mathrm{Var}[X_k - M]$$

$$\xlongequal[X_k i.i.d.]{} \mathrm{Var}_\vartheta\left[X_1 - \frac{1}{n}\sum_{k=1}^{n} X_k\right] = \mathrm{Var}_\vartheta\left[X_1 \cdot \frac{n-1}{n} - \frac{1}{n}\sum_{k=2}^{n} X_k\right]$$

$$\xlongequal[X_k i.i.d.]{} \left(\frac{n-1}{n}\right)^2 \mathrm{Var}_\vartheta[X_1] + \left(\frac{1}{n}\right)^2 \mathrm{Var}_\vartheta[X_1] = \left(\frac{n-1}{n}\right)^2 \nu(\vartheta) + \frac{n-1}{n^2}\nu(\vartheta) = \frac{n-1}{n}\nu(\vartheta)$$

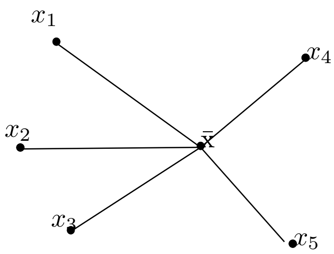
$\Rightarrow \mathbb{E}_\vartheta[V^*] = \nu(\vartheta)$.

$\square$

**Bemerkung 13.** 1) Für große n sind $\frac{1}{n}$ und $\frac{1}{n-1}$ fast gleich. $\Rightarrow$ V ist asymptotisch erwartungstreu.

2) $\mathbb{E}_\vartheta[V] = \frac{n-1}{n}\nu(\vartheta) < \nu(\vartheta), \nu(\vartheta) > 0$.

Der Schätzer V unterschätzt systematisch die Varianz.




Das heißt da

$$\sum_{i=1}^{5} (x_i - \bar{x}) = 0$$

ist $x_1 - \bar{x}$ ist durch die anderen diff. schon bestimmt. Daher Normalisieren mit $\frac{1}{n-1}$.

3) Wenn der Erwartungswert bekannt $\mathbb{E}_\vartheta[X] = \mu$, dann ist V erwartungstreuer Schätzer für die Varianz!!

Ende Vorlesung 2

**Erwartungstreue ist wünschenswert, aber nicht immer „besser".**

**Beispiel 14.** (Fortsetzung Binomialmodell)

$\mathcal{X}=\{0,\ldots,n\}$,$\theta=[0,1]$, $\mathbb{P}_\vartheta=\mathrm{Bin}_{n,\vartheta}$.

T(x)=$\frac{\mathrm{x}}{\mathrm{n}}$ ist Ml-Schätzer für $\theta$.

$$\mathbb{E}_\vartheta[T]=\frac{1}{n}\mathbb{E}_\vartheta[X]=\vartheta\Rightarrow\text{Erwartungstreu}$$

Anderer Schätzer

$S(x)=\frac{x+1}{n+2}$ nicht erwartungstreu

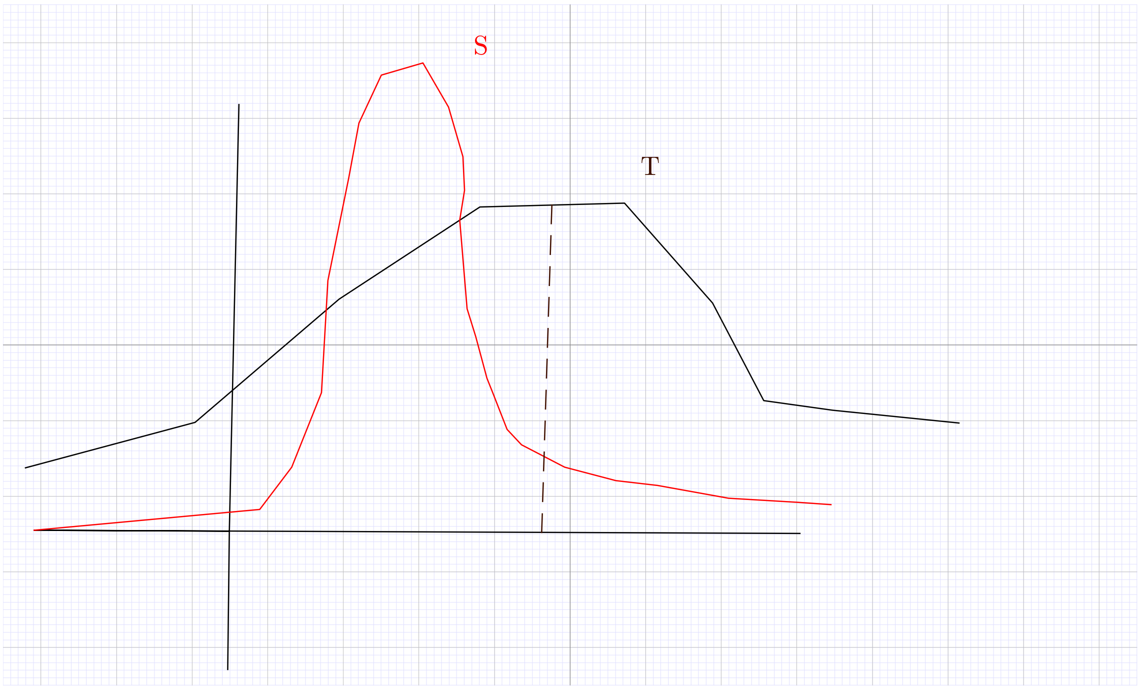
$$\mathbb{B}_\vartheta[S]=\frac{n\vartheta+1}{n+2}-\vartheta=\frac{1-2\vartheta}{n+2}>0$$

Aber was ist mit der mittleren quadraitschen Abweichung?

**Definition 15.** *Der mittlere quadratische Fehler eines Schätzers $T$ für $\tau$ ist*

$$\mathbb{F}_\vartheta[T]:=\mathbb{E}[(T-\tau(\vartheta))^2]=\mathrm{Var}_\vartheta[T]+\mathbb{B}_\vartheta[T]^2$$

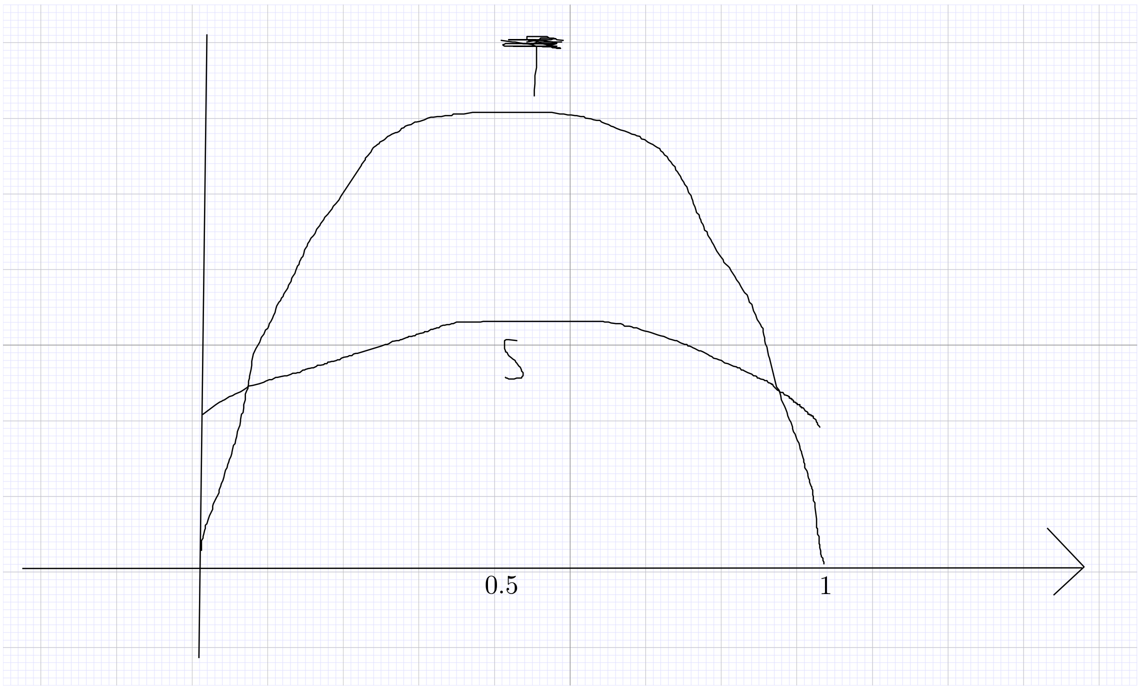
*Wir wollen beide Terme gleichzeitig minimieren.*




$\mathbb{F}_\vartheta[T]=\frac{1}{n^2}\mathrm{Var}_\vartheta[X]=\frac{1}{n^2}n\vartheta(1-\vartheta)=\frac{\vartheta(1-\vartheta)}{n}$

$\mathrm{Var}_\vartheta[S]=\frac{1}{(n+2)^2}\mathrm{Var}_\vartheta[X]$

$\Rightarrow\mathbb{F}_\vartheta[S]=\frac{n\vartheta(1-\vartheta)+(1-2\vartheta)^2}{(n+2)^2}$




Für Zentrale Werte von $\vartheta$: S ist besser als T.

Es gilt für $(\left|\vartheta-\frac{1}{2}\right| \leq \frac{1}{\sqrt{8}} \approx 0.35)$.

Erwartungstreue ist also nicht alles, bleibt aber wichtig (Siehe Kapitel „Beste Schätzer“)

## 2.3 Konsistenz von Schätzern

Ein weiteres Qualitätskriterium ist die Konsistenz.

- Sei $\mathcal{M}=(\mathcal{X},\mathcal{F},(\mathbb{P}_\vartheta))$ ein stat. Modell und $\tau: \theta \to \mathbb{R}$ eine relle Kenngröße.
- Wiederholung der Messung: Sei $(X_n)_{n\geqslant 1}$ eine Folge von ZV auf $(\mathcal{X},\mathcal{F})$

  $X_n$ ist n-te Messung mit Werten in $(E,\mathcal{E})$ $(z.B. \mathcal{X}=E^n)$
- Sei für $n \geqslant 1$ $T_n: \mathcal{X} \to \mathbb{R}$ ein Schätzer für τ

**Definition 16.** *Die Schätzfolge $(T_n)_{n\geqslant 1}$ für $\tau$ heißt konsistent, wenn* $\forall \epsilon > 0, \vartheta \in \theta$

$$\lim_{n\to\infty} \mathbb{P}_\vartheta[|T_n-\tau(\vartheta)| \leq \epsilon] = 1$$

*oder:*

$$\forall \epsilon, \vartheta \in \theta\colon \lim_{n\to\infty} \mathbb{P}_\vartheta[|T_n-\tau(\theta)| \geq \epsilon] = 0$$

*„Konvergenz im Maß (Stochastische Konvergenz)“*

Im folgenden:

Standartfall mit unabhängigen Beobachtungen

$\Rightarrow(\mathcal{X},\mathcal{F},(\mathbb{P}_\vartheta)) = (E^{\mathbb{N}},\mathcal{E}^{\otimes\mathbb{N}},(\mathbb{Q}_\vartheta^{\otimes\mathbb{N}}))$

**Satz 17.** *Im unendlichen Produktmodell seien:*

$M_n = \frac{1}{n}\sum_{k=1}^n X_k, V_n^\star = \frac{1}{n-1}\sum_{k=1}^n (X_k - M_n)^2$

*die Erwartungstreuen Schätzer für m bzw. v.*

*Dann sind die Folgen $(M_n)_{n\geqslant 1}, (V_n^*)_{n\geqslant 1}$ konsistent.*

***Beweis.*** *1.) Nach dem (schwachen) Gesetz der großen Zahlen*

$\frac{1}{n}\sum_{k=1}^n X_k \xrightarrow{\mathbb{P}_\vartheta} \mathbb{E}_\vartheta[X_1] = m(\vartheta)$.

*2.) Sei* $\tilde{V}_n = \frac{1}{n}\sum_{k=1}^n (X_k - m(\vartheta))^2, V_n := \frac{n-1}{n} V_n^*$

$\Rightarrow V_n = \tilde{V} - (M_n - m(\vartheta))^2$ *(Verschiebungsformel / Verschiebungssatz)*

$\tilde{V} \xrightarrow{\mathbb{P}_\vartheta} v(\vartheta)$ *und* $(M_n - m(\vartheta))^2 \xrightarrow{\mathbb{P}_\vartheta} 0$ *(beides Nach g.G.Z.)*

$V_n \xrightarrow{\mathbb{P}_\vartheta} v(\vartheta)$ *und damit* $V_n^* = \frac{n}{n-1} V_k \xrightarrow{\mathbb{P}_\vartheta} v(\vartheta)$. □

Auch M-L-Schätzer sind konsistent:

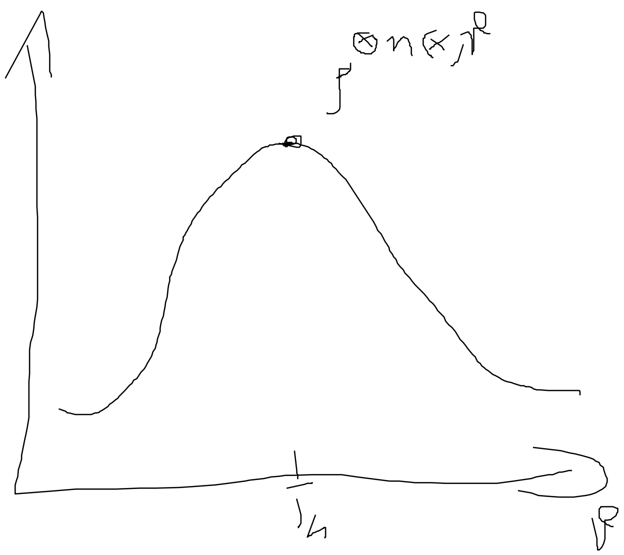
**Satz 18.** *(Konsistenz von M-L-Schätzern)*

*Sei* $(E, \mathcal{E}, \mathbb{Q}_\vartheta)$ *eine einparametriges Standartmodell (d.h.* $\theta \subseteq \mathbb{R}$*) ,mit Likelihood-Funktion* $\rho$.

*Es gelte:*

- $\theta$ *ist offenes Intervall in* $\mathbb{R}$ *und für* $\vartheta \neq \vartheta'$ *ist* $\mathbb{Q}_\vartheta \neq \mathbb{Q}_{\vartheta'}$.
- $\forall n \geqslant 1\ \forall x \in E^n$ *ist*

  $$\rho^{\otimes n}(x, \vartheta) = \prod_{k=1}^n \rho(x_k, \vartheta)$$

  *unimodal, d.h.* $\exists$ *ML-Schätzer* $T_n: E_n \to \mathbb{R}$ *s.d* $\vartheta \mapsto \rho^{\otimes n}(x, \vartheta)$ *ist wachsend für* $\vartheta < T_n(x)$ *und fallend für* $\vartheta > T_n(x)$.
- 

.

Dann ist die Schätzfolge konsistent für $\vartheta$.

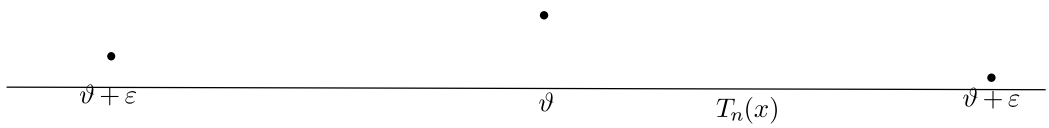
**Beweis.** (grob)

Wir wollen zeigen, dass $\forall \epsilon > 0, \vartheta \in \theta$

$$\mathbb{P}_{\boldsymbol{\vartheta}}[\vartheta - \epsilon \leqslant T_n \leqslant \vartheta + \epsilon] \xrightarrow{n \to \infty} 1$$

Sei $\vartheta \in \theta$ und $\varepsilon > 0$ $(\vartheta \pm \varepsilon \in \theta)$

$\{x : \vartheta - \epsilon \leqslant T_n(x) \leqslant \vartheta + \varepsilon\} \supseteq \{x : \rho_{\vartheta - \varepsilon}^{\otimes n}(x) < \rho_{\vartheta}^{\otimes n}(x), \rho_{\vartheta + \varepsilon}^{\otimes n}(x) < \rho_{\vartheta}^{\otimes n}(x)\}$




$\supseteq \left\{x : \log\left(\frac{\rho_{\vartheta}^{\otimes n}(x)}{\rho_{\vartheta+\varepsilon}^{\otimes n}(x)}\right) > 0, \log\left(\frac{\rho_{\vartheta}^{\otimes n}(x)}{\rho_{\vartheta-\varepsilon}^{\otimes n}(x)}\right) > 0\right\}$

„+“-Fall Sei $f(x) = \frac{\rho_{\vartheta}}{\rho_{\vartheta+\varepsilon}}(x)$ und wir nehmen an, dass $\mathbb{E}_{\vartheta}[\log f] < \infty$.

Dann gilt nach dem G.d.g.Z. ($\mathcal{L}^1$ −Version: $X_i$ p.w. u.i.v. und in $\mathcal{L}^1$, dann $\frac{1}{n}\sum_{i=1}^{n} X_i \to \mathbb{E}[X_1]$)

$$\frac{1}{n}\log\frac{\rho_{\vartheta}^{\otimes n}}{\rho_{\vartheta+\varepsilon}^{\otimes n}}(X) = \frac{1}{n}\sum_{i=1}^{n} \log(X_i) \xrightarrow{\mathbb{P}_{\vartheta}} \mathbb{E}_{\vartheta}[\log f]$$

$\mathbb{E}_{\vartheta}[\log f] = \int \log f \rho_{\vartheta}(x) \mathrm{d}x = \int \log \frac{\rho_{\vartheta}}{\rho_{\vartheta+\varepsilon}}(x) \rho_{\vartheta}(x) \mathrm{d}x =: H(\mathbb{Q}_{\vartheta}; \mathbb{Q}_{\vartheta+\varepsilon})$ (reltative Entropie)

Es gilt $H(\mathbb{Q}_{\vartheta}; \mathbb{Q}_{\vartheta}) > 0$, da wir angenommen haben, dass $\mathbb{Q}_{\vartheta} \neq \mathbb{Q}_{\vartheta'}$ für $\vartheta \neq \vartheta'$. (Beweis Blatt 2)

$\Rightarrow \exists \delta > 0$ s.d.

$$\mathbb{P}_{\vartheta}\left[\frac{1}{n}\log\frac{\rho_{\vartheta}^{\otimes n}}{\rho_{\vartheta+\varepsilon}^{\otimes n}} > \delta\right] \xrightarrow{n \to \infty} 1$$

„-“ Fall genau so…

$\Rightarrow \exists \delta > 0$ s.-d.

$$\mathbb{P}_{\vartheta}[\underbrace{\frac{1}{n}\log\frac{\rho_{\vartheta}^{\otimes n}}{\rho_{\vartheta\pm\varepsilon}^{\otimes n}} > \delta}_{\subseteq \left\{x : \log\left(\frac{\rho_{\vartheta}^{\otimes n}(x)}{\rho_{\vartheta\pm\varepsilon}^{\otimes n}(x)}\right) > 0\right\} \subseteq \{x : \vartheta - \epsilon \leqslant T_n(x) \leqslant \vartheta + \varepsilon\}}] \xrightarrow{n \to \infty} 1$$

Der Fall $\mathbb{E}_\vartheta[\log f] = \infty$ siehe Georgii. □

## 2.4 Beste Schätzer

Wir konzentrieren uns jetzt auf Klasse von Schätzern, die

- erwartungstreu
- am wenigsten streuen (Varianz ist minimal)

**Definition 19.** *Sei* $(\mathcal{X}, \mathcal{F}, (\mathbb{P}_\vartheta : \vartheta \in \theta))$ ein stat. Modell. *Ein erwartungstreuer Schätzer $T$ für eine reelle Kenngröße $\tau(\vartheta)$ heißt varianzminimierend/bester Schätzer, falls für jeden weiteren erwartungstreuen Schätzer $S$*

$$\mathrm{Var}_\vartheta[T] \leqslant \mathrm{Var}_\vartheta[S] \forall \vartheta \in \theta.$$

**Definition 20.** *(Regulär) Ein einparametrisches Standartmodell $(\mathcal{X},\mathcal{F},(\mathbb{P}_\vartheta : \vartheta \in \theta))$ heißt regulär, falls*

*i. $\theta$ ist ein offenes Intervall in $\mathbb{R}$.*

*ii. Die Likelihoodfunktion $\rho$ ist auf $\mathcal{X} \times \theta$ strikt positiv und nach $\vartheta$ stetig differenzierbar.*

*iii. Für jedes $\vartheta \in \theta$ ex. die Varianz:*

$$I(\vartheta) := \mathrm{Var}\left[\frac{d}{d\vartheta} \underbrace{\log \rho(x, \vartheta)}_{\text{diffbar}}\right]$$

*und ist nicht 0. Außerdem gilt die Vertauschungsregel:*

$$\int \frac{d}{d\vartheta} \rho(x, \vartheta) \mathrm{dx} = \frac{d}{d\vartheta} \int \rho(x, \vartheta) \mathrm{dx}.$$

Ende Vorlesung 3

---

**Bemerkung 21.** i. $I(\vartheta)$ heißt auch Fisher-Information des Modells und $U_\vartheta(x) := \frac{d}{d\vartheta} \underbrace{\log \rho(x, \vartheta)}_{\text{diffbar}}$ die Score Funktion. $I(\vartheta) = \mathrm{Var}_\vartheta[U_\vartheta]$

ii. $\mathbb{E}[U_\vartheta] = 0$, denn

$$\mathbb{E}[U_\vartheta] = \int_{\mathcal{X}} \frac{d}{d\vartheta} \log \rho(x, \vartheta) \rho(x, \vartheta) \mathrm{dx}$$

$$= \int_{\mathcal{X}} \frac{d}{d\vartheta} \rho(x, \vartheta) = \frac{d}{d\vartheta} \underbrace{\int \rho(x, \vartheta) \mathrm{dx}}_{1} = 0$$

$\Rightarrow I(\vartheta) = \mathbb{E}[U_\vartheta^2]$

iii. Was bedeutet $I$ ? Falls $I = 0$ auf $\theta_0 \subseteq \theta$, d.h. $U_\vartheta(x) = 0$ für $\vartheta \in \theta_0, \forall x \in \mathcal{X}$.

$\Rightarrow \rho(x,\vartheta) = \text{const}$ für alle $x \in \mathcal{X}$ auf $\theta_0$. Also kann keine Beobachtung die Parameter in $\theta_0$ unterscheiden.

$I$ ist additiv für unabhängige Beobachtungen

**Satz 22.** *Sei $(\mathcal{X},\mathcal{F},(\mathbb{P}_\vartheta: \vartheta \in \theta))$ ein regulläres Modell mit Fisher Information $I$. Dann hat das Produktmodell $\mathcal{M}^{\otimes n}$ die Fisher Information $I^{\otimes n} = n \cdot I$.*

***Beweis.*** *Die Likelihoodfkt. von $\mathcal{M}^{\otimes n}$ ist*

$$\rho_\vartheta^{\otimes n}(x_1, \ldots, x_n) = \prod_{k=1}^{n} \rho_\vartheta(x_k)$$

*und*

$$U_\vartheta^{\otimes n}(x_1, \ldots, x_n) = \frac{d}{d\vartheta}\sum_{k=1}^{n} \log \rho_\vartheta(x_k) = \sum_{k=1}^{n} \frac{\frac{d}{d\vartheta}\rho_\vartheta(x_k)}{\rho_\vartheta(x_k)} = \sum_{k=1}^{n} U_\vartheta(x_k)$$

*Dann* $I^{\otimes n}(\vartheta) = \text{Var}[U_\vartheta^{\otimes n}] = \text{Var}[\sum_{k=1}^n U_\vartheta(x_k)] = \sum_{k=1}^n \text{Var}[U_\vartheta(x_k)] = n \cdot I$

□

Die Fisher Information kann benutzt werden für die Abschätzung der $\text{Var}_\vartheta[T]$ für reguläre erwartungstreue Schätzer T,

$$\int T(x)\frac{d}{d\vartheta}\rho(x,\vartheta)\text{dx} = \frac{d}{d\vartheta}\underbrace{\int T(x)\rho(x,\vartheta)\text{dx}}_{\mathbb{E}_\vartheta[T]}$$

**Satz 23.** *(Informationsungleichung). Sei $\mathcal{M}$ ein reguläres stat. Modell., $\tau: \theta \to \mathbb{R}$ eine zu schätzende stetig diff'bare Funktion mit $\tau' \neq 0$ und $T$ ein regulärer erwartungstreuer Schätzer für $\tau$.*

i. *Es gilt*

$$\text{Var}_\vartheta[T] \geq \frac{\tau'(\vartheta)^2}{I(\vartheta)} \, f\text{ür alle } \vartheta \in \theta. \text{ (Cramér – Rao – Ungleichung)}$$

ii. *Gleichheit gilt für alle $\vartheta \in \theta$ g.d.w.*

$$T - \tau(\vartheta) = \frac{\tau'(\vartheta)}{I(\vartheta)} U_\vartheta \forall_\vartheta$$

*d.h. wenn das Modell die Likelihoodfunktion*

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

*Wobei*

- $a: \theta \to \mathbb{R}$ *ist Stammfunktion von* $\frac{I}{\tau'}$
- $h: \mathcal{X} \to (0,\infty)$ *messbar.*
- $b(\vartheta) := \log(\int e^{a(\vartheta)T(x)}h(x)\text{dx})$ *(Normierugsfunktion)*

***Beweis.*** ***(i)*** $\mathrm{Cov}_\vartheta[T,U_\vartheta] := \mathbb{E}[T\cdot U_\vartheta] - \mathbb{E}[T]\mathbb{E}[U_\vartheta] \overset{\mathbb{E}[U_\vartheta]=0}{=\!=\!=} \mathbb{E}[T\cdot U_\vartheta]$

$$= \int_{\mathcal{X}} T(x)U_\vartheta(x)\rho(x,\vartheta)\mathrm{dx} = \int_{\mathcal{X}} T(x)\frac{d}{d\vartheta}\rho(x,\vartheta)\mathrm{dx}$$

$$= \frac{d}{d\vartheta}\mathbb{E}_\vartheta[T] \overset{T \quad \text{erwartungstrue}}{=\!=\!=\!=\!=} \tau'(\vartheta)$$

$\tau'(\vartheta)^2 = \mathrm{Cov}_\vartheta[T,U_\vartheta]^2 \le \mathrm{Var}_\vartheta[T]\cdot \underbrace{\mathrm{Var}_\vartheta[U_\vartheta]}_{I(\vartheta)}$

$\Rightarrow \mathrm{Var}[T] \ge \frac{\tau'(\vartheta)^2}{I(\vartheta)}$

***(ii)***

*Es gilt Gleichheit g.d.w.* $\exists\lambda\ge 0$ *s.d.*

$$(T-\mathbb{E}_\vartheta[T])^2 = \lambda(U_\vartheta)^2\,\mathbb{P}_\vartheta - f.s.$$

*Es gilt* $\mathbb{E}[T-\mathbb{E}[T]] = \mathrm{Var}_\vartheta[T]$ *und* $\mathbb{E}_\vartheta[\lambda\cdot U_\vartheta^2] = \lambda\mathbb{E}[U_\vartheta^2] = \lambda\cdot I(\vartheta)$

$\Rightarrow \lambda = \frac{\tau'(\vartheta)^2}{I(\vartheta)^2} \ge 0$

$\Rightarrow T - \underbrace{\mathbb{E}_\vartheta[T]}_{\tau(\vartheta)} = \frac{\tau'(\vartheta)}{I(\vartheta)}\cdot U_\vartheta \qquad \mathbb{P}_\vartheta\, f.s.$

*Da* $\rho(x,\vartheta) > 0$ *gilt* $\Rightarrow T - \tau(\vartheta) = \frac{\tau'(\vartheta)}{I(\vartheta)}\cdot U_\vartheta \qquad f.s.$

*Also*

$$\frac{d}{d\vartheta}\log\rho(x,\vartheta) = \frac{I(\vartheta)}{\tau'(\vartheta)}(T(x)-\tau((\vartheta)))$$

*Unbestimmte Integration in* $\vartheta$ *liefert*

$$\log\rho(x,\vartheta) - \underbrace{h(x)}_{\text{Integrationskonstante}} = a(\vartheta)T(x) - \underbrace{b(\vartheta)}_{=\int\frac{I(\tilde\vartheta)}{\tau(\tilde\vartheta)}\tau(\tilde\vartheta)d\tilde\vartheta}$$

$\Rightarrow \rho(x,\vartheta) = \exp\{a(\vartheta)T(x) - b(\vartheta)\}h(x)$

*Da* $\int_{\mathcal{X}}\rho(x,\vartheta)\mathrm{dx} = 1 \Rightarrow b(\vartheta) = \log\int e^{a(\vartheta)T(x)}h(x)\mathrm{dx}$.

*Für die Umkehrung sei* $\rho(x,\vartheta) = \exp\{a(\vartheta)T(x) - b(\vartheta)\}h(x)$

*Dann ist* $U_\vartheta(x) = \frac{d}{d\vartheta}\log\rho(x,\vartheta) = a'(\vartheta)T(x) - b'(\vartheta) = \frac{I(\vartheta)}{\tau'(\vartheta)}T(x) - \underbrace{\frac{I(\vartheta)}{\tau'(\vartheta)}\cdot\tau(\vartheta)}_{(*)}$

$\Rightarrow T(x) - \tau(\vartheta) = \frac{\tau'(\vartheta)}{I(\vartheta)}U_\vartheta$

*Warum gilt* $(*)$ *?*

$(b(\vartheta)) = \log\int e^{a(\vartheta)T(x)}h(x)\mathrm{dx}$

$$b'(\vartheta) = \frac{a'(\vartheta)\int T(x)e^{a(\vartheta)T(x)}h(x)\mathrm{dx}}{\int e^{a(\vartheta)T(x)}h(x)\mathrm{dx}} = \frac{a'(\vartheta)\int T(x)e^{a(\vartheta)T(x)-b(\vartheta)}h(x)\mathrm{dx}}{\underbrace{\int \underbrace{e^{a(\vartheta)T(x)-b(\vartheta)}h(x)}_{=\rho(x,\vartheta)}\mathrm{dx}}_{=1}} = a'(\vartheta)\mathbb{E}_\vartheta[T] = a'(\vartheta)\tau(\vartheta)$$

$$= \frac{I(\vartheta)}{\tau'(\vartheta)}\tau(\vartheta)$$

*ad(**) Wann gilt Gleichheit ?*

$c(\vartheta)=\frac{\tau'(\vartheta)}{I(\vartheta)}$

$0 \leq \mathrm{Var}[T-c(\vartheta)U_\vartheta]=\mathrm{Var}[T]-2c(\vartheta)\mathrm{Cov}[T,U_\vartheta]+c(\vartheta)^2\mathrm{Var}[U_\vartheta]=\mathrm{Var}[T]-2c(\vartheta)\tau'(\vartheta)+c(\vartheta)^2 I(\vartheta)$

$=\mathrm{Var}[T]-2\frac{\tau'(\vartheta)^2}{I(\vartheta)}+\frac{\tau'(\vartheta)^2}{I(\vartheta)}=\mathrm{Var}[T]-\frac{\tau'(\vartheta)^2}{I(\vartheta)}$

$\Rightarrow \mathrm{Var}[T]\geq\frac{\tau'(\vartheta)^2}{I(\vartheta)}$

*Gleichheit gilt g.d.w.* $T(x)-c(\vartheta)U_\vartheta(x)=\mathbb{E}_\vartheta[T]=\tau(\vartheta)\ \mathbb{P}_\vartheta-f.s.\ \ldots$

□

**Bemerkung 24.** Wenn $T$ erwartungstreu regulärer Schätzer , s.d. Gleichheit in Cramèr-Rao gilt, dann ist $T$ bester Schätzer, (zumindestens für reguläre Schätzer).

Wann existieren solche Schätzer ?

Für die exponentielle Familien!

**Definition 25.** *Sei* $\mathcal{M}$ ein einparametriges Standartmodell *mit* $\theta$ *offen. Wenn die Likelihoodfkt. der Form*

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

*mit Funktionen* $a\colon\theta\to\mathbb{R}, a'\neq 0$

$h\colon\mathcal{X}\to(0,\infty)$ *und* $b=\log(\int e^{a(\vartheta)T(x)}h(x)\mathrm{dx})$

*dann heißt* $\mathcal{M}$ exponentielles Modell *und* $(\mathbb{P}_\vartheta,\vartheta\in\theta)$ *heißt* *exponentielle Familie* *bzgl. eine Statistik* $\underbrace{T\colon\mathcal{X}\to\mathbb{R}}_{f.s.\,\mathrm{nicht\,konstant}}$ .

**Beispiel 26.** (Poisson-Verteilung)

$\mathbb{P}_\vartheta$ hat die Dichte ...

i.e. T(x)=x, $a(\vartheta)=\log\vartheta$

Da T erwartungstreu ist, ist T ein bester Schätzer für $\vartheta$.

Ende Vorlesung 4

---

KEINE OFFIZIELLEN MUSTERLÖSUNGEN ZU DEN ÜBUNGSBLÄTTERN!

**Proposition 27.** *(Eigenschaften von exponentiellen Modellen)*

a) *$b(\vartheta)$ ist auf $\theta$ stetig diff'bar mit $b'(\vartheta)=a'(\vartheta)\mathbb{E}_\vartheta[T]$ (Insbesondere existiert der Erwartngswert von $T$).*

b) *Jede Statistik $S\colon\mathcal{X}\to\mathbb{R}$ mit existieremden $\mathbb{E}_\vartheta[S]$ ist regulär. Insbesondere sind $M$ und $T$ regulär und $\tau(\vartheta):=\mathbb{E}_\vartheta[T]$ ist stetig diff'bar mit $\tau'(\vartheta)=a'(\vartheta)\cdot\mathrm{Var}_\vartheta[T]\neq 0\forall\vartheta\in\theta$.*

c) *Es gilt $I(\vartheta)=a'(\vartheta)\tau'(\vartheta)\forall\vartheta\in\theta$.*

*Wir beweisen die Prop. nach folgenden Korollar*

**Folgerung 28.** *(Existenz von besten Schätzern)*
*Für jedes exponentielle Modell $\mathcal{M}$ ist die zugrundeliegende Statistik $T$ für*

$$\tau(\vartheta) := \mathbb{E}_\vartheta[T] = \frac{b'(\vartheta)}{a'(\vartheta)}$$

*und es gilt $I(\vartheta) = a'(\vartheta)\tau'(\vartheta)$*

$$\operatorname{Var}_\vartheta[T] = \frac{\tau'(\vartheta)}{a'(\vartheta)} \forall \vartheta \in \theta.$$

***Beweis.*** *Nach der obigen Prob. 27 ist $M$ und $T$ regulär. Da $\operatorname{Var}_\vartheta[T] = \frac{\tau'(\vartheta)}{a'(\vartheta)} = \underbrace{\frac{\tau'(\vartheta)^2}{I(\vartheta)}}_{\geq \text{Cramèr}-\text{Rao}}$ folgt aus Satz 23 die Behauptung.* □

**Beweis.** (Von Prop. 27)
Wir nehmen an, dass $a(\vartheta) = \vartheta$ (Da $a'(\vartheta) \neq 0$ folgt die allg. Aussage mit Kettenregel).
(Sonst $\left(\tilde{\vartheta} = a(\vartheta) \Rightarrow \frac{d}{d\vartheta}(\ldots\ldots) = a'(\vartheta)\frac{d}{d\tilde{\vartheta}}(\ldots\ldots)\right)$)
Sei S in $\mathcal{L}^1$
Sei $u_S(\vartheta) := e^{b(\vartheta)}\mathbb{E}_\vartheta[S] = \int_\mathcal{X} S(x)h(x)e^{\vartheta T(x)}\mathrm{dx}$.
$u_S(\vartheta)$ ist (reell)-analytisch in $\vartheta$, denn für $\vartheta + t \in \theta$ und $a_k = \int_\mathcal{X} \frac{S(x)h(x)T(x)^k}{k!} e^{\vartheta T(x)}\mathrm{dx}$ gilt

$$(*) \sum_{k=0}^{\infty} |a_k|\,|t|^k \leq \sum_{k=0}^{\infty} \frac{|t|^k}{k!} \int_\mathcal{X} |S(x)|h(x)|T(x)|^k e^{\vartheta T(x)}\mathrm{dx}$$

$$\overset{\text{mon. Konv.}}{=\!=} \int_\mathcal{X} |S(x)|h(x)e^{\vartheta T(x)} \underbrace{\sum_{k=0}^{\infty} \frac{|t|^k}{k!}|T(x)|^k}_{\exp(|\mathrm{tT}(x)|)}\mathrm{dx}$$

$$= \int_\mathcal{X} |S(x)|h(x) \underbrace{e^{\vartheta T(x) + |\mathrm{tT}(x)|}}_{\leq \begin{cases} e^{\vartheta T(x) + \mathrm{tT}(x)} & ,\mathrm{tT}(x) > 0 \\ e^{\vartheta T(x) - \mathrm{tT}(x)} & ,\mathrm{tT}(x) < 0 \end{cases}} \mathrm{dx}$$

$$\int_\mathcal{X} |S(x)| \underbrace{h(x)e^{(\vartheta+t)T(x)}}_{e^{b(\vartheta)}\rho_{\vartheta+t}(x)}\mathrm{dx} + \int_\mathcal{X} |S(x)| \underbrace{h(x)e^{(\vartheta-t)T(x)}}_{e^{b(\vartheta)}\rho_{\vartheta-t}(x)}\mathrm{dx} < \infty$$

Da $S \in \mathcal{L}^1(\mathbb{P}_{\vartheta\pm 1})$.
$\Rightarrow u_S(\vartheta + t) = \int_\mathcal{X} S(x)h(x)e^{(\vartheta+t)T(x)}\mathrm{dx} \overset{(**)}{=\!=} \sum_{k=0}^{\infty} t^k a_k$.
$(**)$ gilt, da $e^{\mathrm{tT}(x)} = \sum_{k=0}^{\infty} t^k \frac{T(x)^k}{k!}$.
und Summe und Integral vertauscht werden dürfen wegen (*).
Also ist $u_S(\vartheta)$ analytisch und ist seine Taylorreihe, d.h.

$$u_S'(\vartheta) = a_1 = \int_\mathcal{X} S(x)T(x)h(x)e^{\vartheta T(x)}\mathrm{dx} = e^{b(\vartheta)}\mathbb{E}_\vartheta[\mathrm{ST}]$$

$$u_S''(\vartheta) = a_2 = \int_\mathcal{X} S(x)T(x)^2h(x)e^{\vartheta T(x)}\mathrm{dx} = e^{b(\vartheta)}\mathbb{E}_\vartheta[\mathrm{ST}^2]$$

Für $S=1$: gilt also $u_1(\vartheta)=e^{b(\vartheta)}$, $u_1'(\vartheta)=e^{b(\vartheta)}\mathbb{E}_\vartheta[T]$, $u_1''(\vartheta)=e^{b(\vartheta)}\mathbb{E}_\vartheta[T^2]$.

$\Rightarrow b'(\vartheta)=\frac{d}{d\vartheta}\log u_1(\vartheta)=\frac{u_1'(\vartheta)}{u_1(\vartheta)}=\mathbb{E}_\vartheta[T]=:\tau(\vartheta)$.

$\Rightarrow$a)

b)

$$\tau'(\vartheta)=b''(\vartheta)=\frac{u_1''(\vartheta)}{u_1(\vartheta)}-\left(\frac{u_1'(\vartheta)}{u_1(\vartheta)}\right)^2$$

$$=\mathbb{E}_\vartheta[T^2]-(\mathbb{E}_\vartheta[T])^2=\mathrm{Var}[T]$$

Allgemeine S (Regularität)

$$\frac{d}{d\vartheta}\int_{\mathcal{X}}S(x)\rho(x,\vartheta)\mathrm{dx}=\frac{d}{d\vartheta}\mathbb{E}_\vartheta[S]=\frac{d}{d\vartheta}[e^{-b(\vartheta)}u_S(\vartheta)]$$

$$=(u_S'(\vartheta)-u_S(\vartheta)b'(\vartheta))e^{-b(\vartheta)}=\mathbb{E}_\vartheta[\mathrm{ST}]-\mathbb{E}_\vartheta[S]\underbrace{\mathbb{E}_\vartheta[T]}_{=:\tau(\vartheta)}$$

$$\mathbb{E}_\vartheta\left[S\left(\underbrace{T-\tau(\vartheta)}_{=(a'(\vartheta))U_\vartheta:\text{Satz }23\text{ (ii)}}\right)\right]=\mathbb{E}_\vartheta[\mathrm{SU}_\vartheta]$$

$$=\int_{\mathcal{X}}S(x)\frac{d}{d\vartheta}\rho(x,\vartheta)\mathrm{dx}$$

$\Rightarrow$ S regulär $\Rightarrow\mathcal{M}$ regulär (Vertauschungsregel gilt).

$\Rightarrow$ T regulär.

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

$\Rightarrow I(\vartheta)=\mathrm{Var}_\vartheta[U_\vartheta]=\mathrm{Var}[T]=\tau'(\vartheta)>0$ (da T f.s. nicht konstant)

(*) $\mathcal{M}$ regulär:

- $\theta$ offen,
- $\rho(x,\vartheta)>0$ und nach $\vartheta$ stetig diff'bar
- $I(\vartheta)>0$ und Vertauschungsregel □

**Beispiel 29.** (Binomialverteilung)

$\rho(x,\vartheta)=\binom{n}{x}\vartheta^x(1-\vartheta)^{n-x}$

$T(x)=\frac{x}{n}$ (ML − Schätzer)

$a(\vartheta)=n\log\left(\frac{\vartheta}{1-\vartheta}\right)$

$b(\vartheta)=-n\ln(1-\vartheta)$

$h(x)=\binom{n}{x}$

$\Rightarrow T$ ist bester Schätzer mit $\mathrm{Var}_\vartheta[T]=\frac{1}{a'(\vartheta)}=\frac{\vartheta(1-\vartheta)}{n}$ (Folgerung 28)

**Bemerkung 30.** (Produktmodelle)

Sei $\mathcal{M}=(\mathcal{X},\mathcal{F},\mathbb{P}_\vartheta:\vartheta\in\theta)$ ein exp. Modell bzgl. einer Statistik $T:\mathcal{X}\to\mathbb{R}$. So ist $\mathcal{M}^{\otimes n}$ mit Statistik

$T_n(x_1,\ldots,x_n) = \frac{1}{n}\sum_{k=1}^n T(x_k)$ und $T_n$ ist bester Schätzer für $\tau(\vartheta) := \mathbb{E}_\vartheta[T]$

**Beweis.** $\rho^{\otimes n}(x,\vartheta) = \prod_{k=1}^n \rho(x,\vartheta)$

$$=\exp(n\,a(\vartheta)T_n - \mathrm{nb}(\vartheta))\prod_{k=1}^n h(x_k)$$

und $\mathbb{E}_\vartheta[T_n] = \mathbb{E}_\vartheta[T]$ Aus Folgerung 28 folgt die Behauptung.

□

## 2.5 Bayes-Schätzer

$\mathcal{M}$ Standartmodell.

- Diesmal ist das Ziel nicht die Minimierung $\mathbb{F}_\vartheta[T]$ für alle $\vartheta$, sondern die Minimierung von dem in $\vartheta$ gemittelten quadratischen Fehler.
- Für gegebenes $\vartheta$ und Schätzer T von $\tau(\vartheta)$ sei

  $$L(\vartheta,T)\,\text{eine}\,''\text{Verlustfunktion}''$$

  $$z.B. L(\vartheta,T) = |T(x) - \tau(\vartheta)|^2$$

  Dann ist $R(\vartheta,T) := \mathbb{E}_\vartheta[L(\vartheta,T)]$ das Risiko und wir wollen es minimieren.
- Aus irgendwelchen Daten nehmen wir an, dass die Werte von $\vartheta$ nicht unbed. gleichhäufig sind, aber haben eine Verteilungsdichte $a(\vartheta)$ (a priori Verteilung).

Ende Vorlesung 5

---

**Definition 31.**

*i. Das Bayesrisiko des Schätzers T bzgl. $\alpha$ und L ist gegeben durch*

$$r(\alpha,T) := \int_\theta \alpha(\vartheta)R(\vartheta,T)d\vartheta = \int_\theta\int_{\mathcal{X}} \alpha(\vartheta)\rho(x,\vartheta)L(\vartheta,T(x))\mathrm{dx}\mathrm{d}\vartheta$$

*ii. Ein Schätzer T heißt Bayes-Schätzer von $\tau(\vartheta)$ bzgl. $\alpha$ und L falls für alle anderen Schätzer S von $\tau(\vartheta)$*

$$r(\alpha,T) \leq r(\alpha,S).$$

Man kann $\alpha(\vartheta)\rho(x,\vartheta)$ so interpretieren

a) zunächst zieht man $\vartheta$ gemäß der Dichte $\alpha$ und dann zieht man $x$ gemäß der Likelihoodfunktion $\rho(x,\vartheta)$.

b) Wenn x gezogen ist, verändert dies die Information über $\alpha$. Statt $\alpha(\vartheta)$ haben wir $\rho(x,\vartheta)\alpha(\vartheta)$.

Man definiert „a posteriori-Dichte“-Dichte

$$\Pi_x(\vartheta) := \frac{\alpha(\vartheta)\rho(x,\vartheta)}{\underbrace{\int_\theta \alpha(\tilde\vartheta)\rho(x,\tilde\vartheta)d\tilde\vartheta}_{=:\rho_\alpha}}$$

(„bedingte Verteilung der Parameter auf Beobachtung x“)

**Satz 32.** *Für den Spezialfall* $L(\vartheta, T) = |T(x) - \tau(\vartheta)|^2$ *ist der Bayes-Schätzer* $T$ *von* $\tau(\vartheta)$ *mit* $\mathbb{E}_\alpha[\tau^2] < \infty$ *bzgl.* $\alpha$ *ist gegeben durch*

$$T(x) := \mathbb{E}_{\Pi_x}[\tau] = \int_\theta \tau(\vartheta)\Pi_x(\vartheta)d\vartheta \quad \forall x\, \rho_a\, f.s.$$

***Beweis.*** *Sei* $r(\alpha, S) = \int_\theta\int_{\mathcal{X}} \alpha(\vartheta)\rho(x,\vartheta)|S(x) - \tau(\vartheta)|^2 \mathrm{dx}\mathrm{d}\vartheta$

$$\Rightarrow r(\alpha,S) - r(\alpha,T) \overset{\text{Fubini}}{=\!=} \int_{\mathcal{X}}\int_\theta \rho_\alpha(x)\Pi_x(\vartheta)[|S(x)-\tau(\vartheta)|^2 - (T(x)-\tau(\vartheta))^2]d\vartheta\mathrm{dx}$$

$$= \int_{\mathcal{X}}\int_\theta \rho_\alpha(x)\Pi_x(\vartheta)[S(x)^2 - 2S(x)\tau(\vartheta) - T(x)^2 + 2T(\vartheta)\tau(\vartheta)]d\vartheta\mathrm{dx}$$

*Da* $\int \Pi_x d\vartheta = 1$

$$\overset{\text{def.}T}{=\!=} \int_{\mathcal{X}} \rho_\alpha(x)\left[\underbrace{S(x)^2 - 2S(x)T(x) - T(x)^2}_{(S(x)-T(x))^2}\right]d\vartheta\mathrm{dx}$$

$\Rightarrow$ *$T$ ist Bayes Schätzer. Gleichheit gilt genau dann wenn* $S(x) = T(x)$ $\rho_\alpha$ *f.s.*

□

**Beispiel 33.** (Auto Versicherung)

$\vartheta =$ Schadenshäufigkeit pro Jahr.

Anfangsbewertung $U_{[0,1]} \Rightarrow \alpha(\vartheta) = 1$ auf $[0,1]$.

Nach $n$ Jahren hat der Kunde $x$ Schaden produziert.

$$\Rightarrow \Pi_x(\vartheta) = \frac{1\cdot \mathrm{Bin}_{n,\vartheta}(x)}{\mathrm{Normalisierung}} = \frac{\binom{n}{x}\vartheta^x(1-\vartheta)^{n-x}}{\int_0^1 \binom{n}{x}\underbrace{\tilde\vartheta^x(1-\vartheta)^{n-x}d\tilde\vartheta}_{=\text{Beta Fkt. mit dem}\int}} = \frac{\vartheta^x(1-\vartheta)^{n-x}}{B(x+1, n-x+1)}$$

Schätzer für $\tau(\vartheta) = \vartheta$

$$T(x) = \int_0^1 \frac{\vartheta^x(1-\vartheta)^{n-x}}{B(x+1,n-x+1)}\vartheta d\vartheta = \frac{x+1}{n+2}.$$

(Aus Beispiel ...)

# 3 Konfidenzbereiche

**Beispiel 34.** Betrachten wir das Binomialmodell

$\mathcal{X} = \{0, \ldots, n\} = \#\text{Erfolge in } n \text{ unab. Versuchen}$

$\mathbb{P}_\vartheta = \mathrm{Bin}_{n,\vartheta}$

$\Rightarrow$ Likelihoodfkt. $=\rho(x,\vartheta) = \binom{n}{x}\vartheta^x(1-\vartheta)^{n-x}$

$\Rightarrow$ Wir haben gesehen, dass ML Schätzer ist gegeben durch $T(x) = \frac{x}{n}$.

Es gibt zwei Personen, Hans und Otto, die die folgenden Ergebnisse bekommen in n=100 Messungen:

Hans: 40 Mal Erfolg $\Rightarrow$Schätzung $T_H = 0.4$

Otto: 55 Mal Efolg $\Rightarrow$Schätzung $T_H = 0.55$

Frage: Wer hat Recht ? Keiner !

Was dann ? Um seriöse Aussagen zu machen, müssen wir Abweichungen zulassen.

z.B. Hans sagt „Mit W'keit 0.9 ist $\vartheta \in [0.32; 0.49]$“

Otto sagt „Mit W'keit 0.9 ist $\vartheta \in [0.46; 0.64]$“

Frage: Wie kommen die beiden auf ihre Aussagen?

**Definition 35.** *Sei $(\mathcal{X}, \mathcal{F}, (\mathbb{P}_\vartheta : \vartheta \in \theta))$ ein stat. Modell, $\Sigma$ eine bel. Menge, $\tau: \theta \to \Sigma$ eine unbekannte Größe und $0 < \alpha < 1$.*

*Eine Abbildung*

$$C: \mathcal{X} \to \mathcal{P}(\Sigma)$$

$$x \mapsto C(x) \subset \Sigma$$

*heißt Konfidenzbereich für $\tau$ zum Irrtumsniveau $\alpha$, wenn*

$$\inf_{\vartheta \in \theta} \mathbb{P}_\vartheta[x \in X : \tau(\vartheta) \in C(x)] \geq 1 - \alpha.$$

*Falls $\Sigma = \mathbb{R}$ und* jedes $C(x)$ *ein Intervall ist, dann spricht man von Konfidenzintervall.*

**Bemerkung 36.**

i. Wir wollen $C(x)$ möglichst klein, aber auch $\alpha$ möglichst klein. Diese zwei Wünsche konkurieren. Je kleiner $\alpha \Rightarrow$ desto größer wird $C(x)$.

   $\alpha = 0 \Rightarrow C(x) = \Sigma$

   $\alpha = 1 \Rightarrow C(x) = \text{ein Punkt}$

ii. Mögliches Missverständnis!

   $\vartheta \in [0.32, 0.49] = C(0.4)$ mit $\alpha = 0.1$. Das bedeutet nicht dass $\vartheta$ in 90% der Fälle in dem Interval liegt: $\vartheta$ ist unbekannt, aber nicht zufällig.

   Das bedeutet, dass in 90% der Beobachtungen (also aller x) ist das $\vartheta \in C(x)$.

Konstruktion von Konfidenzbereichen

Sei $(\mathcal{X}, \mathcal{F}, (\mathbb{P}_\vartheta : \vartheta \in \theta))$ ein stat. Modell. Nehmen wir den Fall $\tau(\vartheta) = \vartheta \Rightarrow \Sigma = \theta$.

Für jedes $\vartheta \in \theta$, sei $C_\vartheta$ eine Untermenge s.d.

$$\mathbb{P}_\vartheta[C_\vartheta] \geqslant 1 - \alpha$$

und $C_\vartheta$ so klein wie möglich. (z.B. Standartmodell)

$C = \{(x, \vartheta) \in \mathcal{X} \times \theta : x \in C_\vartheta\}$

Um für eine geg. $x \in \mathcal{X}$ $C(x)$ zu bestimmen, muss man den vertikalen Schnitt betrachten, d.h.

$$C(x) = \{\vartheta \in \theta : x \in C_\vartheta\}$$

$\Rightarrow \mathbb{P}_\vartheta[x \in \mathcal{X} : \vartheta \in C(x)] = \mathbb{P}_\vartheta[x \in \mathcal{X} : \vartheta \in C_\vartheta] \geq 1 - \alpha.$

$\Rightarrow C$ ist Konfidenzbereich für $\vartheta$ zum Irrtumniveau $\alpha$.

**Beispiel 37.** (Schätzung mittlere Lebenszeit von radioaktiven Zerfall)

$\mathcal{M} = \{\mathbb{R}_+ \mathcal{B}, \mathbb{P}_\vartheta\}$ mit $\mathbb{P}_\vartheta = \underbrace{\frac{1}{\vartheta}}_{\text{Ereignisrate}, \vartheta = \text{mittlere} \quad \text{Lebensdauer}} e^{-\frac{x}{\vartheta}} \quad , x \geq 0$

Sei $x > 0$ Messung.

Für geg. $\vartheta$ suchen wir $C_\vartheta$ $s.d.$

$$\mathbb{P}_\vartheta[C_\vartheta] \geq 1 - \alpha$$

$$\alpha = \int_{x^*}^{\infty} \frac{1}{\vartheta} e^{-x/\vartheta} \mathrm{dx}$$

$$\Rightarrow x^* = -\vartheta \log \alpha$$

$\Rightarrow C_\vartheta = [0, -\vartheta \log \alpha]$

$\Rightarrow C(x) = \{\vartheta : x \in C_\vartheta\}.$

$x \in C_\vartheta \Rightarrow x = -\vartheta \log\alpha \Rightarrow \vartheta = -\frac{x}{\log\alpha}$

$C(x) = \left[-\frac{x}{\log\alpha}, \infty\right)$

Ende Vorlesung 6

---

Wiederholung des Beispieles der Bayesschätzer

**Beispiel 38.** (Auto Versicherung, zweiter Versuch)

Neukunde hat $\vartheta \in [0, 1]$ W-keit mindestens einen Schaden pro Jahr zu produzieren (unbekannt).

Vorbewertung (a priori) des Risikos ist $\mathcal{U}_{[0,1]} \Rightarrow \alpha(\vartheta) = 1$ auf $[0, 1]$.

Nach $n$ Jahren hat der Kunde $x$ Schaden produziert.

Hier steht was anderes??

**Beispiel 39.** Beispiel 37

Jetzt machen wir $n \geq 2$ unabhängige Messungen und sei $x = \frac{1}{n} \sum_{i=1}^{n} x_i$ die mittlere Lebenszeit der Messungen (empirisch).

Dann wegen $\rho_\vartheta(x) = \gamma_{\frac{1}{\vartheta},1}(x)$ (Gamma-Verteilung)

$$\gamma_{b,p}(x) = \begin{cases} \frac{b^p}{\Gamma(p)} x^{p-1} e^{-\mathrm{bx}} & x > 0 \\ 0 & \text{sonst} \end{cases}$$

und $\gamma_{b,r}(x) * \gamma_{b,s}(x) = \gamma_{b,r+s}(x)$ (Faltungshalbgruppe) haben wir:

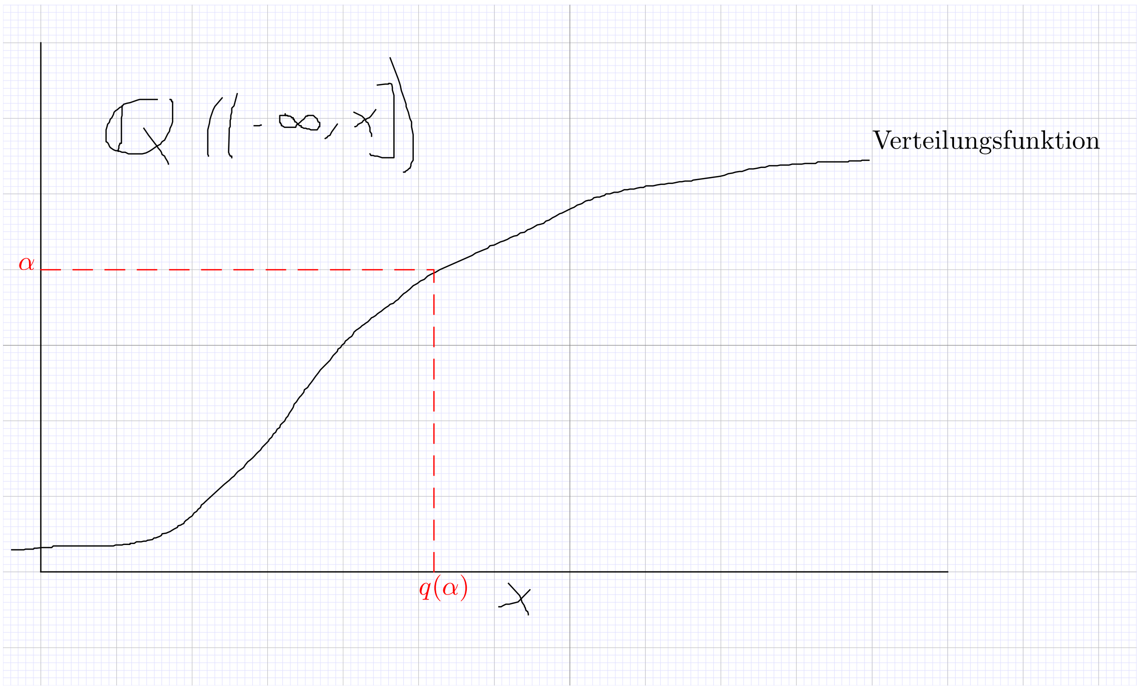
$$\rho_\vartheta^{(n)}(X) = \gamma_{\frac{1}{\vartheta},n}(x) = \left(\frac{x}{\vartheta}\right)^{n-1} \frac{e^{-x/\vartheta}}{(n-1)!} \frac{1}{\vartheta}, x > 0.$$

Für $n \geq 2$ haben wir eine obere Schranke. Nach dem finden des Konfidenzintervalles durch n teilen.

**Definition 40.** *Sei $\mathbb{Q}$ ein W.maß auf $(\mathbb{R}, \mathcal{B})$ und $0 < \alpha < 1$. Jede Zahl $q \in \mathbb{R}$ mit*

$$\mathbb{Q}[(-\infty, q]] \geq \alpha \quad \textit{und} \quad \mathbb{Q}[[q, \infty)] \geq 1 - \alpha$$

ist ein $\alpha$-Quantiel von $\mathbb{Q}$. Ein $\frac{1}{2}$-Quantiel heißt Median. Ein (1-$\alpha$)-Quantil heißt $\alpha$-Fraktil. Ein $\frac{1}{4}$-Quantil heißt unteres Quartiel. Ein $\frac{3}{4}$-Quantil heißt oberes Quartiel.




Im abs. stetigen Fall $\mathbb{Q}[(-\infty, q]] = \alpha = \int_{-\infty}^{q} \rho(x)\mathrm{dx}$. Wenn Dichte $\rho(x) > 0 \Rightarrow$ eindeutig.

Anwendung: Konfidenzintervalle für den Mittelwert im Gauß'schen Produktmodell. Sei das Modell $(\mathcal{X}, \mathcal{F}, (\mathbb{P}_\vartheta: \vartheta \in \theta)) = (\mathbb{R}^n, \mathcal{B}^n, \mathcal{N}_{m,v}^{\otimes n})$

Gesucht: Konfidenzintervall für m:

Für jedes $m \in \mathbb{R}$ wird Menge $C_m$ gesucht, s.d. $\mathcal{N}_{m,v}^{\otimes n}(C_m) \geq 1 - \alpha$ und dann $C(x) = \{m \in \mathbb{R}: x \in C_m\}$.

Sei für jedes $m \in \mathbb{R}$ die Statistik

$$T_m := \frac{M - m}{\sqrt{V^*/n}}$$

wobei:

$$M = \frac{1}{n}\sum x_k, V^* = \frac{1}{n-1}\sum (x_k - M)^2.$$

Wir behaupten: Dann hängt die Verteilung

$$Q := \mathcal{N}_{m,v}^{\otimes n} \circ T_m^{-1}$$

nicht von $(m, v)$ ab. ((Q,$T_m$) ist ein Pivot für m).

**Beweis.** Sei $\mathcal{S}_{m,v} := \left(\frac{X_k - m}{\sqrt{v}}\right)_{k=1,\ldots,m}$. Dann, da $X_k \sim m + \sqrt{v}Z_k, Z_k \sim \mathcal{N}(0,1)$x

$$\mathcal{N}_{m,v}^{\otimes n} \circ \mathcal{S}_{m,v}^{-1} = \mathcal{N}_{0,1}^{\otimes n}$$

Dazu $(M \circ \mathcal{S}_{m,v})(x) = \frac{1}{n}\sum_{k=1}^n \frac{X_k - m}{\sqrt{v}} = \frac{M - m}{\sqrt{v}}$

$$(M \circ \mathcal{S}_{m,v})(x) = \frac{1}{n}\sum_{k=1}^n \left(\frac{X_k - m}{\sqrt{v}} - \frac{M - m}{\sqrt{v}}\right)^2 = V^*/v$$

$\Rightarrow (T_0 \circ \mathcal{S}_{m,v})(x) = \frac{M-m}{\sqrt{v}} \frac{1}{\sqrt{V^*/v \cdot n}} = T_m.$

$\Rightarrow$ Deshalb

$$\mathcal{N}_{m,v}^{\otimes n} \circ T_m^{-1} = \mathcal{N}_{m,v}^{\otimes n} \circ S_{m,v}^{-1} \circ T_0^{-1} = \mathcal{N}_{0,1}^{\otimes n} \circ T_0^{-1} =: Q$$

Welche Verteilung hat Q?

Die Student-t-Verteilung mit (n-1)-Freiheitsgraden $t_{n-1}$:

Für $X, Y_1, \ldots, Y_n$ unabh. $N_{0,1}$-Z.V.

$\frac{\text{X}}{\sqrt{\frac{1}{\text{n}} \underbrace{\sum_{i=1}^n Y_i^2}_{\text{Chi-Quadrat}:=\gamma\left(\frac{n}{2},\frac{1}{2}\right)}}}$ ist $t_n$-verteilt.

□

Ende Vorlesung 7

---

Für gegebenes $\alpha$, sei $-x^*$, das $\frac{\alpha}{2}$-Quantil (wegen Symmetrie ist $x^*$ das $\frac{\alpha}{2}$-Fraktil).

Setzen wir $C_m = T_m^{-1}((-x^*, x^*))$, denn

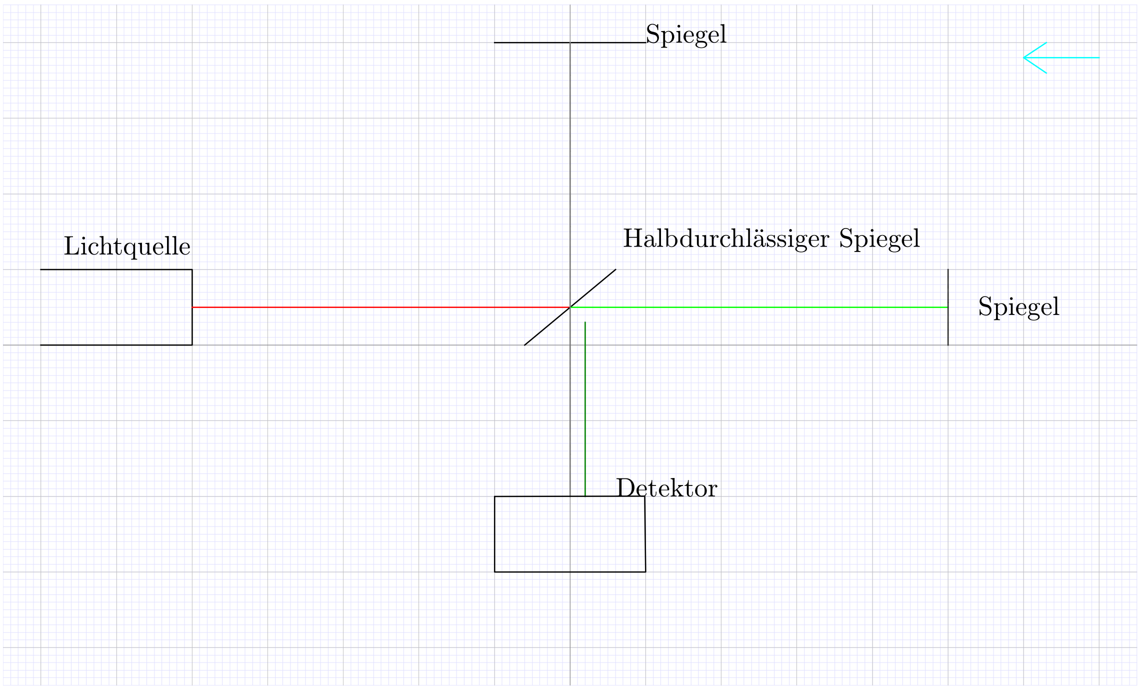
$$\mathcal{N}_{m,v}^{\otimes n}(C_m) = Q((-x^*, x^*)) \geq 1 - \alpha$$

für alle $m, v$.

**Satz 41.** *(Konfidenzintervall für den Mittelwert im Gaußmodell)*
*Sei das n-Produktmodell mit unbekanntem Erwartungswert $m$ und unbekannter Varianz $v$. Sei $x^* := F_Q^{-1}(1 - \alpha/2)$, mit $Q \sim t_{n-1}$ $(F_Q(x) = [(-\infty, x]])$.*
*$\Rightarrow C(x) = \left(M - x^*\sqrt{V^*/n}, M - x^*\sqrt{V^*/n}\right)$ ein Konfidenzintervall für $m$ zum Irrtumsniveau $\alpha$.*

Anwendung: (Bestimmung der Lichtgeschwindigkeit von Michelson und Morely)

Im Jahr 1879 hat Michelson 5 mal eine Reihe von 20 Messungen durchgeführt zur Bestimmung der Lichgeschwindigkeit.




Annahme: Messungen sind Normalverteilt mit unbekannten $m, v$.

Aufgabe: Wie kann man für jede Reihe $(x_{k,1}, \ldots, x_{k,20}), k = 1, \ldots, 5$ ein Konfidenzintervall für $m$ zum Irrtumsniveau $\alpha = 0.02$ bestimmen?

Lösung: Anwenden des Satzes:

1. $x^* = (1 - \alpha/2) -$ Quantil von $t_{n-1}$ mit $n = 20$.

   $\approx 2.54$

2. Berechnen $M(X_{k,1}, \ldots, X_{k,20}) = \begin{cases} 299909\left[\frac{\text{km}}{s}\right] \\ 299856\left[\frac{\text{km}}{s}\right] \\ 299845\left[\frac{\text{km}}{s}\right] \\ 299821\left[\frac{\text{km}}{s}\right] \\ 299832\left[\frac{\text{km}}{s}\right] \end{cases}$

3. Berechnen von $\sqrt{V^*(X_{k,1}, \ldots, X_{k,20})} = \begin{cases} 102\left[\frac{\text{km}}{s}\right] \\ 60\left[\frac{\text{km}}{s}\right] \\ 77\left[\frac{\text{km}}{s}\right] \\ 59\left[\frac{\text{km}}{s}\right] \\ 53\left[\frac{\text{km}}{s}\right] \end{cases}$

   $\Rightarrow C(X_{k,1}, \ldots, X_{k,20}) = \begin{cases} (299851,\ 299966) \\ (299821,\ 299890) \\ (299801,\ 299888) \\ (299787,\ 299854) \\ (299802,\ 299862) \end{cases}$

Aufgabe 2) Bestimmen von Konfidenzintervall mit $\alpha = 0.02$ mit allen Daten.

$x^* = (1-\alpha/2)$-Quantil von $t_4 \approx 2.36$

$M_{\text{alle}} = 299852\,\frac{\text{km}}{s}$

$\sqrt{V_{\text{alle}}^*} = 34\frac{\text{km}}{s}$

$\Rightarrow C(\text{alle}) = (299816, 299888)$

Tatsächlich: 299792

<u>Konfidenzintervalle im Binomialmodell</u>

Sei das Binomialmodell $\mathcal{X} = \{0, \ldots, n\}, \mathbb{P}_\vartheta = \text{Bin}(n, \vartheta), \vartheta \in (0,1)$.

<u>Ges.</u>: Konfidenzintervall für $\vartheta$

<u>3 Methoden:</u>

1. Tchebyschev:

   Der beste Schätzer für $\vartheta$ ist $T(x) = \frac{x}{n}$.

   Ansatz $C(x) = \left(\frac{x}{n} - \varepsilon, \frac{x}{n} + \varepsilon\right)$ mit $\varepsilon > 0$. Die Bedingung ist

   $$\mathbb{P}_\vartheta\left[x\colon \left|\frac{x}{n} - \vartheta\,|\right]\right.$$

   $$\mathbb{P}_\vartheta[|X - n\vartheta \geq \varepsilon n|] \leq \frac{\text{Var}_\vartheta[x]}{\varepsilon^2 n^2} = \frac{n\vartheta(1-\vartheta)}{n^2\varepsilon^2} \leq \frac{1}{4n\varepsilon^2} \leq \alpha$$

   Für $\varepsilon \geq \frac{1}{\sqrt{4n\alpha}}$

   $\alpha = 0.05, n = 1000 \Rightarrow \varepsilon \geq 0.07$

2. Normalapproximation

   $\frac{x}{n} \cong \mathcal{N}\left(\vartheta, \frac{\vartheta(1-\vartheta)}{n}\right)$ für $n \gg 1$

   $$\mathbb{P}_\vartheta\left[\left|\frac{x}{n} - \vartheta\right| \geq \varepsilon\right] = \mathbb{P}_\vartheta\left[\left|\underbrace{\frac{x - n\vartheta}{\sqrt{n\vartheta(1-\vartheta)}}}_{\approx\sim\mathcal{N}(0,1)}\right| \geq \varepsilon\sqrt{\frac{n}{\vartheta(1-\vartheta)}}\right]$$

   $$\approx \Phi\left(-\varepsilon\sqrt{\frac{n}{\vartheta(1-\vartheta)}}\right) + 1 - \Phi\left(\varepsilon\sqrt{\frac{n}{\vartheta(1-\vartheta)}}\right) \overset{\text{symm.}}{=\!=\!=} 2\Phi\left(-\varepsilon\sqrt{\frac{n}{\vartheta(1-\vartheta)}}\right) \leq \alpha$$

   $\Rightarrow \varepsilon_\vartheta \geq -\sqrt{\frac{\vartheta(1-\vartheta)}{n}}\Phi^{-1}(\alpha/2) = \sqrt{\frac{\vartheta(1-\vartheta)}{n}}\Phi^{-1}(1-\alpha/2) = \sqrt{\frac{\vartheta(1-\vartheta)}{n}}\left(\frac{\alpha}{2} - \text{Fraktil von}\,\mathcal{N}(0,1)\right)$.

   Da $\vartheta(1-\vartheta) \leq \frac{1}{4}$ nehmen wir an

   $$\varepsilon \geq \max_\vartheta \varepsilon_\vartheta = \frac{1}{\sqrt{4n}}\Phi^{-1}(1-\alpha/2).$$

   Für $\alpha = 0.05, n = 1000 \Rightarrow \Phi^{-1}(1-\alpha/2) = 1.96$

   $\Rightarrow \varepsilon \geq 0.03$

3. Verwendung von Quantielen

   Lemma: Sei $n \geq 1, \mathcal{X} = \{0, \ldots, n\}$

   a) $\forall \vartheta \in (0,1)\colon x \mapsto \mathrm{Bin}(n,\vartheta)(\{x\})$ strikt wachsend auf $\{0, \ldots, (n+1)\vartheta - 1\}$

      danach stritk fallend $\Rightarrow$ x=(n+1)$\vartheta$

   b) $\forall x \neq 0$ $\vartheta \mapsto \mathrm{Bin}(n,\vartheta)[\{x, \ldots, n\}]$ auf $[0,1]$ stetig und strikt wachsend. Und es gilt

$$\mathrm{Bin}(n,\vartheta)(\{x,\ldots,n\}) = \underbrace{\beta_{x,n-x+1}}_{f_{p,q}(x)=\frac{1}{B(p,q)}x^{p-1}(1-x)^{q-1}}([0,\vartheta])$$

   Das benutzen wir nun als 3. Methode.

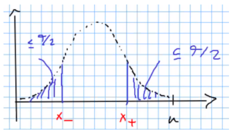



**Abbildung 1.**

   Wir schneiden für jedes $\vartheta \in (0,1)$ $\alpha/2$ von oberen/unterem Teil der Verteilung ab.

   $C_\vartheta = \{x_-(\vartheta), \ldots, x_+(\vartheta)\}$

   wobei $x_-(\vartheta) = \max\{x \in \mathcal{X}\colon \mathrm{Bin}_{n,\vartheta}(\{0,\ldots,x-1\}) \leqslant \alpha/2\}$

   $x_+(\vartheta) = \min\{x \in \mathcal{X}\colon \mathrm{Bin}_{n,\vartheta}(\{x+1,\ldots,n\}) \leqslant \alpha/2\}$

   Und C(x)=$\{\vartheta\colon x \in C_\vartheta\}$ zu finden müssen wir $x \in C_\vartheta$ nach $\vartheta$ auflösen.

   $x \leqslant x_+(\vartheta) \Leftrightarrow \mathrm{Bin}_{n,\vartheta}(\{x,\ldots,n\}) \geq \alpha/2$

   $=\beta_{x,n-x+1}([0,\vartheta])$

   $x \geq x_-(\vartheta) \Leftrightarrow \mathrm{Bin}_{n,\vartheta}(\{0,\ldots,x\}) \geq \alpha/2$

   $=1 - \beta_{x,n-x+1}([0,\vartheta])$

   d.h.$\beta_{x,n-x+1}([0,\vartheta]) > \alpha/2$ und $\beta_{x,n-x+1}([0,\vartheta]) < 1 - \alpha/2$

   Seien $\vartheta_-, \vartheta_+$ $\alpha/2$ Quantil/Fractile von $\beta_{x,n-x+1}$ (eindeutig wegen Lemma b))

   $C(x) = (\vartheta_-, \vartheta_+)$ ist Konfidenzintervall für $\alpha$, weil

$$\mathbb{P}_\vartheta[x\colon \vartheta \in C(x)] = \mathbb{P}_\vartheta[x\colon \vartheta_-(x) < \vartheta < \vartheta_+(\vartheta)] = \mathbb{P}_\vartheta[x\colon x_-(\vartheta) < x < x_+(\vartheta)]$$

   $\geq 1 - \alpha$.

Ende Vorlesung 8

---

**Satz 42.** *Im Binomialmodell,* $\mathcal{X} = \{0,\ldots,n\}, \mathbb{P}_\vartheta = \mathrm{Bin}_{n,\vartheta}, 0 < \vartheta < 1$ *ist*

$$C(x) = (\vartheta_-(x), \vartheta_+(x))$$

*wobei*

$$\vartheta_-(x) = \alpha/2 \quad \text{Quantil von } \beta_{x,n-x+1}$$

$$\vartheta_+(x) = \alpha/2 \quad \text{Fraktil von } \beta_{x+1,n-x}$$

*ein Konfidenzintervall fpr $\vartheta$ zum Irrtumsniveau $\alpha$. Für $\alpha = 0.05$ und $n = 1000$, $\varepsilon(x) = \frac{\vartheta_+(x) - \vartheta_-(x)}{2}$*

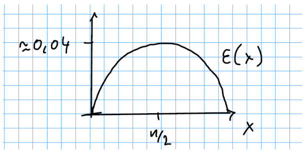



***Abbildung 2.***

**Beweis.** (voherigem Lemma)

a)

Sei $x \geqslant 1$

$$\frac{\mathrm{Bin}_{n,\vartheta}(\{x\})}{\mathrm{Bin}_{n,\vartheta}(\{x-1\})} = \frac{\binom{n}{x}\vartheta^x(1-\vartheta)^{n-x}}{\binom{n}{x-1}\vartheta^{x-1}(1-\vartheta)^{n-x+1}} = \frac{(n-x+1)\vartheta}{x(1-\vartheta)} > 1 \Leftrightarrow x < (n+1)\vartheta$$

b)

Seien $U_1, \ldots, U_n$ u.i.v. mit $\mathcal{U}_{[0,1]}$.

Dann $S_\vartheta := \sum_{k=1}^n \mathbb{1}_{[0,\vartheta]}(U_k)$ ist binomialverteilt $\mathrm{Bin}(n, \vartheta)$.

$\mathrm{Bin}(n,\vartheta)(\{x, \ldots, n\}) = \underbrace{\mathbb{P}[S_\vartheta \geqslant x]}_{\text{mind.}\,x\quad \text{der}\,U_i\quad \text{liegen}\quad \text{in}\,[0,\vartheta]}$

Seien $U_{(1)} < U_{(2)} < U_{(3)} < \cdots < U_{(n)}$ mit $\{U_1, \ldots, U_n\} = \{U_{(1)}, \ldots, U_{(n)}\}$

$\Rightarrow \mathbb{P}[S_\vartheta \geqslant x] = \mathbb{P}[U_{(x)} \leqslant \vartheta]$

$$= n! \int_0^1 \ldots \int_0^1 \mathbb{1}_{[0,\vartheta]}(t_x) \mathbb{1}_{\{t_1 < t_2 < \cdots < t_n\}} \mathrm{d}t_1 \ldots \mathrm{d}t_n$$

$$= n! \int_0^\vartheta \left( \int_0^{t_x} \ldots \int_0^{t_x} \mathbb{1}_{\{t_1 < t_2 < \cdots < t_x\}} \mathrm{d}t_1 \ldots \mathrm{d}t_{x-1} \right) \left( \int_{t_x}^1 \ldots \int_{t_x}^1 \mathbb{1}_{\{t_x < \cdots < t_n\}} \mathrm{d}t_{x+1} \ldots \mathrm{d}t_n \right) \mathrm{d}t_x$$

$$= n! \int_0^\vartheta \frac{t_x^{x-1}}{(x-1)!} \cdot \frac{(1-t_x)^{n-x}}{(n-x)!} \mathrm{d}t_x = \frac{n!}{(x-1)!(n-x)!} \int_0^\vartheta s^{x-1}(1-s)^{n-x} \mathrm{d}s = \beta_{x,n-x+1}([0,\vartheta])$$

□

<u>Ordnungsintervalle</u>

Was kann man machen wenn wir keine Ahnung von $\mathbb{P}_\vartheta$ haben?

Nicht parametrisch.

<u>Gegeben:</u> n unabhängige Beobachtungen $X_1, \ldots, X_n \in \mathbb{R}$ und einer unbekannten Verteilung $Q$.

<u>Annahme:</u>

Q hat eine Dichtefunktion $\Rightarrow Q(\{x\}) = 0 \forall x \in \mathbb{R}$

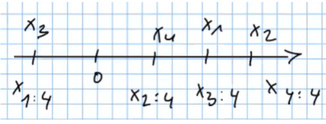
$\Rightarrow$ Mit W'keit 1 hat man unterschiedliche Beobachtungen.

**Definition 43.** *Die Ordnungstatistik $X_{1:n}, \ldots, X_{n:n}$ von Zufallsvariablen $X_1 \ldots, X_n$ ist die steigende Reihenfolge der Werte $X_1, \ldots, X_n$*

$$\{X_1, \ldots, X_n\} = \{X_{1:n}, \ldots, X_{n:n}\}$$

*mit* $X_{k:n} \leqslant X_{k+1_n}$

**Beispiel 44.**




**Definition 45.** *Die empirische Verteilung der* $X_1, \ldots, X_n$ *ist*

$$L = \frac{1}{n} \sum_{k=1}^{n} \delta_{x_k}.$$

$\Rightarrow X_{k:n} = \min \left\{ c \in \mathbb{R} | L((-\infty, c]) \geqslant \frac{k}{n} \right\} =$ das kleinste $\frac{k}{n}$-Quantil von L.

Frage: Was kann man aus den Beobachtungen über den mittleren Wert von $Q$ sagen?

Problem: „Mittelwert" oder „Erwartungswert"?

Für manche W-Maße (z.B. $\rho(x) = \frac{1}{\pi} \frac{1}{(1+x)^2}$) ex. der E-Wert nicht und das emp. Mittel kann stark durch Ausreißer beeinflusst werden.

- Der „mittlere Wert" von $Q$ im Sinne der Ordnungstatistik ist der Median (1/2-Quantil) $\mu(Q)$.
- Wir haben $Q((-\infty, \mu(Q)]) = Q((\mu(Q), \infty)) = \frac{1}{2}$.

Sei $b_n(\alpha)$ das größte $\alpha-$Quantil von $\mathrm{Bin}(n, \frac{1}{2})$, d.h.

$$b_n(\alpha) = \max \left\{ m \in \{1, \ldots, n\} | \mathrm{Bin}\left(n, \frac{1}{2}\right)(\{0, \ldots, m-1\}) \leqslant \alpha \right\}$$

**Satz 46.** *(Konfidenzinterfalle für* $\mu(Q)$*)*

*Seien* $X_1, \ldots, X_n$ *u.i.v. ZV mit unbekannter stetiger Verteilung* $Q$. *Sei* $\alpha \in (0,1)$ *und* $k = b_n(\alpha/2)$. *Dann ist*

$$C(x) = [X_{k:n}, X_{n-k+1:n}]$$

*ein Konfidenzintervall für* $\mu(Q)$ *zum Irrtumsniveau* $\alpha$.

**Beweis.**

$\mathcal{M} = (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), Q^{\otimes n}: Q\,\text{stetig})$

Sei $k = b_n(\alpha/2)$

$\Rightarrow Q^{\otimes n}[X_{k:n} > \mu(Q)] = Q^{\otimes n}[\sum_{l=1}^{n} \mathbb{1}_{\{X_l \leqslant \mu(Q)\}} < k] = \mathrm{Bin}(n, \frac{1}{2})(\{0, \ldots, k-1\}) \leqslant \alpha/2.$

Genauso sieht man :

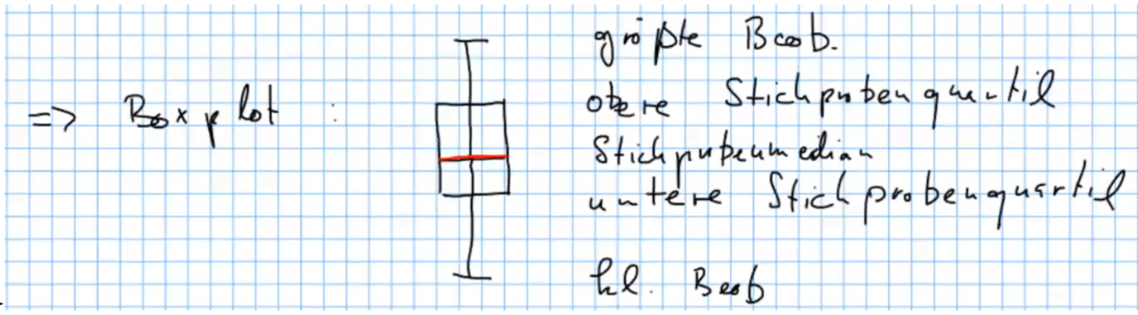
$$Q^{\otimes n}[X_{n-k+1:n} < \mu(Q)] = \mathrm{Bin}\left(n, \frac{1}{2}\right)(\{0, \ldots, k-1\}) \leqslant \alpha/2$$

$\Rightarrow Q^{\otimes n}(\mu(Q) \in [X_{k:n}, X_{n-k+1:n}]) \geqslant 1 - \alpha.$ □

**Bemerkung 47.**

$$\mu(L) = \begin{cases} X_{k+1:n} & \text{fals}\, n = 2k+1 \\ \frac{1}{2}(X_{k:n} + X_{k+1:n}) & \text{falls}\, n = 2k \end{cases}$$

der Stichprobenmedian der emp. Verteilung $L$.

$\Rightarrow$ 


$\Rightarrow$ Die Kiste im Box-Plot def. ein Konfidenzintervall zum Irrtumsniveau $\alpha$

$$\alpha = 2\text{Bin}_{n,\frac{1}{2}}(\{0, \ldots, \lfloor n/4 \rfloor\}) \approx 2(\phi(-\sqrt{n}/2) - \phi(-\sqrt{n})).$$

Ende Vorlesung 9

---

# 4 Normalverteilung, $\chi^2$,t,F-Verteilung

**Lemma 48.** *Sei $X \subseteq \mathbb{R}^n$ offen, $\mathbb{P}$ ein W'maß auf $(X, B(X))$ mit Dichte $\rho$ bzgl. $\mathcal{L}$.*

*Sei $T: X \to Y$ ($Y \subseteq \mathbb{R}^n$ offen) ein Diffeomorphismus (d.h. stetig differenziebare Bijektion, stetig differenzierbare Umkehrabbildung).*

*Dann hat $\mathbb{P} \circ T^{-1}$ die Dichte*

$$\rho_T(y) = \rho(T^{-1}(y))|\det \mathrm{D}T^{-1}(y)| \, \forall y \in Y$$

***Beweis.*** $\mathbb{P} \circ T^{-1}(A) = \int_{T^{-1}(A)} \rho(x) \mathrm{d}x \overset{\text{Trafo.}}{=\!=\!=} \int_A \rho(T^{-1}(y))|\det \mathrm{D}T^{-1}(y)| \, \mathrm{d}y$ □

**Satz 49.** *Seien $X_1, \ldots, X_n$ u.i.v. $\sim \mathcal{N}(0,1)$ und $X = (X_1, \ldots, X_n)^t$.*

*Sei $B \in \mathbb{R}^{n \times n}$ eine reguläre (det $B \neq 0$) Matrix und $m \in \mathbb{R}^n$. Dann $Y = \mathrm{BX} + m$ hat die Dichte*

$$\phi_{m,c}(y) = \frac{1}{(2\pi)^{1/2}|\det c|^{1/2}} \exp\left(-\frac{1}{2}(y-m)^t C^{-1}(y-m)\right)$$

*wobei $C = B \cdot B^t$,*

*und $\mathbb{E}[Y_k] = m_k, \mathrm{Cov}[Y_k, Y_l] = C_{k,l} \quad 1 \leq k, l \leqslant n$.*

***Bemerkung* 50.** *C ist symmetrisch (und pos. definit)* $\Rightarrow$ *diagonalisierbar.*
*Notation* $\forall n\times n$ *pos. def. symmetrisch Matrix C und* $m\in\mathbb{R}^n$ *schreiben wir* $\mathcal{N}_n(m,C)$ *für W-Maße auf* $\mathbb{R}^n$ *mit Dichtefunktion* $\Phi_{m,c}(x)$.

***Beweis.*** *Die Dichte von X ist*

$$\prod_{i=1}^{n}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x_i^2}=\frac{1}{2\pi}^{n/2}e^{-\frac{1}{2}\sum_{i=1}^{n}x_i^2}=\frac{1}{2\pi}^{n/2}e^{-\frac{1}{2}x^Tx}=\Phi_{0,E}$$

*Lemma48* $\Rightarrow$ *Y hat Dichte*

$$\Phi_{0,E}(B^{-1}(y-m))|\det B^{-1}|$$

$$=\frac{1}{\sqrt{2\pi}^n}\frac{1}{\sqrt{|\det C|}}\exp\Big(-\frac{1}{2}(y-m)^TC^{-1}(y-m)\Big)$$

$$\mathbb{E}[Y_i]=\mathbb{E}\Bigg[\sum_{k=1}^{n}B_{i,k}X_k+m_i\Bigg]=\sum B_{i,k}\underbrace{\mathbb{E}[X_k]}_{=0}+m_i=m_i$$

$$\mathrm{Cov}[Y_k,Y_l]=\sum_{i,j=1}^{n}B_{k,i}B_{l,j}\underbrace{\mathrm{Cov}[X_i,X_j]}_{=\delta_{i,j}}=\sum_{i=1}^{n}B_{k,i}B_{l,i}=C_{k,l}$$

□

Ein paar Eigenschaften:

1. $\mathcal{N}(0,E)\circ R^{-1}=\mathcal{N}(0,E)$ $(Y=\mathrm{R}\mathrm{X})$ für R orthogonal (d.h. $R^{-1}=R^T\Rightarrow|\det R|=1$)
   Drehungen + Drehspiegelungen.
2. Sei $X\sim\mathcal{N}(m,C)$, $A\in\mathbb{R}^{k\times n}$ eine Matrix mit Rang k und $a\in\mathbb{R}^k$
   $\Rightarrow Y=\mathrm{Ax}+a\sim\mathcal{N}_k(\mathrm{Am}+a,\mathrm{ACA}^t)$

Besondere Verteilungen:

$\Gamma$-Verteilung

$\beta$-Verteilung

Chi-Quadrat ...

**Satz 51.**
$X\sim N(0,1)\Rightarrow X^2\sim\Gamma_{\frac{1}{2},\frac{1}{2}}$
Beweis: Siehe Blatt 4 Aufgabe 1.

***Satz* 52.** *Seien* $\alpha,r,d>0$
$X\sim\Gamma_{\alpha,r},Y\sim\Gamma_{\alpha,s}$
$\Rightarrow X+Y\sim\Gamma_{\alpha,r+s}$
$\frac{X}{X+Y}\sim\beta_{r,s}$

*und sind unabhängig. Beweis: Übung*

**Folgerung 53.** $\Gamma_{\alpha,r} * \Gamma_{\alpha,s} = \Gamma_{\alpha,r+s}$ *(Faltungshalbgruppe)*

Aus Satz 51 + Korollar53 folgt sofort

**Satz 54.** *(Chi-Quadrat Verteilung)*
*Seien* $X_1, \ldots, X_n$ *u.i.v.*$\sim \mathcal{N}(0,1)$
$\Rightarrow Y = \sum X_k^2 \sim \Gamma_{1/2,n/2} =: \chi_n^2$
*Chi-Quadrat Vert. mit n Freiheitsgraden*
$\gamma_{1/2,n/2}(x) = \frac{x^{n/2-1}}{\Gamma(n/2)2^{n/2}} e^{-x/2}$

**Satz 55.** *(Fisher-Verteilung)*
*Seien* $X_1, \ldots X_m$, $Y_1, \ldots, Y_n$ *u.i.v.* $\mathcal{N}(0,1)$
*Dann hat* $F_{m,n} := \frac{\frac{1}{m}\sum_{k=1}^m X_k^2}{\frac{1}{n}\sum_{k=1}^n Y_k^2}$ *Dichte*

$$f_{m,n}(x) = \frac{m^{m/2} \cdot n^{n/2}}{\underbrace{B(m/2, n/2)}_{\int_0^1 x^{m/2-1}(1-x)^{n/2-1}\mathrm{dx}}} \frac{x^{m/2-1}}{(n + \mathrm{mx})^{(m+\mathrm{xn})/2}}$$

***Beweis.*** $X = \sum_{k=1}^m X_k^2 \sim \Gamma_{1/2,m/2}$ *Satz 54*
$Y = \sum_{k=1}^n Y_k^2 \sim \Gamma_{1/2,n/2}$ *Satz 54*
$\Rightarrow \frac{X}{X+Y} \sim \beta_{m/2}, n/2$ Satz 52
*Außerdem* $F_{m,n} = \frac{n}{m}\frac{X}{Y} = \frac{n}{m}\frac{Z}{1-Z} = T(Z)$
$T(x) = \frac{n}{m}\frac{x}{1-x}: (0,1) \to (0,\infty)$
$T^{-1}(y) = \frac{\mathrm{my}}{n + \mathrm{my}}$
*Lemma 48*
$\Rightarrow f_{m,n} = \beta_{m/2,n/2}\left(\frac{\mathrm{my}}{n+\mathrm{my}}\right) \cdot \frac{\mathrm{mn}}{(n+\mathrm{my})^2}$

$$= \frac{1}{B(m/2,n/2)}\left(\frac{\mathrm{my}}{n+\mathrm{my}}^{m/2-1}\left(\frac{n}{n+\mathrm{my}}\right)^{n/2-1}\frac{\mathrm{mn}}{(n+\mathrm{my})^2}\right)$$ ∎

**Definition 56.** *Die Verteilung* $F_{m,n}$ auf $(0,\infty)$ *mit Dichtefunktion* $f_{m,n}$ *heißt Fisher-Verteilung mit m und n Freiheitsgraden.*

**Bemerkung 57.** $\forall m,n \in \mathbb{N}$ $F_{m,n} = \beta_{m/2}, n/2 \circ T^{-1}$

$$T(x) = \frac{m}{n}\frac{x}{1-x}, d.h.$$

$F_{m,n}((0,c]) = \beta_{m/2,n/2}\left(\left[0, \frac{\mathrm{mc}}{n+\mathrm{mc}}\right)\right)$

**Satz 58.** *(Student Verteilung)*
*Seien* $X, Y_1, \ldots, Y_n$ *u.i.v.* $\mathcal{N}_{0,1}$. *Dann hat*

$$T := \frac{X}{\sqrt{\frac{1}{n}\sum_{k=1}^n Y_k^2}}$$

*Dichte*

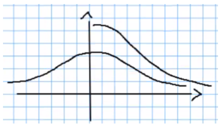
$$\tau_n(x) = \frac{1}{(1+\frac{x2}{n})^{\frac{n+1}{2}}} \frac{1}{\sqrt{n}B(1/2, n/2)}$$

***Beweis.*** *Blatt 4, Aufgabe 1, oder*

$T^2 \sim F_{1,n} \xRightarrow{\text{Lemma48}} |T| = \sqrt{T^2}$ *hat Dichte*

$$f_{1,n}(y^2)2y$$

*$T$ und $-T$ haben gleiche Verteilung (wegen Symmetrie von $\mathcal{N}_{0,1}$)*



$\Rightarrow$ *$T$ hat Dichte $f_{1,n}(y^2)|y|$*

***Abbildung 3.***

$|y| f_{1,n}(y^2) = \frac{n^{n/2}}{B(1/2, n/2)} \frac{|y|^{2(1/2)-1}}{(n+y^2)^{\frac{(n+1)}{2}}} |y|$

$$= \frac{|y|^{-1}}{\left(1+\frac{y^2}{n}\right)^{\frac{n+1}{2}} n^{\frac{1}{2}}} |y| \frac{1}{B(1/2, n/2)} = \tau_n(y)$$

$\square$

**Definition 59.** *Das W-Maß $t_n$ mit Dichte $\tau_n$ heißt Student-t-Verteilung mit $n$ Freiheitsgraden.*

**Satz 60.** *(Student/Gosset 1908)*

*Sei $\mathcal{M} = (\mathbb{R}^n, \mathcal{B}^n, (\mathcal{N}_{m,v}^{\otimes n}, m \in \mathbb{R}, v > 0))$*

*Dann gilt*

i. *$M$ und $V^*$ sind unabhängig*

ii. *$M \sim \mathcal{N}_{m,v/n}, \frac{n-1}{n} V^* \sim \chi^2_{n-1}$*

iii. *$\sqrt{n}(M-m)/\sqrt{V^*} \sim t_{n-1}$*

***Beweis.*** *(i)*

$X = (X_1, \ldots, X_n)^t$ *(m=0,v=1$\rightarrow \frac{X_k - m}{\sqrt{v}}$)*

$$O = \begin{pmatrix} \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \cdots & \frac{1}{\sqrt{n}} \\ & * & & \end{pmatrix} \quad \text{orthogonale}\, n \times n\, \text{Matrix}$$

$Y = \mathrm{OX} \Rightarrow M = \frac{1}{\sqrt{n}} \sum_{k=1}^n \frac{X_k}{\sqrt{n}} = \frac{1}{\sqrt{n}} Y_1$

$|Y| = |X|$ *weil $O$ orthogonal*

$\Rightarrow (n-1)V^* = \sum_{k=1}^n (X_k - M)^2 = \sum_{k=1}^n X_k^2 - \mathrm{nM}^2$

$$= |Y|^2 - Y_1^2 = \sum_{k=1}^n Y_k^2$$

$Y \sim \mathcal{N}(0,E) \Rightarrow Y_k$ u.i.v. $\left(\sim \mathcal{N}\left(0, \frac{1}{n}\right)\right)$. $\square$

Ende Vorlesung 10

---

## 5 Testtheorie

**Beispiel 61.** (Test faire Münze)

Sei $p \in (0,1)$ die W'keit, dass ein Münzwurf Kopf (1) ist.

Werfen n-mal die Münze $X = (X_1, \ldots, X_n)$ und

$$T(x) = \frac{1}{n} \sum_{k=1}^{n} X_k$$

ist bester Schätzer.

Aber: Mit W'keit 1 ($p \notin \mathbb{Q}$) ist $T(x) \neq p$.

Frage:

Wie können wir entscheiden ob die Münze fair ist oder nicht, d.h. die Hypothese $p = \frac{1}{2}$ testen?

In diesem Fall muss man zwischen Nullhypothese $H_0$: $p = \frac{1}{2}$ und Alternative $H_1$: $p \neq \frac{1}{2}$ entscheiden.

Fehler: Es gibt zwei Möglichkeiten einen Fehler zu machen:

1. Verwerfe $H_0$, obwohl $H_0$ vorliegt: Fehler 1. Art
2. Nehme $H_0$ an, obwohl $H_1$ vorliegt: Fehler 2. Art

Gesucht: Ein Test, dessen W'keit einen Fehler 1. Art unterhalb eines geg. Irrtumsniveau $\alpha \in [0,1]$ liegt.

*z.B.*: Falls $\left|T(X) - \frac{1}{2}\right| > \frac{1}{\sqrt{n}}$ nehmen wir $H_1$ an, sonst $H_0$.

Was ist ein Test? Wie sollte man Entscheidungverfahren durchführen?

Stat. Entscheidungsverfahren:

1. Formulierung des stat. Modells

$$(\mathcal{X}, \mathcal{F}, (\mathbb{P}_\vartheta : \vartheta \in \theta))$$

2. Formulierung von Nullhypothese und Alternative zerlegt $\theta$ in $\theta_0$ und $\theta_1$ s.d.

$$\vartheta \in \theta_0 \Leftrightarrow \vartheta \text{ ist akzeptabel, gewünschter Normalfall}$$

$$\vartheta \in \theta_1 \Leftrightarrow \vartheta \text{ ist problematisch, Abweichung vom Normalfall}$$

3. Wahl eines Irrtumsniveau

   Wähle $\alpha \in [0,1]$ und fordert, dass Fehler 1. Art höchstens mit W'keit $\alpha$ passiert.

4. Wahl einer Entscheidungsregel

Man wähle eine Statistik $\varphi: \mathcal{X} \to [0,1]$ wie folgt

$\to$ $\varphi(x)$ ist Grad mit dem man sich für die Alternative entscheidet.

$$\varphi(x) = 0 \Leftrightarrow \text{Nehme die Nullhypothese}$$

$$\varphi(x) = 1 \Leftrightarrow \text{Vewerfe die Nullhypothese und nehme Alternative}$$

$$\varphi(x) \Leftrightarrow \text{Nehme die Nullhypothese mit } W'\text{keit } 1 - \varphi(x)$$

5. Durchführung des Experiments

   Wieso erst jetzt? Sonst Täuschung fast unvermeidbar.

   Beispiel:

   - Nullhypothese und Alternative an Daten anpassen.
   - Niveau und Entscheidungsregel geeignet auswählen.

**Definition 62.** *Sei $(\mathcal{X}, \mathcal{F}, (\mathbb{P}_\vartheta: \vartheta \in \theta))$ ein stat. Modell und $\theta = \theta_0 \cup \theta_1$ eine Zerlegung von $\theta$ in Nullhypothese und Alternative.*

a) *Jede Statistik $\varphi: \mathcal{X} \to [0,1]$ heißt Test von der $\theta_0$ gegen $\theta_1$.*

b) *Ein Test $\varphi$ heißt nicht randomisiert, falls $\varphi(x) \in \{0,1\} \forall x \in \mathcal{X}$.*

   *In diesem Fall heißt $\{x \in \mathcal{X}: \varphi(x) = 1\}$ Ablehnungsbreich oder kritischer Bereich.*

c) *Falls $\varphi(x) \notin \{0,1\}$ für ein $x \in \mathcal{X}$*

   *$\Rightarrow \varphi$ ist randomisiert.*

d) *Effektives Niveau von $\varphi$ ist*

$$\sup_{\vartheta \in \theta_0} \mathbb{E}_\vartheta[\varphi]$$

   *d.h. das sup von W'keiten Fehler erster Art zu begehen.*

e) *Ein Test $\varphi$ hat (Irrtums-)niveau $\alpha$ wenn*

$$\sup_{\vartheta \in \theta_0} \mathbb{E}_\vartheta[\varphi] \leqslant \alpha$$

f) *Gütefunktion $G_\varphi: \theta \to [0,1]$*

$$\vartheta \mapsto G_\varphi(\vartheta) := \mathbb{E}_\vartheta[\varphi].$$

g) *Macht von $\varphi$ bei $\vartheta$: für $\vartheta \in \theta_1$:$G_\varphi(\vartheta)$=W'keit mit der die Alternative erkannt wird, wenn sie vorliegt $\Rightarrow \beta_\varphi(\vartheta) = 1 - G_\varphi(\vartheta)$ für alle $\vartheta \in \theta_1$ ist die W'keit für Fehler 2. Art.*

Wie soll man $\varphi$ wählen?

a) $G_\varphi(\vartheta) \leqslant \alpha$ für alle $\vartheta \in \theta_0$. $\Rightarrow \text{Irrtums} - W'\text{keit}$ für Fehler 1. Art $\leqslant \alpha$.

b) $G_\varphi(\vartheta)$ so groß wie möglich $\forall_{\vartheta \in \theta_1} \Rightarrow$ Fehler 2. Art minimieren.

**Definition 63.** *Ein Test $\varphi$ von $\theta_0$ gegen $\theta_1$ heißt bester Test zum Niveau $\alpha$, wenn Niveau($\varphi$)=$\alpha$ und $\forall$Tests $\psi$ mit Niveau$(\psi) = \alpha$ gilt $G_\varphi(\vartheta) \geq G_\psi(\vartheta) \forall \vartheta \in \theta_1$. Auf englisch: UMP-Test=uniform most powerful test.*

**Beispiel 64.** (Faire Münze)

Wir werfen $n$ mal eine Münze

1. $\mathcal{X} = \{0, \ldots, n\}, \mathbb{P}_p = \mathrm{Bin}_{n,p}$ mit $p \in [0,1]$.

2. $\theta_0 = \left\{\frac{1}{2}\right\}$ und $\theta_1 = [0,1] \setminus \left\{\frac{1}{2}\right\}$

3. $\alpha \in (0,1)$ fest z.B.: $\alpha = 0.05$ (5% Fehler 1. Art möglich)

4. $\varphi(x) = \mathbb{1}_{\{|x - \frac{n}{2}| > c\}}$

   Jetzt wollen wir $c$ berechnen s.d. der Test Niveau $\alpha$ hat.

   $\mathrm{Niveau}(\varphi) := \sup_{\vartheta \in \theta_0} \mathbb{E}_\vartheta[\varphi] = \mathbb{E}_{\frac{1}{2}}[\varphi]$

   $$= \mathbb{E}_{\frac{1}{2}}\left[\mathbb{1}_{\{|x - \frac{n}{2}| \geq c\}}\right] = \mathrm{Bin}_{n,\frac{1}{2}}\left[\left\{0, \ldots, \left\lfloor \frac{n}{2} - c \right\rfloor\right\}\right] + \mathrm{Bin}_{n,\frac{1}{2}}\left[\left\{\left\lceil \frac{n}{2} + c \right\rceil, \ldots, n\right\}\right] \leqslant \alpha$$

   Sei $k_- = \frac{\alpha}{2}$ Quantil von $\mathrm{Bin}_{n,\frac{1}{2}}$ und $k_+$ das (1-$\alpha/2$)-Quantil von $\mathrm{Bin}_{n,\frac{1}{2}} = n - k_-$

   $$\varphi(x) = \begin{cases} 0 & x \in [k_-, k_+] \\ 1 & \text{sonst} \end{cases}$$

   hat Niveau $\alpha$

**Beispiel 65.** $\alpha = 0.05$

$$\begin{pmatrix} n & k_- & \text{Eff. Niveau} \\ 10 & 2 & 2 \cdot 0.0107 = 0.021 \\ 20 & 6 & 0.041? \\ 50 & 18 & 0.033 \\ 100 & 40 & 0.035 \\ 1000 & 469 & 0.046 \\ 1000000 & 499020 & 0.0499 \end{pmatrix}$$

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 10 | 0 | 0.5987 | 0.3487 | 0.1969 | 0.1074 | 0.0563 | 0.0282 | 0.0135 | 0.0060 | 0.0025 | 0.0010 |
| | 1 | 0.9139 | 0.7361 | 0.5443 | 0.3758 | 0.2440 | 0.1493 | 0.0860 | 0.0464 | 0.0233 | 0.0107 |
| | 2 | 0.9885 | 0.9298 | 0.8202 | 0.6778 | 0.5256 | 0.3828 | 0.2616 | 0.1673 | 0.0996 | 0.0547 |
| | 3 | 0.9990 | 0.9872 | 0.9500 | 0.8791 | 0.7759 | 0.6496 | 0.5138 | 0.3823 | 0.2660 | 0.1719 |
| | 4 | 0.9999 | 0.9984 | 0.9901 | 0.9672 | 0.9219 | 0.8497 | 0.7515 | 0.6331 | 0.5044 | 0.3770 |
| | 5 | 1.0000 | 0.9999 | 0.9986 | 0.9936 | 0.9803 | 0.9527 | 0.9051 | 0.8338 | 0.7384 | 0.6230 |
| | 6 | | 1.0000 | 0.9999 | 0.9991 | 0.9965 | 0.9894 | 0.9740 | 0.9452 | 0.8980 | 0.8281 |
| | 7 | | | 1.0000 | 0.9999 | 0.9996 | 0.9984 | 0.9952 | 0.9877 | 0.9726 | 0.9453 |
| | 8 | | | | 1.0000 | 1.0000 | 0.9999 | 0.9995 | 0.9983 | 0.9955 | 0.9893 |
| | 9 | | | | | | 1.0000 | 1.0000 | 0.9999 | 0.9997 | 0.9990 |
| | 10 | | | | | | | | 1.0000 | 1.0000 | 1.0000 |

**Abbildung 4.**

Frage: Können wir nicht einfach einen Test anderer Form wählen

$$\tilde{\varphi}(x) = \mathbb{1}_{\{1, \ldots, k_- - 1\} \cup \{k_+ + 1, \ldots, n-1\}}$$

Dann wäre $\tilde{\varphi}$ ein Test mit kleinerem Niveau als $\varphi$.

Problem: Wenn $x = n$, dann $\tilde{\varphi}(x) = 0$ und die Nullhypothese $H_0$: $p = \frac{1}{2}$ wird behalten $\rightarrow$ nicht gut!

Auch $G_{\tilde{\varphi}}(p = 1) = \mathbb{E}_{p=1}[\tilde{\varphi}(x)] = \mathrm{Bin}_{n,1}(\mathbb{1}_{\{1 \leqslant x \leqslant k_- - 1\} \cup \{k_+ + 1 \leqslant x \leqslant n-1\}}) = 0 < G_{\tilde{\varphi}}\left(p = \frac{1}{2}\right)$.

D.h. im Fall von sehr starker Unfairness nehmen wir mit $\tilde{\varphi}$ $H_0$ an!

Um diese Absurdität zu vermeiden:

**Definition 66.** *Ein Test $\varphi$ heißt <u>unverfälscht</u> zum Niveau $\alpha$ wenn*

$$G_\varphi(\vartheta_0) \leqslant \alpha \leqslant G_\varphi(\vartheta_1) \forall \vartheta_0 \in \theta_0, \vartheta_1 \in \theta_1,$$

*d.h. man entscheidet sich mit größerer W'keit für die Alternative wenn sie richtig ist, als wenn sie falsch ist.*

Ende Vorlesung 11

---

## 5.1 Neyman-Pearson-Test

Jetzt betrachten wir eine besondere Situation, wo wir nur zwischen W'maßen $\mathbb{P}_0$ und $\mathbb{P}_1$ entscheiden müssen.

<u>Annahme:</u>

$(\mathcal{X}, \mathcal{F}, \mathbb{P}_0, \mathbb{P}_1)$ d.h. $\theta = \{0,1\}$, Standardmodell

Nullhypothese $\theta_0 = \{0\}$ Alternative $\theta_1 = \{1\}$

Seien $\rho_0, \rho_1$ die Dichte von $\mathbb{P}_0, \mathbb{P}_1$.

<u>Gesucht:</u> Ein bester Test $\varphi$ von $\mathbb{P}_0$ gegen $\mathbb{P}_1$.

<u>Idee:</u>

Man entscheidet sich für die Alternative, wenn $\rho_1(x)$ hinreichend größer ist als $\rho_0(x)$.

Deshalb definiert man den <u>Likelihood-Quotienten</u>

$$R(x) = \begin{cases} \frac{\rho_1(x)}{\rho_0(x)} & \rho_0(x) > 0 \\ \infty & \rho_0(x) = 0 \end{cases}$$

Hinreichend groß bedeutet, dass $R(x)$ größer als ein vorgegebener Schwellenwert $c$ ist.

$\Rightarrow$ Wir kriegen einen Test der Form

$$\varphi(x) = \begin{cases} 1 & R(x) > c \\ 0 & R(x) < c \end{cases} \tag{1}$$

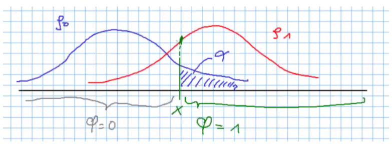



**Abbildung 5.**

**Definition 67.** *Ein Test dieser Form heißt <u>Neyman-Pearson-Test</u> zum Schwellenwert $c$.*

**Satz 68.** *(Neyman-Pearson-Lemma 1932)*

*Sei $(\mathcal{X}, \mathcal{F}, \mathbb{P}_0, \mathbb{P}_1)$ ein Standardmodell mit*

*Nullhypothese: $H_0 \colon \theta = \{0\}$ und*

*Alternative: $H_1 \colon \theta = \{1\}$ und*

*$\alpha \in (0,1)$ ein vorgegebenes Niveau. Dann gilt:*

- *$\exists$ Neyman-Pearson-Test $\varphi$ mit $\mathbb{E}_0[\varphi] = \alpha$.*
- *Jeder N-P-Test $\varphi$ mit $\mathbb{E}_0[\varphi] = \alpha$ ist ein bester Test zum Niveau $\alpha$, und jeder beste Test $\psi$ zum Niveau $\alpha$ ist ununterscheidbar von einem N-P-Test.*

**Bemerkung 69.** Im Beweis werden wir sehen:

Sei $c$ ein $\alpha$-Fraktil von $\mathbb{P}_0 \circ R^{-1}$. Dann

$$\gamma = \begin{cases} \frac{\alpha - \mathbb{P}_0[R > c]}{\mathbb{P}_0[R = c]} & \mathbb{P}_0[R = c] > 0 \\ 0 & \mathbb{P}_0[R = c] = 0 \end{cases}$$

Dann

$$\varphi(x) = \begin{cases} 1 & R(x) > c \\ \gamma & R(x) = c \\ 0 & R(x) < c \end{cases}$$

**Beweis.**

Sei c ein bel. $\alpha$-Fraktil von $\mathbb{P}_0 \circ R^{-1}$, d.h.

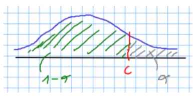



**Abbildung 6.**

$\mathbb{P}_0[R(x) \geqslant c] \geqslant \alpha$,$\mathbb{P}_0[R(x) > c] \leqslant \alpha$

Da $\mathbb{P}_0[R(x) = \infty] = 0$ existiert so ein c.

$$\mathbb{P}_0[R(x) = c] = \mathbb{P}_0[R(x) \geqslant c] - \mathbb{P}_0[R(x) > c] \geqslant \alpha - \mathbb{P}_0[R(x) > c]$$

<u>1. Fall:</u>

$$\mathbb{P}_0[R(x) = c] = 0 \Rightarrow \varphi(x) = \mathbb{1}_{\{R(x) > c\}}$$

ist ein N-P-Test mit $\mathbb{E}_0[\varphi] = \mathbb{P}_0[R(x) > c] = \alpha$.

<u>2. Fall:</u>

$\mathbb{P}_0[R(x) = c] > 0$, dann mit

$$\gamma = \frac{\alpha - \mathbb{P}_0[R(x) > c]}{\mathbb{P}_0[R(x) = c]} \in [0,1]$$

$$\varphi(x) = \begin{cases} 1 & R(x) > c \\ \gamma & R(x) = c \\ 0 & R(x) < c \end{cases}$$

ist ein N-P-Test mit Niveau

$$\mathbb{E}_0[\varphi] = \mathbb{P}_0[R(x) > c] + \gamma \cdot \mathbb{P}_0[R(x) = c] = \alpha.$$

Damit existiert so ein Test.

(ii): Sei $\varphi$ ein N-P-Test mit $\mathbb{E}_0[\varphi] = \alpha$ und Schwellenwert c und $\psi$ ein Test zum Niveau $\alpha$.

$\mathbb{E}_1[\varphi] - \mathbb{E}_1[\psi] = \int \varphi(x) - \psi(x) \rho_1(x) \mathrm{dx}$

1. Fall:

Falls $\varphi(x) > \psi(x) \Rightarrow \varphi(x) > 0$ und deshalb $R(x) \geqslant c$, d.h. $\rho_1(x) \geqslant c\rho_0(x)$.

2. Fall:

Falls $\varphi(x) < \psi(x) \Rightarrow \varphi(x) < 1$ und deshalb $R(x) \leqslant c$ d.h. $\rho_1(x) \leqslant c\rho_0(x)$.

$\Rightarrow \forall x\, \underbrace{(\varphi(x) - \psi(x))\rho_1(x)}_{f_1(x)} \geqslant c(\varphi(x) - \psi(x))\rho_0(x)$

$\Rightarrow \mathbb{E}_1[\varphi] - \mathbb{E}_1[\psi] \geqslant c \int \underbrace{\varphi(x) - \psi(x)\rho_0(x)}_{f_0(x)} \mathrm{dx} = c(\underbrace{\mathbb{E}_0[\varphi]}_{=\alpha} - \underbrace{\mathbb{E}_0[\psi]}_{\leqslant \alpha}) \geqslant 0$

$\Rightarrow \varphi$ ist ein bester Test zum Niveau $\alpha$.

Ununterscheidbar: Sei $\psi$ ein bester Test mit Niveau $\alpha$.

$\Rightarrow \int \varphi(x) - \psi(x)\rho_0(x)\mathrm{dx} = 0$

$\Rightarrow f_1(x) = \mathrm{c}f_0(x)$ bis auf $x$ aus Lebesgue-Nullmenge $N$.

$$d.h.: (\varphi(x) - \psi(x))(\rho_1(x) - c\rho_0(x)) = 0 \forall x \notin N$$

$\Rightarrow \varphi = \psi$ für alle $x \notin N$ mit $R(x) \neq c$

$\Rightarrow$ Behauptung. $\square$

**Beispiel 70.** (Entscheidung zwischen zwei möglichen Werten einer (vermutlich) fairen Münze)

Sei $p$ die W'keit, dass bei einem Münzwurf Zahl rauskommt.

Jemand behauptet, dass die Münze nicht fair ist, sondern $p = 3/4$ gilt.

Er ruft die Polizei, die mit $\underline{n = 10}$ Würfen entscheiden soll, was der Fall ist mit Irrtumsniveau $\alpha = 0.01$.

$H_0$: $p = \frac{1}{2}$ gegen $H_1$: $p = \frac{3}{4}$.

$\Rightarrow \mathbb{P}_0 = \mathrm{Bin}_{n,\frac{1}{2}}$ und $\mathbb{P}_1 = \mathrm{Bin}_{n,\frac{3}{4}}$.

$\Rightarrow R(x) = \frac{\mathrm{Bin}_{n,\frac{3}{4}}(x)}{\mathrm{Bin}_{n,\frac{1}{2}}(x)} = \frac{\binom{n}{x}\frac{3^x}{4^n}}{\binom{n}{x}\frac{1}{2^n}} = \frac{3^x}{2^n}$

ist streng monoton steigend.

Ein Test der Form

$$\varphi(x) = \begin{cases} 0 & x < a \\ \gamma & x = a \\ 1 & x > a \end{cases}$$

ist äquivalent zu

$$\varphi(x) = \begin{cases} 0 & R(x) < c \\ \gamma & R(x) = c \\ 1 & R(x) > c \end{cases}$$

mit $a = R^{-1}(c)$.

Sei $a$ ein $\alpha$-Fraktil von $\mathbb{P}_0 = \mathrm{Bin}_{10,\frac{1}{2}}$

$$\mathbb{P}_0[X \geqslant a] \geqslant \alpha \text{ und } \mathbb{P}_0[X > a] \leqslant \alpha$$

Da $\mathrm{Bin}_{10,\frac{1}{2}}(\{10\}) \cong 0.001 \leqslant 0.01$

und $\mathrm{Bin}_{10,\frac{1}{2}}(\{9\},\{10\}) \cong 0.0107 \geqslant 0.01$

$\Rightarrow a = 9$

$$\gamma = \frac{\alpha - \mathrm{Bin}_{10,\frac{1}{2}}(\{10\})}{\mathrm{Bin}_{10,\frac{1}{2}}(\{9\})} \approx 0.924$$

$$\varphi(x) = \begin{cases} 0 & x < 9 \\ 0.924 & x = 9 \\ 1 & x > 9 \end{cases}$$

Seien $\mathbb{P}_0, \mathbb{P}_1$ zwei W'maße mit Dichten $\rho_0$, $\rho_1$. Dann

$$H(\mathbb{P}_0|\mathbb{P}_1) = \mathbb{E}_{\mathbb{P}_0}\left[\ln\frac{\rho_0}{\rho_1}\right] = \begin{cases} \int \rho_0 \ln\frac{\rho_0}{\rho_1}\mathrm{dx} & \text{falls } \mathbb{P}_0[\rho_1 = 0] \\ \infty & \text{sonst.} \end{cases}$$

heißt relative Entropie von $\mathbb{P}_0$ bzgl. $\mathbb{P}_1$. Insbesondere $H(\mathbb{P}_0|\mathbb{P}_1) \geqslant 0$ und $H(\mathbb{P}_0|\mathbb{P}_1) = 0 \Leftrightarrow \mathbb{P}_0 = \mathbb{P}_1$. (Siehe Blatt 1 oder 2?)

Die statistische Bedeutung von $H$: Je größer $H(\mathbb{P}_0|\mathbb{P}_1)$ desto schneller wächst auch die Macht von optimalen Tests von $\mathbb{P}_0$ gegen $\mathbb{P}_1$ mit der Anzahl der Beobachtungen. Das ist ein Teil vom folgenden Satz

**Satz 71.** *(Lemma von Stein ' 52)*

*Sei* $(E, \mathcal{E}, \mathbb{Q}_0, \mathbb{Q}_1)$ *ein stat. Standardmodell*

$H_0: \theta = \{0\}, H_1: \theta = \{1\}$ *und sei* $(\mathcal{X}, \mathcal{F}, \mathbb{P}_\vartheta: \vartheta \in \{0,1\}) = (E^{\mathbb{N}}, \mathcal{E}^{\otimes\mathbb{N}}, \mathbb{Q}_\vartheta^{\otimes\mathbb{N}}: \{0,1\})$

$\forall n > 1$ *Sei* $\varphi_n$ ein $N-P-$Test *mit* $\mathbb{E}_0[\varphi] = \alpha$, *der nur von den ersten* $n$ *Beobachtungen* $X_1, \ldots,$ $X_n$ *abhängt. Dann*

$$\frac{\ln(1 - \mathbb{E}_1[\varphi_n])}{n} \xrightarrow{n \to \infty} -H(\mathbb{Q}_0|\mathbb{Q}_1),$$

*d.h.*

$$\mathbb{E}_1[\varphi_n] \approx 1 - \underbrace{e^{-\mathrm{nH}(\mathbb{Q}_0|\mathbb{Q}_1)}}_{=\left\{\begin{smallmatrix} 1 & H=0 \\ 0 & H>0 \end{smallmatrix}\right.}.$$

Ende Vorlesung 12

---

Beweis des Lemmas in Georgii.

**Beispiel 72.** (Test für den Erwartungswert zweier Normalverteilungen bei bekannter Varianz)

Sei $\mathbb{Q}_0 = \mathcal{N}_{m_0,v}, \mathbb{Q}_{m_1,v} = \mathcal{N}_{m_1,v}$ mit $m_0 < m_1$ und $v > 0$ fix.

$H_0: m = m_0$, gegen $H_1: m = m_1$

$\Rightarrow R_n = \exp\left(-\frac{1}{2v}\sum_{k=1}^n ((x_k - m_1)^2 - (x_k - m_0)^2)\right)$

$$= \exp\left(-\frac{n}{2v}(2(m_0 - m_1)M_n + m_1^2 - m_0^2)\right)$$

mit $M_n = \frac{1}{n}\sum x_k$

$$\Rightarrow h_n := \frac{-1}{n}\ln R_n = \frac{m_0 - m_1}{v} M_n + \frac{m_1^2 - m_0^2}{2v}$$

$\Rightarrow$ N-P-Test von $m_0$ gegen $m_1$ nach $n$ Beobachtungen

$$\varphi_n(x) := \mathbb{1}_{\{M_n > b_n\}}$$

mit $\mathbb{P}_0\Bigg[\underbrace{M_n > b_n}_{\mathcal{N}_{m_0, v/n}((b_n,\infty))}\Bigg] \stackrel{!}{=} \alpha$

$$\mathcal{N}_{m_0, v/n}((b_n, \infty)) = 1 - \Phi\left(\frac{b_n - m_0}{\sqrt{v/n}}\right)$$

$\Rightarrow b_n = m_0 + \sqrt{v/n}\,\Phi^{-1}(1-\alpha)$

Dann die rel. Entropie:

$$H(\mathbb{P}_0|\mathbb{P}_1) = \mathbb{E}_0[h_n] = m_0 \frac{m_0 - m_1}{v} + \frac{m_1^2 - m_0^2}{2v} = \frac{(m_0 - m_1)^2}{2v}$$

Stein's Lemma:

$$\mathbb{E}_1[1 - \varphi_n] \approx \exp\left(-n\frac{(m_0 - m_1)^2}{2v}\right)$$

**Beispiel 73.**

$\mathbb{Q}_0$: $\mathbb{Q}_0(\{0\}) = \mathbb{Q}_0(\{1\}) = \frac{1}{2}$

$\mathbb{Q}_1$: $\mathbb{Q}_0(\{1\}) = \frac{1}{4}$, $\mathbb{Q}_1(\{1\}) = \frac{3}{4}$

Sei $S_n = \sum x_k$ und $q_n \cong (1-\alpha)$-Quantil von $\mathrm{Bin}_{n,\frac{1}{2}}$

$$\Rightarrow \varphi_n(s) = \begin{cases} 0 & s < q_n \\ \gamma_n & s = q_n \\ 1 & s > q_n \end{cases}$$

mit $\gamma_n = \frac{\alpha - \mathrm{Bin}_{n,\frac{1}{2}}(\{q_n + 1, \ldots, n\})}{\mathrm{Bin}_{n,\frac{1}{2}}(\{q_n\})}$

$\Rightarrow \mathbb{E}_0[\varphi_n] = \alpha$

$\Rightarrow \mathbb{E}_1[\varphi_n] = \gamma_n \mathrm{Bin}_{n,\frac{3}{4}}(\{q_n\}) + \mathrm{Bin}_{n,\frac{3}{4}}(\{q_n + 1, \ldots, n\})$

und $H(\mathbb{Q}_0|\mathbb{Q}_1) = \frac{1}{2}\ln(8/3)$

Stein's Lemma :

$$\mathbb{E}_1[\varphi_n] = 1 - (3/8)^{n/2}$$

## 5.2 Beste einseitige Tests

Die Neyman-Pearson-Tests sind oft zu einfach für Anwendungen. Das ist aber der Grundstein für komplexere Tests (wenn wir eine geeignete Monotonie gilt).

**Definition 74.** *(Likelihood-Quotient $R_{\vartheta';\vartheta}(x)$)*

*Sei $(\mathcal{X}, \mathcal{F}, \mathbb{P}_0 \colon \vartheta \in \theta)$ ein Standardmodell mit $\theta \subseteq \mathbb{R}$. Das Modell hat wachsende Likelihood-Quotienten bzgl. einer Statistik*

$$T \colon \mathcal{X} \to \mathbb{R} \quad \text{wenn } \textit{für} \text{ alle } \vartheta < \vartheta'$$

$$R_{\vartheta' : \vartheta} := \frac{\rho_{\vartheta'}}{\rho_\vartheta}$$

*ist eine wachsende Funktion für T, d.h.*

$$R_{\vartheta' : \vartheta}(x) = f_{\vartheta' : \vartheta}(T(x)) \text{ mit}$$

*$f_{\vartheta' : \vartheta}(y)$ wachsend in $y$.*

**Bemerkung 75.** Jedes (einparametrige) exponentielle Modell hat wachsende Likelihood-Quotienten (bzgl. T oder -T).

In der Tat

$$R_{\vartheta' : \vartheta}(x) = \exp[(a(\vartheta') - a(\vartheta))T(x)] \frac{e^{b(\vartheta)}}{e^{b(\vartheta')}}$$

und $\vartheta \mapsto a(\vartheta)$ ist strikt monoton.

$$\begin{cases} \text{wenn a wachsend} & \text{bzgl T} \\ \text{wenn a fallend} & \text{bzgl. } -T \end{cases}$$

Erinnerung:

In den exponentiellen Modellen sind unter anderem

i. Binomialmodell

ii. Poisson

iii. Normalverteilung mit fester Varianz oder festem Erwartungswert

iv. Alle Produktmodelle von (i)-(iii)

Was ist ein einseitiger Test?

$H_0$: $\vartheta \leqslant \vartheta_0$ gegen $\vartheta > \vartheta_0$ (linksseitig)

$H_0$: $\vartheta \geqslant \vartheta_0$ gegen $\vartheta < \vartheta_0$ (rechtsseitig)

Was ist ein beidseitiger Test?

$H_0 \colon \vartheta \in [\vartheta_{0,1}, \vartheta_{0,2}]$ gegen $H_1 \colon \vartheta \notin [\vartheta_{0,1}, \vartheta_{0,2}]$.

**Satz 76.**

*Sei $(\mathcal{X}, \mathcal{F}, \mathbb{P}_\vartheta \colon \vartheta \in \theta)$ mit $\theta \subseteq \mathbb{R}$ ein Standardmodell mit wachsenden Likelihood-Quotienten bzgl. T und $\vartheta_0 \in \theta$, $0 < \alpha < 1$.*

*Dann ex. ein bester Test $\varphi$ zu dem Niveau $\alpha$ ($\mathbb{E}_{\vartheta_0}[\varphi] \leqslant \alpha$) für $H_0 \colon \vartheta \leqslant \vartheta_0$ gegen $H_1 \colon \vartheta > \vartheta_0$ der Form*

$$\varphi(x) = \begin{cases} 1 & T(x) > c \\ \gamma & T(x) = c \\ 0 & T(x) < c \end{cases} \tag{2}$$

*Außerdem ist die Gütefunktion $G_\varphi$ monoton wachsend.*

$c = \alpha$-*Fraktil von* $\mathbb{P}_{\vartheta_0} \circ T^{-1}$

$\gamma$ *löst* $G_\varphi(\vartheta_0) = \mathbb{P}_{\vartheta_0}[T > c] + \gamma \mathbb{P}_{\vartheta_0}[T = c] \overset{!}{=} \alpha$

**Beweis.**

Sei $R_{\vartheta':\vartheta}(x) = f_{\vartheta':\vartheta}(T(x))$

mit $f_{\vartheta':\vartheta}(y)$ monoton wachsend in $y$.

$$\theta_0 = (-\infty, \vartheta_0], \theta_1 = (\vartheta_0, \infty\}$$

Zuerst berechnen wir $c$ und $\gamma$. Aus $N - P -$ Lemma konstruieren wir ein $\varphi$ der Form 2 mit

$c = \alpha$-Fraktil von $\mathbb{P}_{\vartheta_0} \circ T^{-1}$ und $\gamma \in [0, 1]$ kommt aus der Gleichung

$$\alpha = \mathbb{P}_{\vartheta_0}[T(x) > c] + \gamma \mathbb{P}_{\vartheta_0}[T(x) = c]$$

Sei $\vartheta < \vartheta'$

Wenn $R_{\vartheta':\vartheta} = f_{\vartheta':\vartheta}(T) > f_{\vartheta':\vartheta}(c)$

$\Rightarrow T > c \Rightarrow \varphi = 1$

Wenn $R_{\vartheta':\vartheta} = f_{\vartheta':\vartheta}(T) < f_{\vartheta':\vartheta}(c)$

$\Rightarrow T < c \Rightarrow \varphi = 0$

$\Rightarrow \varphi$ ist ein N-P-test

von $\vartheta$ gegen $\vartheta'$.

Insbeosndere für $\vartheta = \vartheta_0$ und bel. $\vartheta' > \vartheta_0$ folgt aus $N - P -$ Lemma, dass $\varphi$ ein bester Test für $\vartheta_0$ gegen jedes $\vartheta' \in \theta_1$ zum Niveau $\alpha$.

Noch zu zeigen, dass Niveau von $\varphi$ als Test $\theta_0$ gegen $\theta_1$ $\alpha$ ist, d.h.

$$G_\varphi(\vartheta) \leqslant \alpha \forall \vartheta \in \theta_0$$

Da $G_\varphi(\vartheta_0) = \alpha$ müssen wir zeigen, dass $G_\varphi$ monoton wachsend ist.

Für $\vartheta < \vartheta'$, folt aus dem N-P-Lemma, dass $\varphi$ ein bester Test zu dem Niveau $\beta := G_\varphi(\vartheta)$ ist

$\Rightarrow G_\varphi(\vartheta') \geqslant G_\psi(\vartheta')$ für alle Tests $\psi$ mit Niveau $\beta$

$G_\varphi(\vartheta') = \beta = G_\varphi(\vartheta)$

$\Rightarrow G_\varphi$ ist monoton wachsend. □

**Bemerkung 77.** Für einen rechtsseitigen Test

$H_0 \colon \vartheta \geqslant \vartheta_0$ gegen $H_1 \colon \vartheta < \vartheta_0$

müssen wir nur $<$ und $>$ in $\varphi$ tauschen. D.h. $\vartheta \mapsto -\vartheta$ und $T \mapsto -T$ (Anmerkung von Manuel: Nicht wirklich, $\theta = [0, 1]$ als Gegenbsp.)

**Beispiel 78.** (Einseitiger Gauß-Test, bekannte Varianz)

Sei $(\mathbb{R}^n, \mathcal{B}^n, \mathcal{N}_{m,v}^{\otimes n} \colon m \in \mathbb{R}, v \,\text{fest})$

Zu Testen $H_0 \colon m \leqslant m_0$ gegen $H_1$

Der Ablehnungsbereich ist

$$\left\{M_n > m_0 + \sqrt{v/n}\,\Phi^{-1}(1-\alpha)\right\}$$

Übung!

**Beispiel 79.** (Einseitiger Chi-Quadrat Test (bekannter Erwartungswert)

Seien $X_1, \ldots, X_n \sim \mathcal{N}(m, v)$ m fest und $v > 0$.

Testen: $H_0 \colon v \geqslant v_0$ gegen $H_1 \colon v < v_0$.

Dieses Modell ist exponentiell bzgl. der Statistik

$$T_n(x) = \sum_{k=1}^{n} (x_k - m)^2$$

In der Tat, die Likelihoodfunktion

$$\rho_\vartheta(x) = \exp\left(\underbrace{-\frac{1}{2v}}_{a(v)} T(x) - \frac{n}{2} \ln(2\pi v)\right)$$

$a(v)$ ist wachsend in $v$.

$\Rightarrow R_{v' : v}(x) = \frac{\rho_{v'}(x)}{\rho_v(x)}$ ist wachsend in $T_n(x)$.

$\Rightarrow$Rechtsseitiger Test

$$\varphi(x) = \begin{cases} 1 & T_n(x) < c \\ 0 & T_n(x) > c \end{cases}$$

c berechnen:

$G_\varphi(v_0) \overset{!}{=} \alpha$

$$G_\varphi(v_0) = \mathbb{E}_{v_0}[\varphi] = \mathbb{P}_{v_0}[T_n < c]$$

$$\mathbb{P}_{v_0}[T_n < c] = \underbrace{\mathbb{P}_{\vartheta_0}\left[\frac{T_n}{v_0} < \frac{c}{v_0}\right]}_{X_k \sim \mathcal{N}_{m, v_0}} \overset{!}{=} \alpha$$

$Y_k = \frac{X_k - m}{\sqrt{v_0}} \sim \mathcal{N}_{0,1}$

$\frac{T_n(x)}{v_0} = \sum_{k=1}^n Y_k^2 \sim \chi_n^2$

$\Rightarrow c = v_0 \cdot (\alpha\text{-Quantil von } \chi_n^2)$

$\Rightarrow$ Ablehnungsbreich

$$\{x \colon T_n(x) < v_0(\alpha - \text{Quantl von } \chi_n^2)\}$$

Ende Vorlesung 13

Vorlesung über R mit Hilfe von RStudio.

```
#Programmier-Crashkurs R

#Author: M. Braun


#1 Allgemeines

#1.1 Einrichtung

#Download auf rstudio.com (Open Source)
#Empfehlung: Eigenes .R-Script für jedes Projekt

#1.2 Literatur

#Youtube: Statistik am PC, DataCamp, etc.
#M. Luhmann: R für Einsteiger
#G. Grolemund, H. Wickham: R for Data Science


#2 Grundbefehle

#2.1 Rechenoperationen

7+3
7-3
7*3
7/3
sqrt(7)
exp(7)
sin(7)
floor(7.5)
floor(pi)
?floor

#2.2 Variablen

#Zuweisung
x <- 4
y <- x^2
plot(x,y)
curve(sin(x), from=0, to=2*pi)

#Überschreiben
x <- x+3

#Überblick
ls()

#Löschen
rm(x)
rm(y)
```

```r
rm(list=ls())

#Ausgeben
x <- 2
y <- x^2
print("Hallo Welt")
print(x)
print(paste("Das Quadrat von", x, "ist", y))

#2.3 Schleifen und Bedingungen
i <- 1

#if-Bedingung
#Standard !=, <, <=, >, >=
if(i == 5){
  print(paste("Deine Glückszahl ist", i))
} else{
  print("Nö!")
}

#for-Schleife
for(j in 1:10){
  if(!j %% 2){
    print(paste(j, "ist eine gerade Zahl"))
  } else{
    if(j == 5){
      print("Hallo Welt!")
    } else{
      print("Hi.")
    }
  }
}
rm(j)

#while-Schleife
while(i < 60){
  print(i^3)
  i <- i+1
}
rm(i)

#2.4 Funktionen

#Parameter festlegen
square <- function(a){
  b <- a^2
  #Rückgabe des Wertes
  return(b)
}

square(2)

#Auflistung und Zählen natürlicher Zahlen
list.smaller <- function(p){
  q <- 1
  counter <- 0
```

```
  while(q < p){
    print(paste(q, "ist kleiner als", p))
    q <- q+1
    counter <- counter+1
  }
  print(paste("Insgesamt sind", counter, "nat�rliche Zahlen kleiner als", p))
}

list.smaller(1000)

#BMI berechnen
bmi <- function(weight, height){
  if(height > 5){
    height <- height/100
  }
  #BMI = Gewicht[kg]/Größe[m]
  result <- weight/(height)^2
  return(result)
}

#Funktion aufrufen
bmi(80,180)
bmi(80,1.80)

#Alter berechnen
age <- function(year){
  a <- 2021 - year
  return(a)
}

age(1993)

#Lassen Sie dem PC bei umständlichen Berechnungen Zeit
#Bauen Sie ggf. 'Checkpoints' ein
k <- 6

for(n in 1:10^k){
  if(!n %% 10){
    print(paste("Die Marke", n, "ist erreicht!"))
  }
  print(n)
}


#3 Verwalten von Daten

#3.1 Vektoren

#Vektor zuweisen
vector <- c(1,2,3,4,5,6,7,8,9,10)
vector2 <- c(1:10)
vector
table(vector)
vec <- integer(400)
vect <- (1:200)/2
```

```
#Auf Elemente zurückgreifen
vector[1] + vector[5]
#Bei Tabellen bspw. tabelle[zeile,spalte]

#Rechenoperationen auf alle Elemente anwenden
vector <- log(vector)
vector <- sin(vector)

#Vektoren zusammenfügen
vector <- append(vector, c(1))
vector <- append(vector, c(12:24))
vector <- append(c(12:24), vector)

#Grundfunktionen
sum(vector)
length(vector)
mean(vector)
sd(vector)
max(vector)
min(vector)
median(vector)

#Plotten
hist(vector)
barplot(vector)
plot(vector)
boxplot(vector)

#3.2 Erzeugen von Zufallszahlen

#Gleichverteilung
gleich <- runif(100, -1, 1)
gleich
barplot(gleich)

#Normalverteilung
normal <- rnorm(100, 1, 3)
normal
barplot(normal)
mean(normal)
sd(normal)
hist(normal)

#3.3 Data Frames

#Data Frames können alle möglichen Daten zusammenbündeln
person <- c("Lisa", "Kunibert", "Herbert", "Moritz", "Irmgard")
age <- c(36, 66, 90, 41, 50)
vaccine <- c("Johnson", "Astra", "Astra", "Biontech", "Johnson")
df <- data.frame(person,age, vaccine)
df

#3.4 Tabellen einlesen

scorer <- read.table("scorer.txt")
scorer
```

```
nrow(scorer)
length(scorer)

#Operationen auf Spalten durchführen
with(scorer, mean(V3))
with(scorer, sd(V3))

#Abschließendes Beispiel
data <- read.table("bmi-data.txt")

#Berechnung des Durchschnitts-BMI's separiert nach Geschlechtern
experiment <- function(maximum){
  #Counter für die Geschlechter
  females <- 0
  males <- 0
  #Fehlercounter
  error <- 0
  #Durchschnitts-BMI's initialisieren
  fbmi <- 0
  mbmi <- 0
  #Falls Eingabe zu groß
  if(maximum > nrow(data)){
    print(paste("Es wurden nur", nrow(data), "Personen untersucht."))
    print(paste("Sie haben mit", maximum, "eine zu hohe Zahl eingetippt."))
  } else {
    for(i in 1:maximum){
      if(i <= nrow(data)){
        if(data[i,1] == "w"){
          females <- females + 1
          fbmi <- fbmi + bmi(data[i,2], data[i, 3])
        } else {
          if(data[i,1] == "m"){
            males <- males + 1
            mbmi <- mbmi + bmi(data[i,2], data[i, 3])
          } else {
            error <- error + 1
          }
        }
      }
    }
    print(paste("Es wurden", females, "Frauen und", males, "Männer, also insg-
esamt", nrow(data)-error, "Personen untersucht."))
    fbmi <- fbmi/females
    mbmi <- mbmi/males
    print(paste("Der durchschnittliche weibliche BMI ist", fbmi))
    print(paste("Der durchschnittliche männliche BMI ist", mbmi))
    print(paste("Es sind", error, "Fehler aufgetreten"))
  }
}
experiment(7)
experiment(10)
```

Ende Vorlesung 14

Bemerkungsänderung:

$\alpha$ Quantil und Fraktil

**Bemerkung 80.** (Rechtsseitige beste Tests)

Zu Satz 76 Sei $(\mathcal{X}, \mathcal{F}, \mathbb{P}_\vartheta : \vartheta \in \theta)$ mit $\theta \subseteq \mathbb{R}$ ein Standardmodell mit wachsenden Likelihood-Quotienten bzgl. T und $\vartheta_0 \in \theta$, $0 < \alpha < 1$.

Dann ex. ein bester Test $\varphi$ zu dem Niveau $\alpha$ $(\mathbb{E}_{\vartheta_0}[\varphi] \leqslant \alpha)$ für $H_0 : \vartheta \geqslant \vartheta_0$ gegen $H_1 : \vartheta < \vartheta_0$ der Form

$$\varphi(x) = \begin{cases} 1 & T(x) < c \\ \gamma & T(x) = c \\ 0 & T(x) > c \end{cases} \tag{3}$$

Außerdem ist die Gütefunktion $G_\varphi$ monoton fallend.

$c = \alpha$-Quantil von $\mathbb{P}_{\vartheta_0} \circ T^{-1}$

$\gamma$ löst $G_\varphi(\vartheta_0) = \mathbb{P}_{\vartheta_0}[T > c] + \gamma \mathbb{P}_{\vartheta_0}[T = c] \overset{!}{=} \alpha$

<u>Tests im Gaußmodell</u>:

Jetzt behandeln wir den Fall, wo der Erwartungswert <u>und</u> die Varianz unbekannt sind.

Wir wollen aber wie früher nur eine der beiden Unbekannten testen.

$(\mathbb{R}^n, \mathcal{B}^n, \mathcal{N}_{m,v}^{\otimes n} : m \in \mathbb{R}, v > 0)$

<u>Linksseitiger Chiquadrat-Test für die Varianz</u>

Test für die Varianz

$(V-) : H_0 : v \leqslant v_0$ gegen $H_1 : v > v_0$

mit $v_0 > 0$ und Niveau $\alpha$ gegeben.

$\Rightarrow \theta_0 = \mathbb{R} \times [0, v_0], \theta_1 = \mathbb{R} \times (v_0, \infty)$

**Beispiel 81.** Testen ob ein Messgerät präzise genug ist.

Idee: Wenn $m$ fest wäre, dann würde der Ablehnungsbereich

$$\left\{ \sum_{k=1}^{n} (X_k - m)^2 > v_0 (1 - a) \text{Qauntil von } \mathcal{X}_n^2 \right\}$$

Deshalb wäre natürlich zu raten, dass $m$ mit

$$M = \frac{1}{n} \sum_{k=1}^{n} X_k$$

zu ersetzen. Diesmal hätten wir statt $\chi_n^2$ eine $\chi_{n-1}^2$-Verteilung (Satz 60).

Das ist richtig für den linksseitigen Test, für rechtsseitige Tests braucht man die Bedingung <u>unverfälscht</u>.

<u>Likelihood-Quotienten Test:</u>

Bei Beobachtung von $x$ wählen wir die Alternative wenn der Likelihood-Quotient

$$R(x) = \frac{\sup_{\vartheta \in \theta_1} \rho_\vartheta(x)}{\sup_{\vartheta \in \theta_0} \rho_\vartheta(x)}$$

größer als „c“ ist.

$$\varphi = \begin{cases} 1 & R > c \\ 0 & R < c \end{cases}$$

In unserem Fall:

$$R(x) = \frac{\sup_{m \in R, v > v_0} \phi_{m,v}^{\otimes n}(x)}{\sup_{m \in R, v \leqslant v_0} \phi_{m,v}^{\otimes n}(x)} = \frac{\sup_{m \in R, v > v_0} \frac{1}{(2\pi v)^{n/2}} e^{-\frac{n}{2v} \underbrace{\frac{1}{n}\sum_{k=1}^n (X_k - m)^2}_{\to V}}}{\sup_{m \in R, v \leqslant v_0} \frac{1}{(2\pi v)^{n/2}} e^{-\frac{n}{2v}\frac{1}{n}\sum_{k=1}^n (X_k - m)^2}}$$

$$\text{mit } V = \frac{1}{n} \sum_{k=1}^{n} (X_k - M)^2$$

$$R(x) \overset{m=M}{=\!=\!=} \frac{\sup_{v > v_0} e^{-\frac{n}{2} f(v/V)}}{\sup_{v \leqslant v_0} e^{-\frac{n}{2} f(v/V)}}$$

mit $f = \ln(x) + \frac{1}{x}$

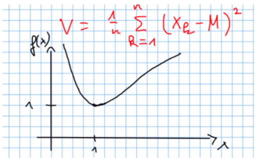



**Abbildung 7.**

Falls $V > v_0$: $\frac{v_0}{V} < 1$

$R(x) = \exp\left(\frac{n}{2}\left(\frac{V}{v_0} - \ln\left(\frac{V}{v_0}\right) - 1\right)\right)$

**Abbildung 8.**

Falls $V < v_0$: $R(x) = \exp\left(-\frac{n}{2}\left(\frac{V}{v_0} - \ln\frac{V}{v_0} - 1\right)\right)$

$\Rightarrow$ R ist eine streng monoton wachsende Funktion in $V$ (und $V^*$)

Dann „anwenden“ von Satz 76 und Satz 60

findet man den Ablehnungsbereich

$$\left\{ \sum_{k=1}^{n} (X_k - M)^2 > v_0(\alpha \,\text{Fraktil von}\, \chi^2_{n-1}) \right\}$$

R (also die Programmiersprache)ist nicht Klausurrelevant.

**Satz 82.** *(Linksseitiger $\chi^2$-Test für die Varianz einer Normalverteilung)*
*Sei das n-fache Gauß'sche Produktmodell*

$$(\mathbb{R}^n, \mathcal{B}^n, \mathcal{N}_{m,v}^{\otimes n}, m \in \mathbb{R}, v > 0)$$

*Dann ist der Test*

$$\varphi(x) = \begin{cases} 1 & \sum_{k=1}^n (X_k - M)^2 > v_0(\alpha \,\text{Fraktil von}\, \chi_{n-1}^2) \\ 0 & \text{sonst} \end{cases}$$

*ist ein <u>bester Test</u> von*

$$H_0\colon v \leqslant v_0 \,\text{gegen}\, H_1\colon v > v_0 \,\text{zum Niveau}\, \alpha$$

***Beweis.***

*Idee: Reduktion zu einem 1-parametrigem Problem!*

*Für ein festes $\vartheta_1 = (m_1, v_1) \in \theta_1$*

*Sei das W-Maß: $v \leqslant v_1$*

$$\widetilde{\mathbb{P}}_v := \int \mathbb{P}_{m,v} \mathrm{dw}_v(m)$$

*Mit $w_v = \mathcal{N}_{m_1, \frac{v_1 - v}{n}}$ (für $v = v_1, w_v = \delta_{m_1}$)*
*Es gilt:*

$$\widetilde{\mathbb{P}}_v \circ M^{-1} = \mathbb{P}_{\vartheta_1} \circ M^{-1}$$

$$\widetilde{\mathbb{P}}_v \circ M^{-1} = \int \mathcal{N}_{m,v/n} d\mathcal{N}_{m_1, \frac{v_1 - v}{n}}(m)$$

$$= \mathcal{N}_{m_1, \frac{v_1 - v}{n}} * \mathcal{N}_{0, v/n} = \mathcal{N}_{m_1, v_1/n} = \mathbb{P}_{\vartheta_1} \circ M^{-1}$$

*Das bedeutet dass man durch Beobachtung des emp. Mittels $\widetilde{\mathbb{P}}_v$ nicht von $\mathbb{P}_{\vartheta_1}$ unterscheiden kann.*
*Die Likelihood-Funktion von $\widetilde{\mathbb{P}}_v$ ist*

$$\tilde{\rho_v}(x) = \int_{\mathbb{R}} \prod_{k=1}^n \phi_{m_1, \frac{v_1 - v}{n}}(m) \mathrm{dm} = c(v) \int_{\mathbb{R}} \underbrace{e^{-\sum_{k=1}^n \frac{(X_k - m)^2}{2v}}}_{e^{-\frac{\mathrm{nV}}{2v} - \frac{n(m - M_v)^2}{2v}}} e^{-\frac{(m - m_1)^2}{2(v_1 - v)/n}} \mathrm{dm}$$

$$= c(v) e^{-\frac{\mathrm{nV}}{2v}} \underbrace{\int_{\mathbb{R}} \exp\left( -\frac{n(m - m_1)^2}{2(v_1 - v)} - \frac{n}{2v}(m - M_n)^2 \right) \mathrm{dm}}$$

$$= \tilde{c}(m_1) \phi_{0, \frac{v_1 - v}{n}} * \phi_{M_n, \frac{v_1}{n}}(m_1)$$

$$= \tilde{\tilde{c}}(m_1) \exp\left( -\frac{(M_n - m_1)^2}{2v_1/n} \right)$$

$$\tilde{\rho_v}(x) = c'(v, m_1) \exp\left( -\frac{n - 1}{2v} V^* - \frac{n(m_1 - M_n)^2}{2v_1} \right)$$

*für* $v \leqslant v_1$.

$\Rightarrow\{\tilde{\mathbb{P}}_v: 0 < v \leqslant v_1\}$ *ist* <u>*exponentielle Familie*</u> *bzgl. der Statistik* $T = V^*$ *mit* <u>*wachsendem Koeffizienten*</u> $a(v) = -\frac{n-1}{2v}$

$\Rightarrow$ *Aus Satz 76 gibt es einen besten Test* $\varphi$ *von* $\{\tilde{\mathbb{P}}_v: v \leqslant v_0\}$ *gegen* $\{\tilde{\mathbb{P}}_{v_1}\}$ *mit Niveau* $\alpha$.

$$\varphi = \mathbb{1}_{\{V^* > c\}} \text{ mit } c \text{ aus Bedingung } (\alpha - \text{Fraktil von } \widetilde{\mathbb{P}_{v_0}} \circ (V^*)^{-1})$$

$$\tilde{G}_\varphi(v_0) \tilde{\mathbb{P}}_{v_0}[V^* > c] \Longrightarrow \alpha$$

$$\forall v \leqslant v_1: \tilde{\mathbb{P}}_{v_0}[V^* > c] = \textstyle\int_{\mathbb{R}} \underbrace{\mathbb{P}_{m,v}[V^* > c]}_{\frac{n-1}{v}V^* \sim \chi^2_{n-1}} \mathrm{dw}_v(m) = \int \chi^2_{n-1}\big(\big(\frac{n-1}{v}c, \infty\big)\big) \mathrm{dw}_v(m) = \chi^2_{n-1}\big(\big(\frac{n-1}{v}c, \infty\big)\big)$$

*Für* $v = v_0$: $c = \frac{v_0}{n-1} \underbrace{\chi^2_{n-1}: 1-\alpha}_{\alpha - \text{Fraktil von } \chi^2_{n-1}}$.

*Für bel.* $\vartheta = (m, v) \in \theta_0. \Rightarrow v \leqslant v_0 \leqslant v_1$

$$G_\varphi(\vartheta) = \chi^2_{n-1}\left(\left(\frac{n-1}{v}c, \infty\right)\right) \leqslant \alpha$$

*weil* $\frac{n-1}{v}c > \frac{n-1}{v_0}c$

$\Rightarrow \varphi$ *ist ein Test von* $\theta_0$ *gegen* $\theta_1$ *vom Niveau* $\alpha$.

<u>*Schließlich:*</u>

$\varphi$ *ist ein bester Test von* $\theta_0$ *gegen* $\theta_1$*:*

*Sei* $\psi$ *ein anderer Test von* $\theta_0$ *gegen* $\theta_1$ *mit Niveau* $\alpha \Rightarrow$ *für* $v \leqslant v_0$

$$\widetilde{G_\psi}(v) = \int \underbrace{G_\psi(m, v)}_{\leqslant \alpha} \mathrm{dw}_v(m) \leqslant \alpha$$

*d.h.* $\psi$ *als Test von* $\{\tilde{\mathbb{P}}_v: v \leqslant v_0\}$ *gegen* $\{\tilde{\mathbb{P}}_{v_1}\}$ *hat Niveau* $\alpha$.

$\varphi$ *ist aber optimal* $\Rightarrow G_\psi(m, v_1) = \widetilde{G_\psi}(v_1) \leqslant \widetilde{G_\varphi}(v_1) = G_\varphi((m_1, v_1))$

$\square$

Ende Vorlesung 15

---

Klausurankündigungen: (unbenotet)

1. Klausur: 03.08.2021 (über Ecampus) open Book Klausur:

online um 08:45 geschrieben von 09:00 bis 11:00.

Upload um 11:15 auf ecampus.

2. Klausur 23.09.2021 gleiche Zeiten.

Wiederholung: von Beweis von Satz 76:

$0 < \alpha < 1$

$H_0\colon \vartheta \leqslant \vartheta_0$ gegen $H_1\colon \vartheta > \vartheta_0$.

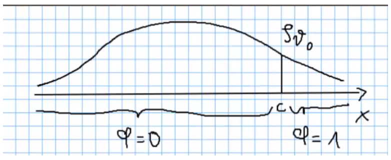



**Abbildung 9.**

$\vartheta < \vartheta'$, T Statistik (typischer Weise Schätzer für Parameter)

$$R_{\vartheta':\vartheta}(x) = f_{\vartheta':\vartheta}(T(x))$$

$f_{\vartheta':\vartheta}$ monoton steigend.

Idee: Konstruiere NP-Test $\varphi$ für $\vartheta_0$ gegen bel. $\vartheta' > \vartheta_0$

$\Rightarrow \varphi(x) = \begin{cases} 1 & R_{\vartheta',\vartheta_0}(x) > c \\ \gamma & R_{\vartheta',\vartheta_0}(x) = c \\ 0 & R_{\vartheta',\vartheta_0}(x) < c \end{cases}$

$\xRightarrow{\text{Monoton}} \varphi(x) = \begin{cases} 1 & T(x) > c' \\ \gamma & T(x) = c \\ 0 & T(x) < c' \end{cases}$ mit $c'$ $\alpha$-Fraktil von $\mathbb{P}_{\vartheta_0} \circ T^{-1}$

und $\alpha = G_\varphi(\vartheta_0) = \mathbb{P}[T > c'] + \vartheta \mathbb{P}[T = c']$

Unabhängig von $\vartheta'$!

$\Rightarrow$NP $-$ Lemma $\varphi$ ist bester Test von $\{\vartheta_0\}$ gegen $\{\vartheta'\}$ für alle $\vartheta' \in \theta_1$-

$$\Rightarrow G_\varphi(\vartheta') \geqslant G_\psi(\vartheta') \forall \vartheta' \in \theta_1$$

und $\psi$ Test mit $\sup_{\vartheta \in \theta_0} G_\psi(\vartheta) \leqslant \alpha$.

Zu zeigen: $\varphi$ hat Niveau $\alpha$.

Sei $\vartheta < \vartheta_0 \Rightarrow \varphi$ ist bester Test ist NP-Test für $\{\vartheta\}$ gegen $\{\vartheta_0\}$

mit Niveau

$$\mathbb{E}_\vartheta[\varphi] = G_\varphi(\vartheta) =: \beta$$

NP-Lemma$\Rightarrow$ $\varphi$ bester Test von $\{\vartheta\}$ gegen $\{\vartheta_0\}$ mit Niveau $\beta$.

$\Rightarrow \alpha = G_\varphi(\vartheta) \geqslant G_\psi(\vartheta_0) \forall \psi\colon G_\psi(\vartheta) \leqslant \beta$

$= \beta$ (konstanter Test $\psi = \beta$)

---

Linksseitiger $\chi^2$-Test für Varianz im Gauß-Produktmodell:

$(\mathbb{R}^n, \mathcal{B}^n, \mathcal{N}_{m,v}^{\otimes n}\colon m \in \mathbb{R}, v > 0)$

$H_0\colon v \leqslant v_0$ gegen $H_1\colon v > v_0$.

$\theta_0 = \mathbb{R} \times (0, v_0], \theta_1 = \mathbb{R} \times (v_0, \infty)$

$\vartheta_1 = (m_1, v_1) \in \theta_1$

$$\tilde{\mathbb{P}}_v := \int \mathbb{P}_{m,v} \mathrm{dw}_v(m), w_v = \mathcal{N}_{m_1, \frac{v_1 - v}{n}}$$

Für $v = v_1 {:}\Rightarrow \tilde{\mathbb{P}}_v = \mathbb{P}_{m_1, v_1} = \mathbb{P}_{\vartheta_1}$

$\Rightarrow \tilde{\rho}_v(x) = \int_{\mathbb{R}} \prod_{k=1}^n \phi_{m,v}(x_k) \phi_{m1, \frac{v_1 - v}{n}}(m) \mathrm{dm}$

$\tilde{\mathbb{P}}_v \circ M_n = \mathbb{P}_{\vartheta_1} \circ M_n = \mathbb{P}_{m_1, v_1} \circ M_n$

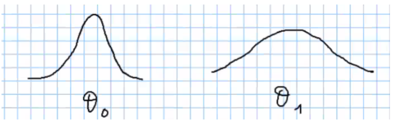



**Abbildung 10.**

Für $T = V^*$ bildet $\{\tilde{\mathbb{P}}_v : 0 \leqslant v \leqslant v_1\}$ eine exponentielle Familie.

Satz 76 $\Rightarrow \varphi = \mathbb{1}_{\{V^* > c\}}$ ist ein bester Test von $\{v \leqslant v_0\}$ gegen $\{v_1\}, \alpha = \tilde{\mathbb{P}}_{v_0}[V^* > c]$.

$\forall v \leqslant v_1 {:}\, \tilde{\mathbb{P}}_v[V^* > c] = \chi^2_{n-1}\big(\big(\frac{n-1}{v} c, \infty\big)\big)$

$\Rightarrow c = \frac{v_0}{n-1} \cdot \chi^2_{n-1:1-\alpha}$

$\varphi$ ist ein Test von $\theta_0$ gegen $\{\vartheta_1\}$ zum Niveau $\alpha$:

$\forall v \leqslant v_0 {:}\, G_\varphi((m, v)) = \chi^2_{n-1}\left(\left(\underbrace{\frac{n-1}{v}}_{\geqslant \frac{n-1}{v_0}} c, \infty\right)\right) \leqslant \alpha.$

$\varphi$ ist ein bester Test von $\theta_0$ gegen $\theta_1$ mit Niveau $\alpha$.

$\psi$ Test von $\theta_0$ gegen $\theta_1$ zu $\alpha$

$\Rightarrow$ für alle $v \leqslant v_0$

$\widetilde{G_\psi}(v) = \int_{\mathbb{R}} G_\psi(m, v) \mathrm{dw}_v(m) \leqslant \alpha$

$\Rightarrow \psi$ hat als Test von $\{\tilde{\mathbb{P}}_v : v \leqslant v_0\}$ gegen $\{\tilde{\mathbb{P}}_{v_1}\}$ Niveau $\alpha$.

$\varphi$ ist optimal

$\Rightarrow G_\psi(\vartheta_1) = \mathbb{E}_{\mathbb{P}_{\vartheta_1}}[\psi] = \mathbb{E}_{\tilde{\mathbb{P}}_{v_1}}[\psi] \leqslant \mathbb{E}_{\tilde{\mathbb{P}}_{v_1}}[\varphi] = \mathbb{E}_{\mathbb{P}_{\vartheta_1}}[\varphi] = G_\varphi(\vartheta_1)$

$\vartheta_1 \in \theta_1$ bel $\Rightarrow$ Behauptung.

---

Rechtsseitiger Chiquadrat-Test für die Varianz

**Satz 83.**

*Sei* $(\mathbb{R}^n, \mathcal{B}^n, \mathcal{N}^{\otimes}_{m,v} : M \in \mathbb{R}, v > 0)$. *Dann ist der Test mit Ablehnungsbereich*

$$\{\Sigma_{k=1}^n (x_k - M_n)^2 < v_0 \chi^2_{n-1:\alpha}\}$$

*ein bester unverfälschter Test von Niveau $\alpha$ von*

$$(V+) H_0 {:}\, v \geqslant v_0 \,\text{gegen}\, H_1 {:}\, v < v_0.$$

Wieso nur unverfälscht?

Wieso nicht mehr gleichmäßig?

Erinnerung:

$$\sup_{(m,v)\in\theta_0} G_\varphi((m,v)) \leqslant \alpha \leqslant \inf_{(m,v)\in\theta_1} G_\varphi((m,v))$$

Für $m\in\mathbb{R}$, sei

$$\varphi_m := \mathbb{1}_{\{\sum(x_k-m)^2\leqslant v_0 c\}}, c=\chi^2_{n:\alpha}$$

Für bel. $(m',v)\in\theta_0$

$\Rightarrow G_{\varphi_m}((m',v)) = \mathbb{E}_{\mathcal{N}^{\otimes n}_{m',v}}[\varphi_m] \leqslant \mathbb{E}_{\mathcal{N}^{\otimes n}_{m,v}}[\varphi_m]$ (Letzter Schritt ist Übung).

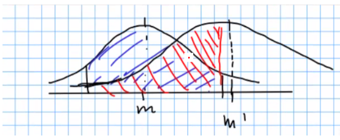



**Abbildung 11.**

$=\chi^2_n\left(\left[0,\frac{v_0 c}{v}\right]\right)\leqslant\alpha$ mit $\frac{v_0}{v}\leqslant 1$.

$\Rightarrow\varphi$ hat auf $\theta_0=\mathbb{R}\times[v_0,\infty)$ Niveau $\alpha$.

$\varphi_m$ ist bester Test von $\{m\}\times[v_0,\infty)$ gegen $\{m\}\times(0,v_0)$

denn $\varphi_m$ unter allen Tests $\psi$ mit $\mathbb{E}_{(m,v_0)}[\psi]\leqslant\alpha$ an allen Stellen $(m,v)$ mit $v<v_0$ die größte Macht

$\Rightarrow\mathbb{E}_{m,v}[\varphi_m]\geqslant\mathbb{E}_{m,v}[\psi]\forall v<v_0$.

Fixiere $v$ und variiere in $m$

$\Rightarrow\varphi_m$ ändert sich

$\Rightarrow$ es gibt keinen (gleichmäßig) besten Test :(

Nachteil von $\varphi_m$:

$$\forall v\colon G_{\varphi_m}((m',v)) = \mathbb{E}_{\mathcal{N}^{\otimes n}_{(m',v)}}[\varphi_m] = \mathcal{N}^{\otimes n}_{(m',v)}\left[\underbrace{\sum_{k=1}^n (x_k-(m-m'))^2}_{\to\infty} < v_0 c\right] \xrightarrow{|m'|\to\infty} 0$$

aus dominierte Konvergenz.

$\Rightarrow$ $\varphi_m$ ist verfälscht!

$\varphi=\mathbb{1}_{\{\sum_{k=1}^n x_k-M\}^2<v_0\chi^2_{n-1:\alpha}}$

ist unverfälscht!

$(m,v)\in\theta_1\quad(v<v_0)$

$$G_\varphi(m,v)=\chi^2_{n-1}\left(\left[0,\frac{v_0}{v}\chi^2_{n-1:\alpha}\right]\right)>\alpha.$$

<u>Einseitiger t-Test für Erwartungswert</u>

Sei $(\mathbb{R}^n,\mathcal{B}^n,\mathcal{N}^{\otimes n}_{m,v}\colon m\in\mathbb{R},v>0)$ gegeben.

$(M-)\colon H_0\colon m\leqslant m_0$ gegen $H_1\colon m>m_0$

$\Rightarrow \theta_0 = (-\infty, m_0] \times \mathbb{R}_+$, $\theta_1 = (m_0, \infty) \times \mathbb{R}_+$

**Beispiel 84.** Minigurken

Welchen Test würden wir aus dem MLP wählen?

$R(x) = \frac{\sup_{m>m_0, v>0} \phi_{m,v}^{\otimes n}(x)}{\sup_{m \leqslant m_0, v>0} \phi_{m,v}^{\otimes n}(x)} = \frac{\sup_{m>m_0} \tilde{V}_m^{-n/2}}{\sup_{m \leqslant m_0} \tilde{V}_m^{-n/2}}$

Wobei $\widetilde{V_m} := \frac{1}{n}\sum_{k=1}^n (X_k - m)^2$. Für gegebenes $m$ ist $\phi_{m,v}^{\otimes n}(x)$ maximal für $v = \frac{1}{n}\sum_{k=1}^n (X_k - m)^2 = \widetilde{V_m}$.

$$\phi_{m,\tilde{V}_m}^{\otimes n}(x) = \frac{1}{\sqrt{2\pi \tilde{V}_m}} e^{-\frac{\sum (X_k - m)^2}{2\tilde{V}_m}} = \text{const} \cdot \tilde{V}_m$$

$$\frac{d}{\text{dm}} \tilde{V}_m = -\frac{2}{n} \sum_{k=1}^n (X_k - m) = -2(M_n - m)$$

$\Rightarrow$

$$R(x) = \begin{cases} (\tilde{V}_{m_0} / V)^{-n/2} & \text{falls } M_n \leqslant m_0 \\ (V / \tilde{V}_{m_0})^{-n/2} & \text{falls } M_n > m_0 \end{cases}$$

Wobei V die emp. Varianz ist.

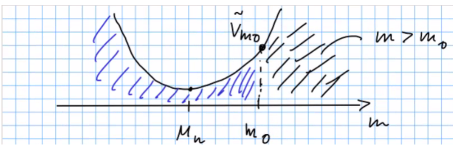



**Abbildung 12.**

Ende Vorlesung 16

---

Aus letzter VL:

$$R(x) = \begin{cases} (\tilde{V}_{m_0} / V)^{-n/2} & \text{falls } M_n \leqslant m_0 \\ (V / \tilde{V}_{m_0})^{-n/2} & \text{falls } M_n > m_0 \end{cases}$$

$\widetilde{V_{m_0}} = \frac{1}{n}\sum_{k=1}^n (X_k - m)^2$

$V = \frac{1}{n}\sum (X_k - M_n)^2$

Verschiebungsformel:

$$\tilde{V}_{m_0} = V + (M - m_0)^2$$

$\Rightarrow \frac{\tilde{V}_{m_0}}{V} = 1 + \frac{(M_n - M)^2}{V}$

Mit $\quad V^*$:

$$T_{m_0} = \frac{\sqrt{n}(M - m_0)}{\sqrt{V^*}} \, (\text{folgt aus Student})$$

$$\frac{\tilde{V}_{m_0}}{V} = 1 + \frac{T_{m_0}^2}{n-1}$$

$\Rightarrow R$ ist strikt wachsend in T.

$\Rightarrow$Satz 76 hat Gestalt $\varphi = \mathbb{1}_{\{T_{m_0} > t\}}$.

Für jedes $\mathbb{P}_{m_0,v}$ hat $T_{m_0}$ die $t_{n-1}$-Verteilung

$\Rightarrow$ Für geg. $\alpha$ wählt man

$$t = t_{n-1:1-\alpha}$$

**Satz 85.**

*Im n-fachen Gauß'schen Produktmodell ist der Test*

$$\varphi(x) = \mathbb{1}_{\left\{\frac{M_n - m_0}{\sqrt{V^*/n}} > t_{n-1:1-\alpha}\right\}}$$

*ein bester unverfälschter Test zum Niveau $\alpha$ von*

$H_0\colon m \leqslant m_0$ *gegen* $H_1\colon m > m_0$.

***Beweis.***

*a) o.B.d.A.* $m_0 = 0$

*Reparametrisieren*

$$\mu = \frac{m\sqrt{n}}{v}, \eta = \frac{1}{2v}$$

$\Rightarrow \rho_{\mu,\eta}(x) = \frac{1}{\sqrt{2\pi v}}\mathrm{n}e^{-\frac{1}{2v}\sum(X_k - m)^2} = \mathrm{const}(m,v)e^{\mu \tilde{M}_n - \eta S}$

*mit* $\widetilde{M_n} = \sqrt{n} M_n$

*und* $S = \sum_{k=1}^n X_k^2$

*Test in* $(\mu, \eta)$:

$H_0$: $\mu \leqslant 0$ *gegen* $H_1$: $\mu > 0$

$T_0 = M_n \sqrt{\frac{n}{V^*}} = \frac{\tilde{M}_n}{\sqrt{S - \tilde{M}_n^2}}\sqrt{n-1}$.

$\Rightarrow \{T_0 > t_{n-1:1-\alpha}\} = \left\{\frac{\tilde{M}_n}{\sqrt{S - \tilde{M}_n^2}} > \frac{t_{n-1:1-\alpha}}{\sqrt{n-1}}\right\} = \{\tilde{M}_n > f(S)\}$

$$f(S) = r\sqrt{\frac{S}{1+r^2}}, r = \frac{t_{n-1:1-\alpha}}{\sqrt{n-1}}.$$

*b)*

$(*)$ *Behauptung: Sei $\psi$ unverfälscht zum Niveau $\alpha$*

$\forall h\colon [0,\infty) \to \mathbb{R}$ *mit* $h(x)e^{-\delta x} \xrightarrow[x\to\infty]{} 0 \forall \delta > 0$.

$$\mathbb{E}_{(m,\eta)}[h(S)(\varphi - \psi)] = 0$$

*Es gilt* $\mathbb{E}_{(m,\eta)}[\psi] = \alpha$ *für alle* $\eta > 0$, *denn*

$$\forall \varepsilon > 0\colon \mathbb{E}_{(0,\eta)}[\psi] \leqslant \alpha \leqslant \mathbb{E}_{(\varepsilon,\eta)}[\psi]$$

*und die Verteilung* $\mathbb{P}_{\mu,\eta}$ *ist stetig in* $\mu$.

$\Rightarrow \mathbb{E}_{(0,\eta)}[\psi] = \alpha$ *für alle* $\eta$.

*Satz von Student* $\Rightarrow \mathbb{E}_{(0,\eta)}[\varphi] = \mathbb{P}_{(0,\eta)}[T_0 > t_{n-1:1-\alpha}] = \alpha$

$\Rightarrow \mathbb{E}_{(0,\eta)}[\varphi - \psi] = 0$ *für alle* $\eta$.

*Da* $\rho_{\mu,\eta} = c(m,v)e^{-\eta S + \mu \tilde{M}_n}$ *gilt* $\forall k \in \{0,1,\dots\}$

$0 = \mathbb{E}_{(0,\eta+k)}[\varphi - \psi] = \mathbb{E}_{(0,\eta)}[e^{-\mathrm{kS}}(\varphi - \psi)]\tilde{c}(v,k)$

$\Rightarrow \mathbb{E}_{(0,\eta)}[g(e^{-S})(\varphi - \psi)] = 0$

*für alle Polynome Polynome* $g$ *(Linearität im Erwartungswert).*

*Weierstraß Approximationssatz:*

$\Rightarrow \mathbb{E}_{(0,\eta)}[g(e^{-S})(\varphi - \psi)] = 0$

*für alle stetigen Funktionen* $g: [0,1] \to \mathbb{R}, 0 \leqslant e^{-S} \leqslant 1$.

*Setzte* $g_\delta(x) = \begin{cases} x^\delta h(\ln(1/x)) & x \in (0,1] \\ 0 & x = 0 \end{cases}$

*für* $\delta > 0$ *und* $h$ *gegebene stetige Funktion*

$h: [0,\infty) \to \mathbb{R}$ *mit* $\lim_{x\to\infty} h(x)e^{-\delta x} = 0 \forall \delta > 0$

$\Rightarrow \mathbb{E}_{(0,\eta)}[e^{-\delta S}h(S)(\varphi - \psi)] = 0 \forall \delta > 0$.

$\Rightarrow \mathbb{E}_{(0,\eta)}[h(S)(\varphi - \psi)] = 0$

$\Rightarrow (*)$, *damit ist die Behauptung bewiesen.*

*c):* $\varphi$ *bester Test:*

*Sei* $(\mu,\eta) \in \theta_1 = (0,\infty)^2$ *gegeben:*

$\Rightarrow$ *Likelihood-Quotient*

$$R_{(\mu,\eta):(0,\eta)}(x) = \frac{\rho_{\mu,\eta}(x)}{\rho_{0,\eta}(x)} = c(m,v)e^{\mu \tilde{M}_n(x)}$$

*Ist eine strikt wachsende Funktion von* $\tilde{M}_n$.

$\Rightarrow \varphi(x) = \mathbb{1}_{\{\tilde{M}_n > f(S)\}} = \mathbb{1}_{\left\{R_{(\mu,\eta):(0,\eta)} > \underbrace{c \cdot e^{\mu f(S)}}_{h(S)}\right\}}$.

$h(x)e^{-\delta x} = e^{\mu\sqrt{x} - \delta x} \xrightarrow[x\to\infty]{} 0 \forall \delta > 0$.

*Aber:*

$$\mathbb{E}_{(\mu,\eta)}[\varphi - \psi] = \int (\varphi - \psi)\rho_{\mu,\eta}(x)\mathrm{dx} = \int (\varphi - \psi)c \cdot e^{\mu \tilde{M}_n(x)}\rho_{0,\eta}(x)\mathrm{dx} = \mathbb{E}_{(0,\eta)}[(\varphi - \psi)R_{(\mu,\eta):(0,\eta)}]$$

$$\overset{(*)}{=\!=} \mathbb{E}_{(0,\eta)}[(\varphi - \psi)(R_{(\mu,\eta):(0,\eta)} - h(S))] \geqslant 0 \,\text{weil}$$

*wenn* $R > h(S) \Rightarrow \varphi = 1 \Rightarrow \varphi - \psi \geqslant 0$

*wenn* $R < h(S) \Rightarrow \varphi = 0 \Rightarrow \varphi - \psi \leqslant 0$. □

**Bemerkung 86.** Analoges gilt für den rechtsseitigen Test!

$H_0: m \geqslant m_0$ gegen $H_1: m < m_0$

$$\varphi(x) = \mathbb{1}_{\left\{\frac{M_n - m_0}{\sqrt{V^*/n}} < t_{n-1:\alpha}\right\}}$$

ist bester unverfälschter Test zum Niveau $\alpha$.

Zweiseitige Tests:

$H_0\colon m = m_0$ gegen $H_1\colon m \neq m_0$

$\Rightarrow \begin{cases} \theta_0 = \{m_0\} \times (0,\infty) \\ \theta_1 = (\mathbb{R} \setminus \{m_0\}) \times (0,\infty) \end{cases}$

**Satz 87.** *(Zweiseitiger t-test)*

*Im n-fachen Produktmodell ist der Test*

$$\varphi(x) = \mathbb{1}_{\left\{|M_n - m_0|\sqrt{\frac{n}{V^*}} > t_{n-1:1-\alpha/2}\right\}}$$

*ein bester unverfälschter Test zum Niveau von*

$$H_0\colon m = m_0 \text{ gegen } H_1\colon m \neq m_0$$

*Beweis (Georgii).*

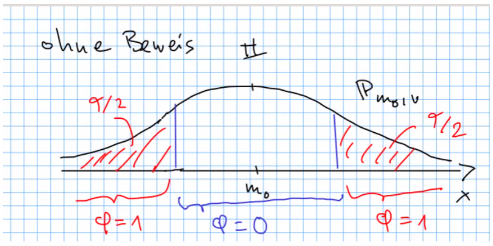



***Abbildung 13.***

Beidseitiger Chiquadrat Test: Blatt 9 Aufgabe 5

---

Anmerkung zu Aufgabe 2 auf Blatt 9:

2- Stichproben Problemm:

$\underbrace{X_1,\ldots,X_k}_{\sim\mathcal{N}(m,v)}$ und $\underbrace{Y_1,\ldots,Y_l}_{\sim\mathcal{N}(m',v)}$ unabhängig.

$H_0\colon m = m'$ gegen $H_1\colon m \neq m'$

Likelihood Quotienten Test:

$\varphi = \mathbb{1}_{\{|T|>c\}}$

mit $T = \sqrt{\frac{1}{1/k+1/l}}\frac{\bar{X}-\bar{Y}}{\sqrt{V^*}} \sim t_{k+l-2}$ und $V^* = \frac{1}{k+l-2}(\sum(x_i-\bar{x})^2 + \sum(y_i-\bar{y})^2)$

---

# 6 $\chi^2$-Anpassungstest

**Beispiel 88.** (Mendels Erbsen 1865)

Beobachtung bei Erbsen:

a) Form

b) Farbe

a) Form $\in$\{rund, kantig\} mit rund <u>dominant</u>

b) Farbe $\in$\{gelb, grü$n$\} mit gelb <u>dominant</u>

d.h. 4 verschiedene Genotypen für die Form:

$\underbrace{(\text{rund}, \text{rund})\,(\text{rund}, \text{kantig})\,(\text{kantig}, \text{rung})}_{\text{Phaenotyp:rund}}$

und $\underbrace{(\text{kantig}, \text{kantig})}_{\text{Phaenotypkantig}}$.

Ähnlich für die Farbe.

Beobachtungsergebnis : $n = 556$

| absolut | gelb | grün |
|---|---|---|
| rund | 315 | 108 |
| kantig | 101 | 32 |

**Tabelle 1.**

Zu testen:

| theoretisch | gelb | grün |
|---|---|---|
| rund | $\frac{9}{16} \approx 0.562$ | $\frac{3}{16} \approx 0.188$ |
| kantig | $\frac{3}{16} \approx 0.188$ | $\frac{1}{16} \approx 0.063$ |

**Tabelle 2.**

vs:

| relativ | gelb | grün |
|---|---|---|
| rund | 0.567 | 0.194 |
| kantig | 0.182 | 0.058 |

**Tabelle 3.**

Dieses Bspo. ist ein besonderer Fall von Folgendem:

Seien $\{1, \ldots, s\}$ die möglichen ausgänge einer Messung und $\rho(k)$ die W'keit Ausgang $k$ zu messen. $\sum \rho(k) = 1$, $\rho(k) > 0$.

Wir wiederholen die Messung n-mal (unabhängig). Dann sind die Häufigkeiten

$$h_n(k) = |\{m \in \{1, \ldots n\}: X_m = k\}|$$

Dann ist die Verteilung

$$\mathbb{P}\left[\bigcap_{k=1}^{s} \{h_n(k) = a_k\}\right] = \binom{n}{a_1, \ldots, a_s} \left(\prod_{k=1}^{s} \rho(k)^{a_k}\right) \mathbb{1}_{\{\sum_{k=1}^{s} a_k = n\}}$$

$\underbrace{(, \ldots, k, \ldots, )}_{n}$ mit $\binom{n}{a_1, \ldots, a_s} = \frac{n!}{\prod_{i=1}^{s} a_i!}$.

Multinomialverteilung: $s = 2 \Rightarrow$ Binomialverteilung.

Ende Vorlesung 17

---

$E = \{1, \ldots, s\}$

Vermutete Verteilung $\{\rho(i)\}_{i=1}^{s}, \sum \rho(i) = 1$, $\rho(i) > 0$.

n-unabhängige Wiederholungen

$h_n(k) := |\{m \in \{1, \ldots, n\} \colon X_m = k\}|$

ist multinomialverteilt.

Multi$(n, \rho(1), \ldots, \rho(s))$

Wir betrachten die Stichprobenfunktion

$$D_{n,\rho}(X_1, \ldots, X_n) = \sum_{k=1}^{n} \frac{(h_n(k) - n\rho(k))^2}{n\rho(k)}$$

Wir sind interessiert an der Verteilung von $D_{n,\rho}$ und werden einen Test einführen der Gestalt:

$$\varphi = \mathbb{1}_{\{D_{n,\rho} > c\}}$$

zum Niveau $\alpha$.

In Mendels Erbsen Beispiel:

1. (gelb,rund) $\rho(1) = 9/16$

2. (gelb,kantig) $\rho(2) = 3/16$

3. (grün,rund) $\rho(3) = 3/16$

4. (grün,kantig) $\rho(4) = 1/16$

Sei $h_n^*(k) = \frac{h_n(k) - n\rho(k)}{\sqrt{n\rho(k)}}$, $k = 1, \ldots, s$.

Dann ist der Vektor $h_n^*$ stets in der Hyperebene

$$H_\rho := \left\{ x \in \mathbb{R}^s \colon \sum_{k=1}^{s} \sqrt{\rho(k)} x_k = 0 \right\}$$

In der Tat

$$\sum_{k=1}^{s} \sqrt{\rho(k)} \frac{h_n(k) - n\rho(k)}{\sqrt{n\rho(k)}} = \frac{1}{\sqrt{n}} \sum_{k=1}^{s} (h_n(k) - n\rho(k)) = \frac{1}{\sqrt{n}} (n - n) = 0.$$

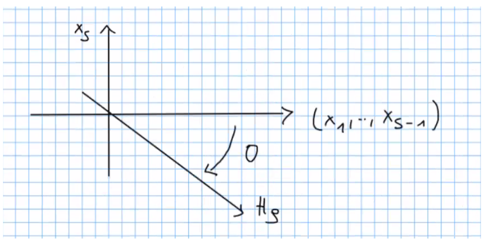



**Abbildung 14.**

Sei $O$ eine orthogonale Matrix mit letzter Spalte $\left(\sqrt{\rho(k)}\right)_{k=1}^{s}$

$$O = \begin{pmatrix} * & * & \sqrt{\rho(1)} \\ * & * & \vdots \\ * & * & \sqrt{\rho(k)} \end{pmatrix}$$

Dann ist $O$ eine Drehung s.d. $O\{x\in\mathbb{R}^s\colon x_s=0\}=H_\rho$ denn für $y\in H_\rho$:

$$\mathrm{O}^T y=\begin{pmatrix} * & * & * \\ * & * & * \\ \sqrt{\rho(1)} & \dots & \sqrt{\rho(k)}\end{pmatrix}\begin{pmatrix} y_1 \\ \vdots \\ y_s\end{pmatrix}$$

$$=\begin{pmatrix} * \\ 0\end{pmatrix}\in\{x\in\mathbb{R}^s\colon x_s=0\}$$

$\Rightarrow O^T H_\rho=\{x\in\mathbb{R}^s\colon x_s=0\}$.

Wir wollen herausfinden welche Verteilung hat $h_n^*$ im Limes. Da $h_n^*\in H_\rho$, suchen wir zunächst die orthogonale Projektion von $x\in\mathbb{R}^s$ auf $H_\rho$. Diesen Punkt nennen wir $\pi_\rho(x)$.

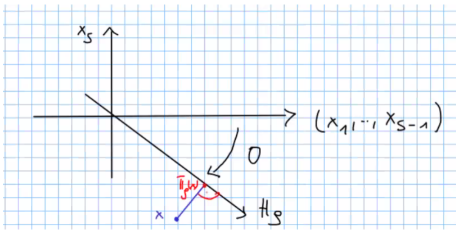



**Abbildung 15.**

$$\Rightarrow\pi_\rho:=\underbrace{O\begin{pmatrix}1 & & & \\ & \ddots & 0 & \\ & 0 & 1 & \\ & & & 0\end{pmatrix}\mathrm{O}^T}$$

denn $\pi_\rho(x)\in H_\rho=O\{x\in\mathbb{R}^s\colon x_s=0\}$

und

$$<x-\pi_\rho(x),H_\rho>=<O^T x-\begin{pmatrix}1 & & & \\ & \ddots & 0 & \\ & 0 & 1 & \\ & & & 0\end{pmatrix}O^T x,\{x_s=0\}>$$

$$=<(0,\dots,(O^T x)_s),\{x_s=0\}>=0$$

**Definition 89.** *Die* *Standartnormalverteilung auf $H_\rho$*

$$\mathcal{N}_\rho:=\mathcal{N}_{0,1}^{\otimes s}\circ\pi_\rho^{-1}$$

*ist die Normalverteilung unter der Projektion $\pi_\rho$.*

**Bemerkung 90.**
$$\underbrace{\left(\underbrace{\mathcal{N}_{0,1}^{\otimes s-1}\otimes\delta_0}_{\mathcal{N}_s\left(0,\begin{pmatrix}1 & & & \\ & \ddots & 0 & \\ & 0 & 1 & \\ & & & 0\end{pmatrix}\right)}\right)\circ\mathrm{O}^T=\mathcal{N}_\rho}_{\overset{\text{Trafo}}{=\!=}\mathcal{N}_s\left(0,\underbrace{O\begin{pmatrix}1 & & & \\ & \ddots & 0 & \\ & 0 & 1 & \\ & & & 0\end{pmatrix}O^T}_{\pi_\rho}\right)}$$

**Satz 91.**

*Es gilt*

$$\lim_{n\to\infty} h_n^* \overset{\mathcal{D}}{=\!=} \mathcal{N}_\rho.$$

*ohne Beweis.*

Daraus erhalten wir folgendes Korollar:

**Folgerung 92.**

$\lim_{n\to\infty} D_{n,\rho} \overset{\mathcal{D}}{=\!=} \chi^2_{s-1}$.

***Beweis.***

$$\mathbb{P}[D_{n,\rho} \leqslant c] = \mathbb{P}\left[\sum_{i=1}^{s} (h_n^*(i))^2 \leqslant c\right]$$

$$= \mathbb{P}[h_n^* \in \{x \in \mathbb{R}^s \|x\|^2 \leqslant c\}]$$

$$\underbrace{\xrightarrow{n\to\infty}}_{\text{Satz 91}} \mathcal{N}_\rho(\{x \in \mathbb{R}^s\colon \|x\|^2 \leqslant c\})$$

$$\overset{\text{geht, weil } O^T \text{ orhtogonal}}{=\!=\!=\!=\!=} (\mathcal{N}_{0,1}^{\otimes s-1} \otimes \delta_0)(\{x \in \mathbb{R}^{s-1}\colon \|x\|^2 \leqslant c\}) = \mathcal{N}_{0,1}^{\otimes s-1}(\{x \in \mathbb{R}^{s-1}\colon \|x\|^2 \leqslant c\})$$

$$\Rightarrow \chi^2_{s-1}([0, c])$$

□

Anwendung: $\chi_2$-Anpassungstest

$E = \{1, \ldots, s\}$

$(E^{\mathbb{N}}, \mathcal{P}(E)^{\mathbb{N}}, \mathbb{P}_\vartheta = \vartheta^{\otimes\mathbb{N}}\colon \vartheta \in \theta)$

$\theta = \{\vartheta = (\vartheta(i))_{i=1}^s\colon \sum_{i=1}^s \vartheta(i) = 1\}$.

**Definition 93.** *Ein Test für das Problem*

*$H_0\colon \vartheta = \rho$ gegen $H_1\colon \vartheta \neq \rho$*

*mit Ablehnungsbereich $\{D_{n,\rho} > c\}$ heißt $\chi^2$-Anpassungstest nach $n$ Beobachtungen.*

*Man setzte $c = \chi^2_{s-1:1-\alpha}$, dann*

$$\chi^2_{s-1}((c, \infty)) = \alpha$$

**Bemerkung 94.**

Faustregel: Die Approximation für den Fall

$$n \geqslant \frac{5}{\min(\rho(i))}$$

ist die ausreichend gut.

Für kleinere $n$: $\mathbb{P}[D_{n,\rho} \leqslant c] = \mathbb{P}[\sum_{i=1}^n (h_n^*(i))^2 \leqslant c]$.

Es gilt:

$$D_{n,\rho} = \sum_{k=1}^{s} \frac{(h_n(k) - n\rho(k))^2}{n\rho(k)} = \frac{1}{n}\sum_{k=1}^{s} h_n(k)^2 / \rho(k) - n.$$

**Beispiel 95.**

$D_{556,\rho} = \frac{16}{556}\left(\frac{315^2}{9} + \frac{108^2 + 101^2}{3} + 32^2\right) - 556 \approx 0.47$

Für $\alpha = 0.1$: $\chi^2_{3:0.9} = 6.3$

$\Rightarrow D_{556,\rho} < 6.3 \Rightarrow \varphi = 0$

$\Rightarrow$ Wir lehnen die Nullhypothese nicht ab.

**Beispiel 96.** Test eines Pseudozufallsgenerators:

$n = 10^4$, Ziffern $\in \{0, \ldots, 9\}$

$\Rightarrow \rho(i) = \frac{1}{10}, i = 0, \ldots, 9.$

| Ziffern | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|---|
| $H$äufigkeiten | 1007 | 987 | 928 | 986 | 1010 | 1029 | 987 | 1006 | 1034 | 1026 |

$\alpha = 5\% \Rightarrow \chi^2_{9,0.95} \approx 17$

und $D_{10^4,\rho} = \frac{1}{104}\sum_{i=0}^{9} \frac{h_n(i)}{1/10} - 10^4 \approx 8.6 < 17.$

Test lehnt die Hypothese der Gleichverteilung nicht ab mit 5% Fehlerniveau.

**Beispiel 97.** Test auf Lottozahlen

n=4854 Ziehungen 6 aus 49.

Sind alle Zahlen gleichverteilt? (Daten aus dem Übungsblatt)

1. Test: 13 ist seltene Zahl: „13 ist eine Unglückszahl“.

   $H_0$: $p \leqslant p_0 = \frac{6}{49}$ gegen $H_1$: $p > p_0$

   $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, X_i = \begin{cases} 1 & \text{falls } 13 \\ 0 & \text{sonst} \end{cases}$

   $\sim \mathcal{N}\left(p, \frac{p(1-p)}{n}\right)$ approximativ.

   $\varphi(x) = \mathbb{1}_{\{\bar{X} > c\}}$ $\mathbb{P}_0[\bar{X} > c] = \alpha \Rightarrow c \approx \frac{6}{49} + \sqrt{\frac{p_0(1-p_0)}{n}}\Phi^{-1}(1-\alpha)$

   $\alpha = 10\% \Rightarrow c \approx 0.129 > \frac{6}{49} \approx 0.122$

   Relative H'keit der 13: $\frac{532}{4854} \approx 0.110.$

   $\Rightarrow$Die Nullhypothese wird akzeptiert.

2. Test.: „13 ist keine Unglückszahl“

   $H_0$: $p \geqslant p_0 = \frac{6}{49}$ gegen $H_1$: $p < p_0$.

   $\Rightarrow \varphi(x) = \mathbb{1}_{\{\bar{X} < c\}}$

   $\mathbb{P}_=[\bar{X} < c] = \alpha$

   $\Rightarrow c = p_0 + \sqrt{\frac{p_0(1-p_0)}{n}}\Phi^{-1}(\alpha)$

Mit $\alpha = 0.1 \Rightarrow c \approx 0.116 > 0.11$

Die Nullhypothese wird abgelehnt.

**Bemerkung 98.** Diese Tests sind künstlich.

Wir haben zuerst bemerkt, dass 13 weniger häufig gezogen wurde und dann haben wir den Test durchgeführt. $\Rightarrow$ Das Ergebnis war von vornherein klar!

3. „Zahlen sind gleichverteilt“

   $H_0$: $p(i) = \frac{1}{49}$ vs $H_1$: $p(i) \neq \frac{1}{49}$.

   $N = 6 \cdot n, n = 4856$

$$D_{n,\rho} = \frac{1}{N} \sum_{k=1}^{49} \frac{h_n(k)^2}{\rho(k)} - N \approx 43$$

   $\chi^2_{48:0.9} = 61$

   $\Rightarrow$Unser Anpassungstest lehnt die Hypothese der Gleichverteilung nicht ab!

Ende Vorlesung 18

---

# 7 $\chi^2$-Test auf Unabhängigkeit

**Beispiel 99.** „Umweltbewusstsein und Bildungsstand“

In einer Umfrage werden $n = 2004$ Leute nach ihrem Umweltbewusstsein und ihrem Bildungsstand gefragt.

| Umwelt und Bildung | 1 | 2 | 3 | 4 | 5 | $\sum$ |
|---|---|---|---|---|---|---|
| Kein | 212 | 434 | 169 | 79 | 45 | 939 |
| Wenig | 85 | 245 | 146 | 93 | 49 | 638 |
| Ziemlich | 38 | 85 | 74 | 56 | 48 | 301 |
| Viel | 20 | 35 | 30 | 21 | 20 | 126 |
| $\sum$ | 355 | 799 | 419 | 249 | 182 | 2004 |

**Tabelle 4.**

Frage: Gibt es eine Korrelation zwischen Ausbildungsgrad und Umweltbewusstsein?

**Bemerkung 100.** Auch wenn die Unabhängigkeit abgelehnt wird, bedeutet das nicht, dass es einen kausalen Zusammenhang gibt. Zum Beispiel könnten beide Faktoren von einem dritten abhängen.

Seien $A = \{1, \ldots, a\}$ und $B = \{1, \ldots, b\}$ die Mengen der beiden Eigenschaften, die auf Unabhängigkeit getestet werden.

Eine Beobachtung ist $X \in E = A \times B$.

Unbekannte Parameter: Wahrscheinlichkeiten der Beobachtungen.

$$\theta = \left\{ \vartheta = (\vartheta_{i,j})_{(i,j) \in E} \in (0,1)^E : \sum_{(i,j) \in E} \vartheta_{i,j} = 1 \right\}$$

Modell: $\mathcal{X}=E^{\mathbb{N}}, \mathbb{P}_\vartheta=\vartheta^{\otimes\mathbb{N}}, \vartheta\in\theta$.

Seien $\vartheta^A=(\vartheta^A(i)=\sum_{j\in B}\vartheta(i,j))_i$ und analog $\vartheta^B=(\vartheta^B(j)=\sum_{i\in A}\vartheta(i,j))_j$.

die beiden Randverteilungen/Marginale von $\vartheta$ auf $A$ bzw. $B$.

Wenn die Eigenschaften unabhängig sind

$\Rightarrow\vartheta=\vartheta^A\otimes\vartheta^B$: $H_0$

$\Rightarrow\theta_0=\{\alpha\otimes\beta=(\alpha(i),\beta(j))_{(i,j)\in E}: \alpha\in\theta^A, \beta\in\theta^B\}$

$\Rightarrow\theta_1=\theta\setminus\theta_0$

Wie testen?

Nach n Beobachtungen $X_1,\ldots,X_n$ erhalten wir Kontingenztafeln

$$h_n(i,j)=|\{k\in 1,\ldots,n\}: X_k=(i,j)|$$

und die empirischen gemeinsamen Verteilungen

$$L_n(i,j)=\frac{h_n(i,j)}{n}.$$

Wir können die empirischen Randverteilungen betrachten

$$L_n^A(i)=\frac{1}{n}h_n^A(i)=\frac{1}{n}|\{k\in 1,\ldots,n\}: X_k=(i,*)|$$

$$L_n^B(j)=\frac{1}{n}h_n^B(j)=\frac{1}{n}|\{k\in 1,\ldots,n\}: X_k=(*,j)|$$

Frage:

Ist $L_n(i,j)$ ungefähr $L_n^A(i)L_n^B(j)$?

Wir betrachten die Funktion

$$\tilde{D}_n=\sum_{(i,j)\in E}\frac{(h_n(i,j)-\mathrm{n}\mathrm{L}_n^A(i)L_n^B(j))^2}{\mathrm{n}\mathrm{L}_n^A(i)L_n^B(j)}=n\left[\sum_{(i,j)\in E}\frac{L_n(i,j)^2}{L_n^A(i)L_n^B(j)}-1\right].$$

**Satz 101.** *Für jedes $\rho=\alpha\otimes\beta$ in der Nullhypothese $\theta_0$ gilt:*

$$\lim_{n\to\infty}\tilde{D}_n=\chi^2_{(a-1)(b-1)}\,\text{bzgl.}\,\mathbb{P}_\rho,$$

*d.h.* $\forall c>0$

$$\lim_{n\to\infty}\mathbb{P}_{\alpha\otimes\beta}[\tilde{D}_n\leqslant c]=\chi^2_{(a-1)(b-1)}([0,c]).$$

Wieso $(a-1)(b-1)$ Freiheitsgrade?

Wieso nicht $\mathrm{ab}-1$?

S

Schätzen von $\alpha$ verbraucht wegen $\sum_{i=1}^{a} \alpha(i) = 1$ $a-1$ Freiheitsgerade.

Schätzen von $\beta$ verbraucht wegen $\sum_{j=1}^{b} \beta(j) = 1$ $b-1$ Freiheitsgerade.

$\Rightarrow \underbrace{\text{ab}-1}_{\text{Gesammtzahl Freiheitsgerade}} - (a-1) - (b-1) = (a-1)(b-1)$

**Folgerung 102.** *($\chi^2$-Test auf Unabhängigkeit)*

*Sei $\alpha \in (0,1)$ ein Irrtumsniveau, $c = 1-\alpha$ Quantil von $\chi^2_{(a-1)(b-1)} = \chi^2_{(a-1)(b-1):(1-\alpha)}$.*

*Dann hat der Test*

*$H_0: \vartheta \in \theta_0$ gegen $H_1: \vartheta \in \theta_1$ mit Ablehnungsbereich $\{\tilde{D}_n > c\}$ das asymptotische Niveau $\alpha$.*

**Beispiel 103.** Fortsetzung: Umwelt und Bildung

Umweltbewusstsein $a = 4$, Ausbildung $b = 5$.

$\Rightarrow$Freiheitsgerade =12.

$\alpha = 1\% \Rightarrow \chi^2_{12:0.99} = 26.22$

$\Rightarrow \varphi(x) = \mathbb{1}_{\{\tilde{D}_n > 26.22\}}$

$\Rightarrow \tilde{D}_n = 125.01$

$\Rightarrow$Die Unabhängigkeit wird abgelehnt!

Das bedeutet nicht, dass es einen kasualen Zusammenhang gibt!

Das soll das nächste Beispiel zeigen:

**Beispiel 104.** (Simpson's Paradox)

Für das Wintersemester 1973 in Berkeley gab es insgesammt 12763 Bewerber. Davon 8442 Männer und 4321 Frauen.

| Geschlecht und Annahme | Angenommen | Abgelehnt |
|---|---|---|
| Männer | 3738 | 4704 |
| Frauen | 1494 | 2827 |

**Tabelle 5.**

$\Rightarrow$ Zulassungsquote $M = 0.44$ und $F = 35\%$.

$\Rightarrow \tilde{D}_n = 114$, Freiheitsgerad=1

$\Rightarrow \alpha = 1\% \Rightarrow \chi^2_{1:0.99} = 6.63$.

$\Rightarrow$ Hypothese der Unabhängigkeit wird abgelehnt.

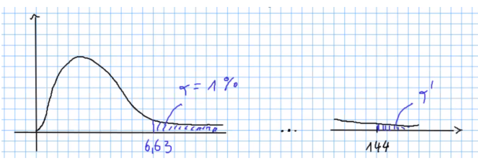



**Abbildung 16.**

$\alpha' < 2.2 \cdot 10^{-16} \Rightarrow$ Hypothese sehr unplausibel.

**Bemerkung 105.** $\alpha'$ ist der p-Wert unserer Beobachtung $x$ (bzw. $\tilde{D}_n(x)$):

D.h. $p(x)$ ist das größte Niveau $\alpha$ bei dem $x$ noch zur Annahme der Nullhypothese führt.

$\varphi = \mathbb{1}_{\left\{\underbrace{T}_{\tilde{D}_n} > c\right\}}$ und $\mathbb{P}_{\vartheta_0}[T \leqslant c] = \underbrace{F(c)}_{\chi^2}$

$\Rightarrow p(x) = 1 - F(T(x))$ (siehe Übungsblatt 10)

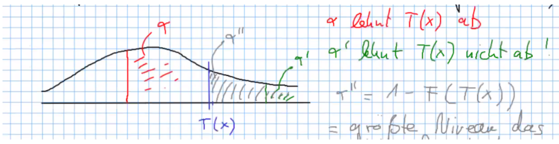



**Abbildung 17.**

=Wkeit der Richtigkeit der Nullhypothese einen „extremeren“ Wert als $T(x)$ zu erhalten.

Frage: Haben die Frauen einen Nachteil?

Wenn man die Zulassungsquoten innerhalb der Departments betrachtet, dann verschwindet die Benachteiligung, da die Frauen sich häufiger bewerben bei Departments mit (absolut) niedriger Aufnahmequote als Männer.

Hypothetisches Beispiel (Die genauen Zahlen sind nicht öffentlich, warum?)

| | Angenommen | Abgelehnt |
|---|---|---|
| M (Technisch) | 200 | 200 |
| F (Technisch) | 200 | 200 |
| M (social science) | 50 | 100 |
| F (social science) | 150 | 300 |
| M ($\sum$) | 250 | 300 |
| F ($\sum$) | 250 | 400 |

**Tabelle 6.**

Technische Bereiche: 50%, Social science: 33%, Insgesammt 45% für Männer und 38% für Frauen.

75% der Frauen bewerben sich bei social science mit 33% Zulassungsquote.

Beweis von Satz 101:

**Beweis.**

Sei $s = a \cdot b, r = (a-1)(b-1)$.

$\rho \in \alpha \otimes \beta \in \theta_0$ fest.

Dann $\tilde{h}_n(i,j) = \frac{h_n(i,j) - \mathrm{n}\mathrm{L}_n^A(i) L_n^B(j)}{\sqrt{\mathrm{n}\mathrm{L}_n^A(i) L_n^B(j)}}$ und $\tilde{D}_n = \|\tilde{h}_n\|^2$

Zu zeigen, $\tilde{D}_n \xrightarrow[n \to \infty]{} \chi_r^2$ bzgl. $\mathbb{P}_\rho$ in Verteilung.

Wegen Satz 91 $\tilde{h}_{n,\rho}^* = \frac{h_n - n\rho}{\sqrt{n\rho}}$ gilt

$$h_{n,\rho}^* \xrightarrow{\mathcal{D}} \mathcal{N}_\rho \,\text{unter}\, \mathbb{P}_\rho.$$

Aber $\tilde{D}_n \neq \|h_{n,\rho}^*\|^2$

Wir zeigen $\tilde{h}_n$ ist ungefähr gleich der Projektion $h_{n,\rho}^*$ auf einen geeigneten Unterraum $L \subseteq \mathbb{R}^s$.

a) Sei $h_n^0(i,j) = \frac{h_n(i,j) - \mathrm{nL}_n^A(i) L_n^B(j)}{\sqrt{n\alpha(i)\beta(j)}}$ und

$$a_l = \left(\sqrt{\alpha(i)}\delta_{\mathrm{jl}}\right)_{(i,j)\in E}, l = 1, \ldots, b$$

$$b_k = \left(\sqrt{\beta(j)}\delta_{\mathrm{ik}}\right)_{(i,j)\in E}, k = 1, \ldots, a$$

□

Ende Vorlesung 19

---

Weiter im Beweis von Satz 101.

**Beweis.** a)

Sei $h_n^0(i,j) = \frac{h_n(i,j) - \mathrm{nL}_n^A(i) L_n^B(j)}{\sqrt{n\alpha(i)\beta(j)}}$ und

$$a_l = \left(\sqrt{\alpha(i)}\delta_{\mathrm{jl}}\right)_{(i,j)\in E}, l = 1, \ldots, b$$

$$b_k = \left(\sqrt{\beta(j)}\delta_{\mathrm{ik}}\right)_{(i,j)\in E}, k = 1, \ldots, a$$

Es gilt $h_n^0 \perp a_l$ und $h_n^0 \perp b_k$.

In der Tat

$$h_n^0 \perp a_l = \sum_{(i,j)\in E} h_n^0(i,j) a_l(i,j) = \sum_{i\in A} h_n^0(i,l)\sqrt{\alpha(i)}$$

$$= \frac{1}{\sqrt{n\beta(l)}} \underbrace{\sum_{i\in A} h_n(i,l)}_{\mathrm{nL}_n^B(l)} - \mathrm{nL}_n^A(i) L_n^B(j) = 0$$

Genau so für $b_k$.

$$L^\perp := \mathrm{span}(\{a_l, b_k \colon l \in B, k \in A\})$$

$$L = \mathrm{orthogonales\,Komplement\,von}\, L^\perp$$

$\Rightarrow h_n^0 \in L$

Dann gilt $\dim(L^\perp) = a + b - 1$.

Denn es gilt

$$\sum_{l\in B} \sqrt{\beta(l)} a_l = \left(\sqrt{\alpha(i)}\sqrt{\beta(j)}\right)_{(i,j)\in E} = \sum_{k\in A} \sqrt{a(k)} b_k$$

$\Rightarrow$ $\{a_l\}$, $\{b_k\}$ sind linear abhängig. $\Rightarrow$ $\dim(L^\perp) \leqslant a + b - 1$.

Weil $a_l \cdot a_{l'} = \delta_{l,l'}$, $b_k \cdot b_{k'} = \delta_{k,k'}$ und $a_l \cdot b_k = \sqrt{\alpha(k)\beta(l)} > 0$

$\Rightarrow a + b - 1$ Vektoren linear unabhängig $\Rightarrow \dim(L^\perp) = a + b - 1$.

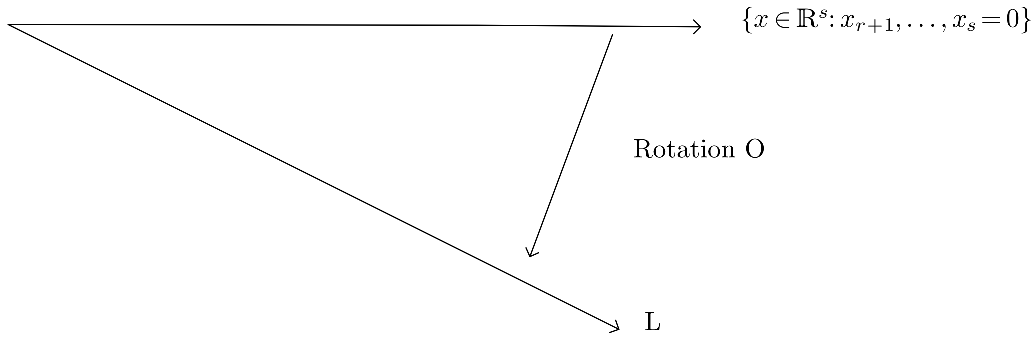
Sei $u_s = \left(\sqrt{\alpha(i)\beta(j)}\right)_{(i,j)\in E}$

Gram Schmidt $\Rightarrow \underbrace{u_{r+1},\ldots,u_s}_{s-r=a+b-1}$ ONB von $L^\perp$

und dann weiter $u_1,\ldots,u_r,u_{r+1},\ldots,u_s$ ONB von $\mathbb{R}^s = L \oplus L^\perp$.

Sei nun

$$O = \begin{pmatrix} u_1 & u_2 & \ldots & u_s \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \end{pmatrix} \text{orthogonal}$$




Und $\Pi = O\begin{pmatrix} 1_r & 0 \\ 0 & 0_{s-r}\end{pmatrix}O^T$.

b)

Als nächstes zeigen wir, dass

$$h_n^0 \approx \underbrace{\Pi \circ h_{n,\alpha\beta}^*}\, f\ddot{u}r\, \text{groß}e n$$

Wir berechnen:

$$\Pi \circ h_{n,\alpha\beta}^*(i,j) = \underbrace{\frac{h_n(i,j) + n\alpha(i)\beta(j) - \alpha(i)\mathrm{n}\mathrm{L}_n^B(j) - \beta(j)\mathrm{n}\mathrm{L}_n^A(i)}{\sqrt{n\alpha(i)\beta(j)}}}_{\in L, \text{denn} \perp \text{zu} (a_l) \text{ und } (b_k), \text{Klammer} - h_{n,\alpha\beta}^*(i,j) \in L^\perp}$$

Daraus folgt dann

$$h_n^0(i,j) = (\Pi \circ h_{n,\alpha\beta}^*(i,j))(i,j) + \eta_n^A(i)\eta_n^B(j)$$

mit $\eta_n^A(i) = \frac{\sqrt[4]{n}}{\sqrt{\alpha(i)}}(L_n^A(i) - \alpha(i))$ und $\eta_n^B(j) = \frac{\sqrt[4]{n}}{\sqrt{\beta(j)}}(L_n^B(j) - \alpha(j))$

$\eta_n^A(i)\eta_n^B(j) = \frac{\sqrt{n}}{\sqrt{\alpha(i)\beta(j)}}(L_n^A(i)L_n^B(j) + \alpha(i)\beta(j) - L_n^B(j)\alpha(i) - L_n^A(i)\beta(j))$

Es gilt jedes $h_n^A(i)$ ist Binomialverteilt $\mathrm{Bin}(n,\alpha_i)$.

$\Rightarrow \mathbb{E}[h_n^A(i)] = n\alpha_i$

$\Rightarrow \mathrm{Var}[h_n^A(i)] = \alpha_i(1-\alpha_i)$

Mit Tschebishev

$$\mathbb{P}_{\alpha\otimes\beta}[|\eta_n^A(i)| \geqslant \varepsilon] \leqslant \frac{\mathrm{Var}[\eta]}{\varepsilon^2} = \frac{1-\alpha(i)}{\sqrt{n}\varepsilon^2}$$

D.h. mit $\mathrm{Var}[\eta] = \frac{\sqrt{n}}{\alpha_i}\frac{1}{n^2} n\alpha_i(1-\alpha_i) = \frac{(1-\alpha(i))}{\sqrt{n}}$

$\Rightarrow |\eta_n^A| \xrightarrow[n\to\infty]{} 0 \, \mathbb{P}_{\alpha\otimes\beta} -$ stochastisch. Genauso für $\eta_n^B$.

$\Rightarrow |h_n^0 - \Pi \circ h_{n,\alpha\beta}^*(i,j)| \xrightarrow[n\to\infty]{} 0 \, \mathbb{P}_{\alpha\otimes\beta} -$ stochastisch

$\Rightarrow |h_n^0 - \Pi \circ h_{n,\alpha\beta}^*(i,j)| \xrightarrow[n\to\infty]{\mathcal{D}} 0$ bzgl. $\mathbb{P}_{\alpha\otimes\beta}$.

c)

Jetzt zeigen wir

$$\widetilde{h_n} \xrightarrow[n\to\infty]{} \left( \underbrace{\mathcal{N}_{0,1}^{\otimes r} \otimes \delta_0^{\otimes s-r}}_{\mathcal{N}_s\left(0, \left(\begin{smallmatrix} 1_r & 0 \\ 0 & 0_{s-r} \end{smallmatrix}\right)\right)} \right) \circ O^T$$

Satz 91 $\Rightarrow h_{n,\alpha\beta}^* \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}_{\alpha\otimes\beta} =: \mathcal{N}_{0,1}^{\otimes s} \circ \Pi^{-1}$

$O^T = \left( \begin{smallmatrix} * & * & * \\ \sqrt{\alpha_1\beta_1} & \dots & \sqrt{\alpha_b\beta_a} \end{smallmatrix} \right)$

$u_s^T$ als letzte Zeile

$\xRightarrow{\Pi\,\text{stetig}} \Pi h_{n,\alpha\beta}^* \to \mathcal{N}_{\alpha\otimes\beta} \circ \Pi^{-1} \xlongequal{\text{Trafo}} \mathcal{N}_s\left(0, \left(\begin{smallmatrix} 1_r & 0 \\ 0 & 0_{s-r} \end{smallmatrix}\right)\right) \circ O^T$

und $O^T\Pi^{-1} = \left(O\left(\begin{smallmatrix} 1_r & 0 \\ 0 & 0_{s-r} \end{smallmatrix}\right)\right)^{-1}$

$\xRightarrow{b} h_n^0 \to \mathcal{N}_{\alpha\otimes\beta} \circ \Pi^{-1}$

$\Rightarrow \tilde{h}_n \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}_{\alpha\otimes\beta} \circ \Pi^{-1}$

, weil $L_n^A(i) \xrightarrow{G\,\mathrm{dgZ}} \alpha(i), \mathbb{P}$ f.s.. genauso für $L_n^B(j)$.

$\Rightarrow \mathbb{P}_{\alpha\otimes\beta}[\tilde{D}_n \leqslant c] = \mathbb{P}_{\alpha\otimes\beta}[\|\widetilde{h_n}\|^2 \leqslant c] \xrightarrow{n\to\infty} \mathcal{N}_{\alpha\otimes\beta} \circ \Pi^{-1}(\{x \in \mathbb{R}^s \colon \|x\|^2 \leqslant c\})$

$$= \underbrace{\mathcal{N}_s\left(0, \begin{pmatrix} 1_r & 0 \\ 0 & 0_{r-s} \end{pmatrix}\right)}_{\mathcal{N}_{0,1}^{\otimes r} \otimes \delta_0^{\otimes s-r}} (\{x \in \mathbb{R}^s \colon \|x\|^2 \leqslant c\}) = \chi_r^2([0,c])$$

□

# 8 Regression

**Beispiel 106.** In einem Versuch misst man die Winkelverformung $\vartheta$ von MgSi-Zylinderblock mit verschiedenen Längen $l$.

Die Verformung wird mit einer konstanten Kraft ausgeübt.

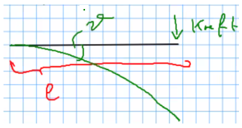



**Abbildung 18.**

Laut Elastizitätstheorie gilt

$$\vartheta = \frac{k \cdot l}{G}$$

mit $k = 4.03 \cdot 10^8 \left[\frac{J}{m^4}\right]$, $(l, \vartheta)$ werden experimentell bestimmt.

G= materialabhängige Konstante

Frage: Was ist der Wert von G unter der Annahme, dass die Messfehler unabhängig und $\mathcal{N}(a, \sigma^2)$-verteilt sind.

Allgemeiner:

Seien $X_1, \ldots, X_n$ Messungen mit Parametern $t_1, \ldots, t_n$.

Wenn man einen linearen Zusammenhang vermutet kann man die $X_k's$ durch eine lineare Regressionsgleichung beschreiben:

$$X_k = \underbrace{\gamma_0 + \gamma_1 + t_k}_{\text{tatsächlicher Zusammenhang}} + \underbrace{\sqrt{v}\,\zeta_k}_{\text{Störung/zufälliger Fehler}}$$

- $\gamma_0$, $\gamma_1$ sind zu ermittelnde unbekannte Parameter
- $v > 0$ ist unbekannte für den Messfehler.
- $\zeta_1, \ldots, \zeta_n$ ZV mit $\mathbb{E}[\zeta_k] = 0$ und $\operatorname{Var}[\zeta_k] = 1$

**Bemerkung 107.** Man kann nicht unterscheiden zwischen $\gamma_0$ und dem Fall $\mathbb{E}[\zeta_k] \neq 0$, aber meistens ist mehr an $\gamma_1$ interessiert.

Ende Vorlesung 20

---

Heute der vorletzte Zettel der Zulassungsrelevant ist.

Lineare Regressionsgleichung

$$X_k = \underbrace{\gamma_0 + \gamma_1 + t_k}_{\text{tatsächlicher Zusammenhang}} + \underbrace{\sqrt{v}\,\zeta_k}_{\text{Störung/zufälliger Fehler}}, k = 1, \ldots, n$$

Vekorielle Form:

$$X = \gamma_0 \mathbb{1} + \gamma_1 \cdot t + \sqrt{v}\,\zeta$$

wobei $X = (X_1, \ldots, X_n)^T$, $t = (t_1, \ldots, t_n)^T$, $\zeta = (\zeta_1, \ldots, \zeta_n)^T$

Für $\gamma = (\gamma_0, \gamma_1)$ sei $\mathbb{P}_{\gamma, v}$ die Verteilung $\gamma_0 \mathbb{1} + \gamma_1 \cdot t + \sqrt{v}\,\zeta$

$\Rightarrow (\mathbb{R}^n, \mathcal{B}^n, \mathbb{P}_{\gamma, v}: (\gamma, v) \in \mathbb{R}^2 \times (0, \infty))$.

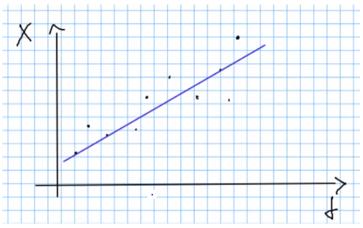



**Abbildung 19.**

Frage: Wie findet man die Parameter $(\hat{\gamma}_0, \hat{\gamma}_1)$ s.d. die Gerade $\hat{\gamma}_0 + \hat{\gamma}_1 t$ „am besten" in den Messpunkten liegt.

Prinzip der kleinsten Quadrate

$\hat{\gamma} := (\hat{\gamma}_0, \hat{\gamma}_1)$ wird so bestimmt, dass der mittlere quadratische Fehler

$$F_\gamma := \frac{1}{n}\sum_{k=1}^{n} (X_k - (\hat{\gamma}_0 + \hat{\gamma}_1 t_k))^2 = \frac{|X - \gamma_0 \mathbb{1} - \gamma_1 \cdot t|^2}{n}$$

für $\gamma = \hat{\gamma}$ minimal ist.

$\hat{\gamma}$ bestimmen: Für $|\hat{\gamma}_0|$, $|\hat{\gamma}_1|$ groß wird der Fehler groß.

$$\Rightarrow \begin{cases} 0 = \frac{\partial}{\partial \gamma_0} F_\gamma = -\frac{2}{n}\sum_{k=1}^{n} (X_k - (\hat{\gamma}_0 + \hat{\gamma}_1 t_k)) \\ 0 = \frac{\partial}{\partial \gamma_0} F_\gamma = -\frac{2}{n}\sum_{k=1}^{n} t_k (X_k - (\hat{\gamma}_0 + \hat{\gamma}_1 t_k)) \end{cases}$$

$$\Rightarrow \text{Normalgleichungen:} \begin{cases} M(X) = \hat{\gamma}_0 + \hat{\gamma}_1 M(t) \\ \hat{\gamma}_0 M(t) + \hat{\gamma}_1 \frac{1}{n}\sum_{k=1}^{n} t_k^2 = \frac{1}{n}\sum_{k=1}^{n} t_k X_k \end{cases}$$

Mittelwert für $x \in \mathbb{R}^n M(x) = \frac{1}{n}\sum_{k=1}^{n} x_k$

Varianz $V(t) = \frac{1}{n}\sum_{k=1}^{n} t_k^2 - M(t)^2$ Bonus: In VL 25 nachgefragt, ist das gleiche wie $\mathrm{Var}(t)$

Kovarianz: $c(t, X) = \frac{1}{n}\sum_{k=1}^{n} t_k X_k - M(t) M(X)$

$\Rightarrow \hat{\gamma}_0 M(t) + \hat{\gamma}_1 \frac{1}{n}\sum_{k=1}^{n} t_k^2 = \hat{\gamma}_0 M(t) + \hat{\gamma}_1 (V(t) + M(t)^2) = c(t, X) + M(t) M(X)$

$$\Rightarrow \underbrace{\begin{pmatrix} 1 & M(t) \\ M(t) & \mathrm{Var}(t) + M(t)^2 \end{pmatrix}}_{Q} \begin{pmatrix} \hat{\gamma}_0 \\ \hat{\gamma}_1 \end{pmatrix} = \begin{pmatrix} M(X) \\ c(t, X) + M(t) M(X) \end{pmatrix}$$

$\det Q = \mathrm{Var}(t) > 0$, wenn $|\{t_1, \ldots, t_n\}| \geqslant 2$

$$Q^{-1} = \frac{1}{\mathrm{Var}(t)} \begin{pmatrix} \mathrm{Var}(t) + M(t)^2 & -M(t) \\ -M(t) & 1 \end{pmatrix}$$

$$\Rightarrow \begin{pmatrix} \hat{\gamma}_0 \\ \hat{\gamma}_1 \end{pmatrix} = \begin{pmatrix} \frac{1}{\mathrm{Var}(t)}\{(\mathrm{Var}(t) + M(t)^2 - M(t))(c(t, X) + M(t) M(X))\} \\ \frac{1}{\mathrm{Var}(t)}\{-M(t) M(X) + c(t, X) + M(t) M(X)\} \end{pmatrix}$$

**Satz 108.**

$\hat{\gamma}_0 := M(X) - \frac{M(t)}{\mathrm{Var}(t)} c(t, X)$

$\hat{\gamma}_1 := \frac{c(t, X)}{\mathrm{Var}(t)}$

*sind die eindeutig bestimmten kleinste-Quadrate-Schätzer für die Parameter $\gamma_0$ und $\gamma_1$. Sie sind erwartungstreu.*

***Beweis.***

$$\mathbb{E}[\hat{\gamma_0}] = \mathbb{E}\left[\frac{1}{n}\sum_{k=1}^{n} X_k - \frac{M(t)}{\text{Var}(t)} \cdot \frac{1}{n}\sum_{k=1}^{n} t_k X_k + \frac{M(t)^2}{\text{Var}(t)}\frac{1}{n}\sum_{k=1}^{n} X_k\right]$$

$E[X_k] = \gamma_0 + \gamma_1 t_k + 0$

$$\Rightarrow \mathbb{E}[\hat{\gamma_0}] = \frac{1}{n}\sum_{k=1}^{n} (\gamma_0 + \gamma_1 t_k)\left(1 - \frac{M(t)}{\text{Var}(t)} t_k + \frac{M(t)^2}{\text{Var}(t)}\right)$$

$$= \gamma_0 + \gamma_1 \underbrace{\left(M(t) - \frac{M(t)}{\text{Var}(t)} \underbrace{M(t^2)}_{\text{Var}(t)+M(t)^2} + \frac{M(t)^2}{\text{Var}(t)} M(t)\right)}_{=0} = \gamma_0$$

$$\mathbb{E}[\hat{\gamma_1}] = \frac{1}{\text{Var}(t)}\mathbb{E}\left[\frac{1}{n}\sum_{k=1}^{n} X_k t_k - M(t)\frac{1}{n}\sum_{k=1}^{n} X_k\right]$$

$$= \frac{1}{\text{Var}(t)}\frac{1}{n}\sum_{k=1}^{n} (\gamma_0 t_k + \gamma_1 t_k^2 - M(t)\gamma_0 - M(t)\gamma_1 t_k) = \frac{1}{\text{Var}(t)}\gamma_1 (M(t^2) - M(t)^2) = \gamma_1 \qquad \square$$

**Bemerkung 109.**

Sei $L := \{\gamma_0 \mathbb{1} + \gamma_1 t | \gamma_0, \gamma_1 \in \mathbb{R}\} \subseteq \mathbb{R}^n$

Wenn $\Pi_L$ die Orthogonale Projektion auf L ist, dann gilt:

$$\Pi_L X = \hat{\gamma_0}\mathbb{1} + \hat{\gamma_1} t.$$

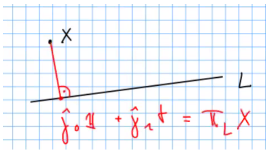



**Abbildung 20.**

Nach dem Prinzip der kleinsten Quadrate hat der Vektor $\hat{\gamma_0}\mathbb{1} + \hat{\gamma_1} t \in L$ den kleinsten euklidischen Abstand zum Punkt X, *d.h.* $X - \hat{\gamma_0}\mathbb{1} - \hat{\gamma_1} t \perp$L.

Was gewinnt man? Man kann Vorhersagen machen über $X$ bei noch nicht gemessenen Werten t.

Aber, aufgepasst: Wenn das richtige Verhältnis zwischen $X$ und $t$ nur approximativ ist, wird eine Extrapolierung für t weit außerhalb problematisch.

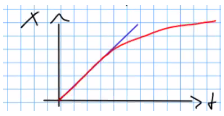



**Abbildung 21.**

Außerdem: Es kann weitere wichtige Einflussfaktoren außer $t$ geben! (Siehe Simpson's Paradox)

(Aufgabe 2 Blatt 11!)

**Beispiel 110.** Fortsetzung Beispiel:

Folgendes Ergebnisse haben wir gemessen:

| l[mm] | $\vartheta[10^{-3}\text{rad}]$ |
|---|---|
| 390 | 9.7 |
| 370 | 9.05 |
| 350 | 9 |
| 330 | 8.5 |
| 310 | 8.2 |
| 290 | 7.55 |
| 260 | 6.25 |
| 230 | 6.1 |

**Tabelle 7.**

$\underbrace{\vartheta_i}_{X_i} = \underbrace{\frac{k}{G}}_{\gamma_1} l_i + \underbrace{a}_{\gamma_0} + \underbrace{x_i}_{v\zeta_i}$ $(x_i \sim \mathcal{N}(0, \sigma^2))$

Schätzung für $\gamma_1$ ist

$$\hat{\gamma_1} = \frac{c(l, V)}{\mathrm{Var}(l)} = \frac{\frac{1}{8}\sum l_i\vartheta_i - \frac{1}{8^2}\sum l_i \sum \vartheta_i}{\frac{1}{8}\sum l_i^2 - \frac{1}{8^2}(\sum l_i)^2} = \frac{\sum l_i\vartheta_i - \frac{1}{8}\sum l_i \sum \vartheta_i}{\sum l_i^2 - \frac{1}{8}(\sum l_i)^2} = 2.36 \cdot 10^{-2}$$

$\Rightarrow G = \frac{k}{\hat{\gamma_1}} = \frac{4.038}{2.36} 10^{10} = 1.71 \cdot 10^{10} \left[\frac{J}{m^3}\right]$

Wert aus Physiktafeln: $17.3 \cdot 10^{10}$.

Absoluter Unterschied $= 2 \cdot 10^8$

Relativer Unterschied $= \frac{2 \cdot 10^8}{1.73 \cdot 10^{10}} = 1.1\%$

Das lineare Modell:

**Definition 111.** *Seien $1 \leqslant s < n \in \mathbb{N}$. Ein lineares Modell für m reellwertige Beobachtungen mit s unbekannten Parametern $\gamma = (\gamma_1, \ldots, \gamma_s) \in \mathbb{R}^s$ und mit einem unbekannten Parameter $v > 0$ besteht aus dem folgenden:*

*i.* $\underbrace{n \times s \,\text{Matrix}\, A}_{\text{Designmatrix}}$, $\mathrm{rang}(A) = s$

*ii. Zufallsvektor* $\underbrace{\zeta = (\zeta_1, \ldots, \zeta_n)^T}_{\text{Fehlergröße}}$ *standardisierte Zufallsvariablen*

$X = A\gamma + \sqrt{v}\zeta$ *Beobachtungsvektor*

Modell: $(\mathbb{R}^n, \mathcal{B}^n, \mathbb{P}_{\gamma,v} : \gamma \in \mathbb{R}^s, v > 0)$ *mit* $\mathbb{P}_{\gamma,v}$ *der Verteilung von* $A\gamma + \sqrt{v}\zeta$

Ziel ist: Aufgrund einer Realisierung $(X_1, \ldots, X_n)$ wollen wir $\gamma$ schätzen oder Hypothesen aufstellen und testen.

**Beispiel 112.**

1. $s = 1$, $A = \mathbb{1}$ $\gamma = m \in \mathbb{R}$, $\zeta_k \sim \mathcal{N}(0, 1)$ u.i.v.

   $\Rightarrow \mathbb{P}_{\gamma,v} = \mathcal{N}_{m,v}^{\otimes n} \Rightarrow$ Gauß'sches Produktmodell mit unbekanntem Erwartungswert und unbekannter Varianz.

2. $s = 2, A = \begin{pmatrix} 1 & t_1 \\ \vdots & \vdots \\ 1 & t_n \end{pmatrix}, \gamma = (\gamma_0, \gamma_1)^T$

   Wir brauchen mindestens zwei verschiedene Werte für die ts. sonst $\mathrm{rang}(A) = 1 < s = 2$.

Das ist die einfache lineare Regression aus dem Beispiel.

3. Polynomielle Regression

   $t = (t_1, \ldots, t_n)^T$ mind. $d+1$ verschiedene

   $s = d+1, \gamma = (\gamma_0, \ldots, \gamma_s)^T$ und

$$A = \begin{pmatrix} 1 & t_1 & t_1^2 & \ldots & t_1^d \\ \vdots & \vdots & \vdots & \ldots & \vdots \\ 1 & t_n & t_n^2 & \ldots & t_n^d \end{pmatrix}$$

   $\Rightarrow X_k = \gamma_0 + \gamma_1 t_k + \gamma_2 t_k^2 + \cdots + \gamma_d t_k^d + \sqrt{v}\zeta_k.$

4. Mehrfache lineare Regression:

   mehrere Einflussgrößen

   $s = d+1$

$$A = \begin{pmatrix} 1 & t_{1,1} & \ldots & t_{1,d} \\ \vdots & \vdots & & \vdots \\ 1 & t_{n,1} & \ldots. & t_{n,d} \end{pmatrix} \text{mit Rang } s$$

   $X_k = \gamma_0 + \gamma_1 t_{k,1} + \cdots + \gamma_d t_{k,d} + \sqrt{v}\zeta_k, 1 \leqslant k \leqslant n$

   d Anzahl der Einflussgrößen

   $t_{k,i}$ Wert der i-ten Einflussgröße beim $k-$ten Experiment.

Ende Vorlesung 21

---

$X = \underbrace{A}_{\text{Designmatrix} \mathbb{R}^{n \times s} \text{mit} \quad \text{rang}(A) = s, s < n} \gamma + \sqrt{v} \underbrace{\zeta}_{\mathbb{E}[\zeta_i] = 0, \text{Var}[\zeta_i] = 1}$

Beobachtungsvektor.

Modell: $(\mathbb{R}^n, \mathcal{B}^n, \mathbb{P}_{\gamma,v}: \gamma \in \mathbb{R}^s, v > 0)$

mit $\mathbb{P}_{\gamma,v}$ die Verteilung von $A\gamma + \sqrt{v}\zeta$.

Prinzip der kleinstesn Quadrate:

Sei $L := \{A\gamma | \gamma \in \mathbb{R}^s\} \subseteq \mathbb{R}^n$.

Wir suchen ein Element von $L$ mit kleinstem Abstand zum Beobachtungsvektor.

Sei $\Pi_L$ die Orthogonale Projektion auf $L$.

$\Rightarrow \Pi_L x \in L,\ |x - \Pi_L x| = \min_{u \in L} |x - u| \forall x \in \mathbb{R}^n$.

bzw

$$x - \Pi_L \perp L \forall x \in \mathbb{R}^n$$

**Bemerkung 113.**

- Die $s \times s$ Matrix $A^T A$ ist invertierbar.
- $\Pi_L = A(A^T A)^{-1} A^T$ (Bonus: Moore-Penrose-Pseudoinverse)und $\hat{\gamma} := (A^T A)^{-1} A^T x$ ist die einzige Lösung von $\Pi_L x = A\hat{\gamma}$

**Beweis.** Wenn für einen Vektor $c \in \mathbb{R}^s, c \neq 0$ gilt $A^T \mathrm{Ac} = 0$ gilt

$|\mathrm{Ac}|^2 = c^T A^T \mathrm{Ac} = 0 \Rightarrow \mathrm{Ac} = 0$, aber das ist ein Widerspruch zu $\mathrm{rang}(A) = s \Rightarrow (A^T A)^{-1}$ existiert.

$$\forall x \in \mathbb{R}^n \colon \underbrace{A(A^T A)^{-1} A^T x}_{y \in \mathbb{R}^s} \in L$$

$$A^T \left( x - \underbrace{A(A^T A)^{-1} A^T x}_{\Pi_L} \right) = A^T x - A^T x = 0$$

Daraus folgt, dass $\Pi_L = A^T (A^T A)^{-1} A^T$, denn

$\Rightarrow \forall u \in L \colon (u = \mathrm{Ay})\ (x - \Pi_L x) \cdot u = 0 \Leftrightarrow A^T (x - \Pi_L x) y = 0$

□

**Satz 114.** *(Schätzer im linearen Modell)*

*Im linearen Modell mit Fehlergrößen $\zeta_1, \ldots, \zeta_n$ unkorreliert $(\mathbb{E}[\zeta_i] = 0, \mathrm{Var}[\zeta_i] = 1)$ gilt*

i. *Der kleinste Quadrate Schätzer*

$$\hat{\gamma} := (A^T A)^{-1} A^T x$$

*ist ein erwartungstreuer Schätzer für $\gamma$.*

ii. *Satz von Gauß-Markov: Sei $\tau \colon \mathbb{R}^s \to \mathbb{R}$ eine lineare zu schätzende Kenngröße, d.h.*

$$\tau(\gamma) = c \cdot \gamma \forall \gamma \in \mathbb{R}^s$$

*für ein $c \in \mathbb{R}^s$. Dann ist $T = c \cdot \hat{\gamma}$ ein erwartungstreuer Schätzer für $\tau$, der als einziger unter allen linearen erwartungstreuen Schätzern von $\tau$ die kleinste Varianz hat.*

iii. *Die (korrigierte) Stichprobenvarianz*

$$V^* := \frac{|X - \Pi_L X|^2}{n - s} = \frac{|X|^2 - |\Pi_L X|^2}{n - s} = \frac{|X - A\hat{\gamma}|^2}{n - s}$$

*ist ein erwartungstreuer Schätzer für $v$.*

***Beweis.***

*(i) Wegen Linearität*

$$\mathbb{E}[\hat{\gamma}] = (A^T A)^{-1} A^T \underbrace{\mathbb{E}[X]}_{A \cdot \gamma} = \gamma$$

*(ii)*

*$T = c\hat{\gamma}$ ist erwartungstreu wegen (i) und Linearität.*

*Sei $a = A(A^T A)^{-1} c \in L$*

$\Rightarrow A^T a = c \Rightarrow c^T = a^T A \Rightarrow \tau(\gamma) = c^T \gamma = a^T A \gamma$

$T = c^T \hat{\gamma} = c^T (A^T A)^{-1} A^T X = a^T \underbrace{A(A^T A)^{-1} A^T}_{\Pi_L} X = a^T X + \underbrace{a^T \left( \underbrace{\Pi_L X - X}_{\in L^\perp} \right)}_{=0} = a^T X.$

*Sei $S$ ein weiterer linearer erwartungstreuer Schätzer für $\tau$.*

$\Rightarrow S = b \cdot X$ *für ein* $b \in \mathbb{R}^n$.

*Dann* $b^T A\gamma = \mathbb{E}[b^T X] = \mathbb{E}[S] = \tau(\gamma) \overset{\text{erwartungstreu}}{=\!=\!=} a^T A\gamma$

$\Rightarrow b \cdot u = a \cdot u \forall u \in L$

$\Rightarrow b - a \in L^{\perp}$, *d.h.*

$\Rightarrow a = \Pi_L b$ *und deshalb* $|a|^2 \leqslant |b|^2$.

$$\mathrm{Var}[S] - \mathrm{Var}[T] = \mathbb{E}\left[\left(b^T \underbrace{(X - A\gamma)}_{\sqrt{v}\zeta}\right)^2\right] - \mathbb{E}\left[\left(a^T\left(\underbrace{X - A\gamma}_{\sqrt{v}\zeta}\right)\right)^2\right]$$

$$= v\left(b^T \quad \underbrace{\mathbb{E}[\zeta \cdot \zeta^T]}_{E_n(\text{unkoreliertheit})} \quad b - a^T \mathbb{E}[\zeta \cdot \zeta^T] a\right) = v(|b^2| - |a|^2) \geqslant 0.$$

*(iii) Aus Pythagoras folgt* $|x - \Pi_L x|^2 + |\Pi_L x|^2 = |x|^2$. *Zusammen mit* $\Pi_L x = A\hat{\gamma}$ *erhalten wir die Gleichheiten.*

*Zu zeigen: Erwartungstreue*

$$\Pi_L = O \cdot E_s \cdot O^T, \text{wobei}$$

$O = \left(\begin{array}{c|c|c} u_1 & \cdots & u_n \end{array}\right)$ *orthogonale Matrix mit* $L = \mathrm{span}\{u_1, \ldots, u_s\}$ *und* $u_1, \ldots, u_n$ *ONB von* $\mathbb{R}^n$.

$E_s = \begin{pmatrix} \mathbb{1}_s & 0 \\ 0 & 0_{n-s} \end{pmatrix}$.

$$(n-s)V^* = \left|\underbrace{A\gamma + \sqrt{v}\zeta}_{X} - \Pi_L\left(\underbrace{A\gamma + \sqrt{v}\zeta}_{X}\right)\right|^2 \overset{\Pi_L A\gamma = A\gamma}{=\!=\!=\!=} v|\zeta - \Pi_L \zeta|^2$$

$$= v\left|O\left(\underbrace{O^T\zeta}_{\eta} - E_s \underbrace{O^T\zeta}_{\eta}\right)\right|^2 = v|\eta - E_s\eta|^2 = v \sum_{k=s+1}^{n} \eta_k^2$$

$$\mathbb{E}[\eta_k^2] = \sum_{i,j=1}^{n} \underbrace{O_{i,k} O_{j,k}}_{\delta_{i,k}} \underbrace{\mathbb{E}[\zeta_i, \zeta_j]}_{\delta_{i,j}} = 1$$

$\Rightarrow \mathbb{E}[(n-s)V^*] = v \sum_{k=s+1}^{n} \mathbb{E}[\eta_k^2] = v(n-s)$.

□

<u>Das lineare Gaußmodell:</u>

$\zeta \sim \mathcal{N}_n(0, E) = \mathcal{N}_{0,1}^{\otimes n} \Rightarrow \mathbb{P}_{\gamma,v} \overset{\text{Trafo.}}{=\!=\!=} \mathcal{N}_n(A\gamma, \mathrm{vE})$.

Insbesondere sind $X_1, \ldots, X_n$ unabhängig.

$\Rightarrow$ <u>Lineares Gaußmodell</u>

Ziel:

- Konfidenzbereiche angeben
- Hypothesen testen

**Satz 115.** *(Verallgemeinerung von Satz von Student)*

*Im linearen Gaußmodell gelten bzgl.* $\mathbb{P}_{\gamma,v}$

i. $\hat{\gamma}$ *ist* $\mathcal{N}_s(\gamma, v(A^TA)^{-1})$*-verteilt*

ii. $\frac{n-s}{v}V^* \sim \chi^2_{n-s}$

iii. $\frac{1}{v}|A(\hat{\gamma}-\gamma)|^2 = \frac{1}{v}|\Pi_L X - \mathbb{E}[X]|^2 \sim \chi^2_s$ *und ist unabhängig von* $V^*$.

   *Es folgt* $\frac{|A(\hat{\gamma}-\gamma)|^2}{\mathrm{s}\mathrm{V}^*} \sim \mathcal{F}_{s,n-s}$

   *(Vgl. Satz von Student:* $T = \frac{\sqrt{n}(M-m)}{\sqrt{V^*}}$*)*

iv. *Sei* $H \subset L$ *ein linearer Teilraum mit* $\dim H = r < s$ *und* $A\gamma \in H$.

   *Dann gilt:* $\frac{1}{v}|\Pi_L X - \Pi_H X|^2 \sim \chi^2_{s-r}$ *und ist unabhängig von* $V^*$.

   *Insbesondere*

$$F_{H,L} = \frac{n-s}{s-r}\frac{|\Pi_L X - \Pi_H X|^2}{|X - \Pi_L X|^2} = \frac{|A\hat{\gamma} - \Pi_H X|^2}{(s-r)V^*} \sim \mathcal{F}_{s-r,n-s}.$$

***Beweis.***

*(i) Erinnerung:* $Z = \mathrm{QY} + a$ $Q$: $k \times n$ *Matrix mit rang* $k < n$.

$Y \sim \mathcal{N}_n(m, C) \Rightarrow Z \sim \mathcal{N}_k(\mathrm{Qm} + a, \mathrm{QCQ}^T)$ *(Satz 9.5 Georgii)*

$X \sim \mathcal{N}_n(A\gamma, \mathrm{vE}_n) \Rightarrow \hat{\gamma} = \underbrace{(A^TA)^{-1}A^T}_{Q: n \times s}X$

*ist normalverteilt mit* $\mathbb{E}[\hat{\gamma}] = \gamma$ *(erwartungstreu) und Kovarianzmatrix*

$$(A^TA)^{-1}A^T(\mathrm{vE})A(A^TA)^{-T} = v(A^TA)^{-T} \overset{A^TA\,\text{symmetrisch}}{=\!=\!=\!=\!=} v(A^TA)^{-1}$$

*(ii)*

*Sei* $(u_1, \ldots, u_n)$ *eine ONB von* $\mathbb{R}^n$ *mit*

$H = \mathrm{span}\{u_1, \ldots, u_r\}, L = \mathrm{span}\{u_1, \ldots, u_s\}$ $r < s$:

$$O = \left(\begin{array}{c|c|c} u_1 & \ldots & u_n \\ & & \\ & & \end{array}\right) \text{orth. Matrix}$$

*Da* $\zeta \sim \mathcal{N}_n(0, E) \Rightarrow \eta := O^T\zeta \sim \mathcal{N}_n(0, E)$

$\Rightarrow \frac{n-s}{v}V^* = \frac{1}{v}|X - \Pi_L X|^2 = \frac{1}{v}|A\gamma + \sqrt{v}\zeta - \Pi_L(A\gamma + \sqrt{v}\zeta)|^2 = \left|\zeta - \underbrace{\Pi_L}_{\mathrm{OE}_s O^T}\zeta\right|^2 = \sum_{k=s+1}^n \eta_k^2 \sim \chi^2_{n-s}$.

*(iii)* $\frac{1}{v}|A(\hat{\gamma}-\gamma)|^2 = \frac{1}{v}|\Pi_L X - A\gamma|^2 = \frac{1}{v}|\Pi_L A\gamma + \sqrt{v}\Pi_L\zeta - A\gamma|^2 = |\Pi_L\zeta|^2 = \sum_{k=1}^s \eta_k^2 = \chi^2_s$.

*und ist unabhängig von $V^*$.*

*Dann gilt auch* $\frac{|A(\hat{\gamma}-\gamma)|^2}{\mathrm{sV}^*}$

$$=\frac{\frac{1}{s}\sum_{k=1}^{s}\eta_k^2}{\frac{1}{n-s}\sum_{k=s+1}^{n}\eta_k^2}\sim\mathcal{F}_{s,n-s}.$$

*(iv)* $\frac{1}{v}|\Pi_L X-\Pi_H X|^2=|\Pi_L\zeta-\Pi_H\zeta|^2=\sum_{k=r+1}^{s}\eta_k^2\sim\chi_{s-r}^2$ *und ist unabhängig von $V^*$.*

*Also* $F_{H,L}:=\frac{\frac{1}{s-r}\sum_{k=r+1}^{s}\eta_k^2}{\frac{1}{n-s}\sum_{k=s+1}^{n}\eta_k^2}\sim\mathcal{F}_{s-r,n-s}.$

□

Ende Vorlesung 22

---

Mit Satz 115 kann man Konfidenzbereiche und Tests konstruieren.

**Satz 116.** *Sei $\alpha\in(0,1)$ ein Irrtumsniveau.*

*i.* *Konfidenzbereich für $\gamma$*

$$C_{\hat{\gamma}}(X)=\{\gamma\in\mathbb{R}^s\colon|A(\gamma-\hat{\gamma})|^2<\mathrm{sV}^*\mathcal{F}_{s,n-s:1-\alpha}\}.$$

*ii.* *Konfidenzintervall für $\tau(\gamma)=c\gamma$*

$$C_\tau(X)=(c\hat{\gamma}-\delta\sqrt{V^*},c\hat{\gamma}+\delta\sqrt{V^*})$$

*mit* $\delta:=t_{n-s:1\alpha}\sqrt{c^T(A^TA)^{-1}c}.$

*iii.* *Konfidenzintervall für Varianz v:*

$$C_v(X)=\left((n-s)\frac{V^*}{q_+},(n-s)\frac{V^*}{q_-}\right)$$

*wobei* $q_-=\chi^2_{n-s:\alpha/2}, q_+=\chi^2_{n-s:1-\alpha/2}$.

***Beweis.*** *(i) Satz 115* $\frac{|A(\gamma-\hat{\gamma})|^2}{\mathrm{sV}^*}\sim\mathcal{F}_{s,n-s}$

$$\Rightarrow\mathbb{P}_{\gamma,v}\left[x\in\mathbb{R}^n\colon\underbrace{\frac{|A(\gamma-\hat{\gamma})|^2}{\mathrm{sV}^*}}_{\sim\mathcal{F}_{s,n-s}}\leqslant\mathcal{F}_{s,n-s:1-\alpha}\right]=1-\alpha$$

*Erinnerung(Konfidenzbereich):*$\mathbb{P}_{\gamma,v}[x\in\mathbb{R}^n\colon(\gamma,v)\in C(X)]\geqslant 1-\alpha\forall(\gamma,v)\in\mathbb{R}^s\times(0,\infty)$

*(ii)* $z:=c^T\hat{\gamma}\sim\mathcal{N}(c^T\gamma,\mathrm{v}c^T(A^TA)^{-1}c)$

$\Rightarrow z^*=\frac{c^T\hat{\gamma}-c^T\gamma}{\sqrt{c^T(A^TA)^{-1}c}}\sim\mathcal{N}(0,1)$

$\Rightarrow T:=\frac{z^*}{\sqrt{V^*/v}}\sim t_{n-s}$, *da* $z^*$

*eine Funktion von $A\hat{\gamma}$ (aus dem Beweis Satz 114 (ii)) ist und daher nach Satz 115 (iii) unabhängig von $V^*$ ist.*

$$\Rightarrow 1-\alpha = \mathbb{P}_{\gamma,v}[|T| \leqslant t_{n-s:1-\alpha/2}]$$

$$= \mathbb{P}_{\gamma,v}\left[|c^T\gamma - c^T\hat{\gamma}| \sqrt{\frac{V^*}{v}} \sqrt{\mathrm{v}c^T(A^TA)^{-1}c} t_{n-s:1-\alpha/2}\right]$$

$$= \mathbb{P}_{\gamma,v}[].$$

*(iii)* $\mathbb{P}_{\gamma,v}\Big[v \in \Big((n-s)\frac{V^*}{q_+}, (n-s)\frac{V^*}{q_-}\Big)\Big]$

$$= \mathbb{P}_{\gamma,v}\Big[q_- < \frac{n-s}{v} V^\times < q_+\Big] = 1-\alpha.$$

□

## Abschließende Anwendung auf Tests:

**Satz 117.** *Sei $0 < \alpha < 1$ ein Irrtumsniveau.*

*i.* *t-Test von Hypothese $c\gamma = m_0$*

$H_0$: $c\gamma = m_0$ gegen $H_1$: $c\gamma \neq m_0$

$$\Rightarrow \varphi = \mathbb{1}_{\left\{|c^T\hat{\gamma} - m_0| > \sqrt{c^T(A^TA)^{-1}c}\sqrt{V^*} t_{n-s:1-\alpha/2}\right\}}$$

*ist ein Test mit Irrtumsniveau $\alpha$.*

*ii.* *$\mathcal{F}$-Test der linearen Hpyothese $A\gamma \in H$*

*Sei $H \subset L$ ein lin. Unterraum mit $\dim H = r < s$-*

*$H_0$: $A\gamma \in H$ gegen $H_1$: $A\gamma \notin H$*

$$\Rightarrow \varphi(x) = \mathbb{1}_{\{F_{H,L} > \mathcal{F}_{s-r,n-s:1-\alpha}\}}$$

*ist ein Test mit Irrtumsniveau $\alpha$.*

*iii.* *$\chi^2$-Test für die Varianz: Für $v_0 > 0$*

*$H_0$: $v \leqslant v_0$ gegen $H_1$: $v > v_0$*

$$\Rightarrow \varphi(x) = \mathbb{1}_{\{(n-s)V^* > v_0\chi^2_{n-s:1-\alpha}\}}$$

*ist ein Test mit Irrtumsniveau $\alpha$.*

***Beweis.*** *(i) Ergibt sich aus Satz 116 (ii)*

*(ii) Für alle $\gamma \in H_0, v > 0$ hat*

$$F_{H,L} := \frac{n-s}{s-r} \frac{|\Pi_L X - \Pi_H X|^2}{|X - \Pi_L X|^2}$$

*unter $\mathbb{P}_{\gamma,v}$ die F-Verteilung (Satz 115 (iv))*

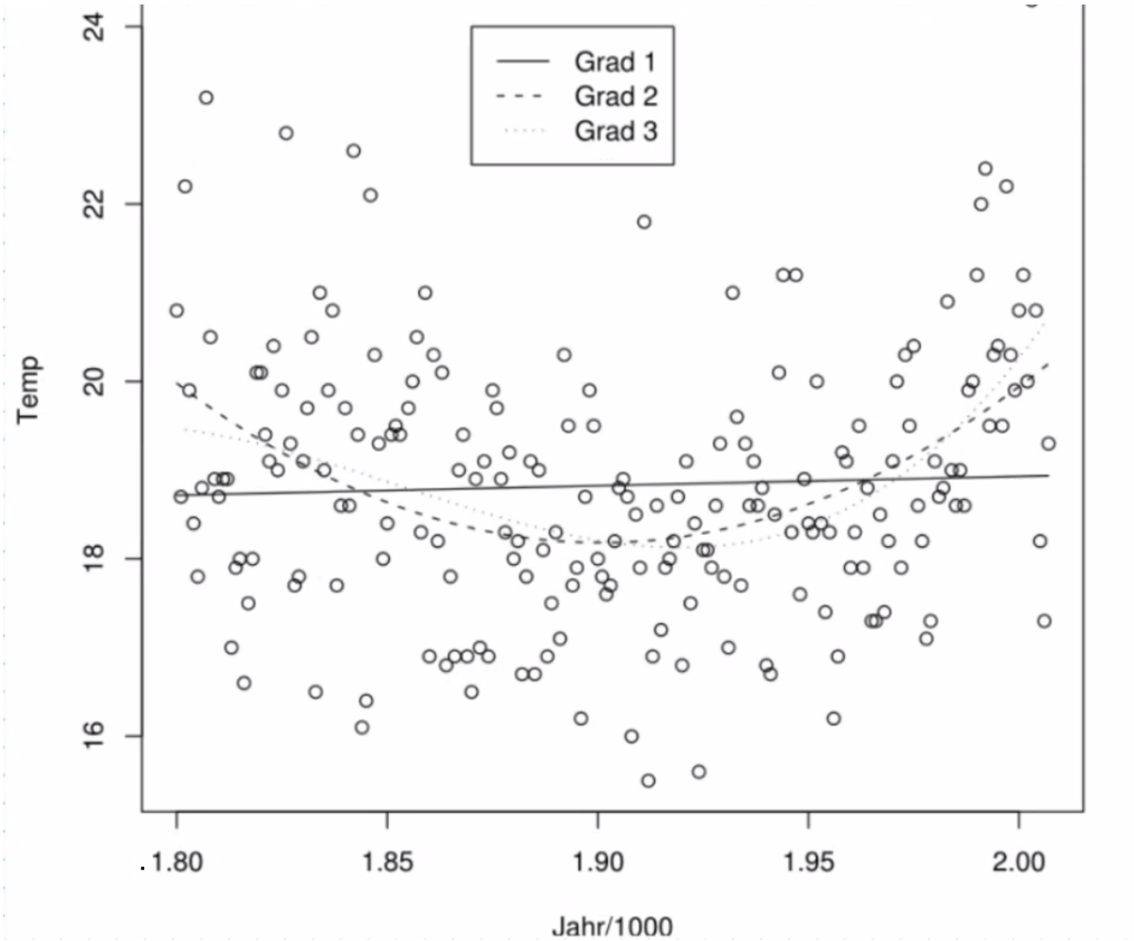
$\Rightarrow \mathbb{P}_{\gamma,v}[F_{H,L} > \mathcal{F}_{s-r,n-s:1-\alpha}] = \alpha$

*(iii) Ergibt sich aus dem Satz 116 (iii).* □

**Beispiel 118.** Durchschnittstermperaturen in Karlsruhe im August in den Jahren 1800 bis 2007 (n=206: Beobachtungen der Jahre 1854, 1945 fehlen.)

$t_1, \ldots, t_n$ Jahre in Jahren/1000

$X_1, \ldots, X_n$ mittlere Temperaturen im August in Karlsruhe




(i) $A = \underbrace{(1,t)}_{s=2}$: $X_k = \gamma_0 + \gamma_1 t_k + \sqrt{v}\zeta_k, \zeta_k u.i.v. \quad \mathcal{N}(0,1)$

KQS: $\hat{\gamma_0}$=18.7, $\hat{\gamma_1} = 0.1$

Frage: Was sagt der Test

$H_0$: $\gamma_1 \leqslant 0$ gegen $H_1$: $\gamma_1 > 0$

Satz 117 $\Rightarrow$ t-Test!

Wir berechnen $A^T A$ und $(A^T A)^{-1}$

$$A^T A = \begin{pmatrix} 1 & \ldots & 1 \\ t_1 & \ldots & t_n \end{pmatrix} \begin{pmatrix} 1 & t_1 \\ \vdots & \vdots \\ 1 & t_n \end{pmatrix} = \begin{pmatrix} n & \sum_{k=1}^n t_k \\ \sum_{k=1}^n t_k & \sum_{k=1}^n t_k^2 \end{pmatrix}$$

$$\Rightarrow (A^T A)^{-1} = \frac{1}{\mathrm{nVar}(t)} \begin{pmatrix} \frac{1}{n} t_k^2 & \frac{1}{n}\sum_{k=1}^n t_k \\ \frac{1}{n}\sum_{k=1}^n t_k & 1 \end{pmatrix}, \text{ weil } \det(A^T A) = n^2 \mathrm{Var}\, t$$

$$c = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \Rightarrow c^T (A^T A)^{-1} c = \frac{1}{\mathrm{nVar}\, t}$$

Hypthese wird abgelehnt, wenn

$$\hat{\gamma}_1 > \underbrace{t_{n-2:1-\alpha}\sqrt{\frac{V^*}{n\operatorname{Var}(t)}}}_{\approx 0.29}$$

$\alpha = 0.1 \Rightarrow t_{n-2:1-\alpha} \approx 1.65$

$\Rightarrow$ Die Hypothese wird nicht abgelehnt!

(ii) Ist die lineare Interpolation ausreichend?

Sei $X_k = \underbrace{\gamma_0 + \gamma_1 t_k + \gamma_2 t_k^2 + \gamma_3 t_k^3 + \gamma_4 t_k^4}_{s=5} + \sqrt{v}\zeta_k$

<u>Test:</u> $H_0$: $\gamma_2 = \gamma_3 = \gamma_4 = 0, \alpha = 0.05$

Satz 117 $\Rightarrow F$-Test

$L_1 = \operatorname{span}\{1, t\}$ „linear“

$L_4 = \operatorname{span}\{1, t, t^2, t^4\}$ „polynomiell“

$\Rightarrow F_{L_1, L_4} = \frac{n-s}{s-r}\frac{|\Pi_{L_4}X - \Pi_{L_1}|^2}{|X - \Pi_{L_4}|^2}$

$$= \frac{n-5}{5-2}\frac{\sum_{k=1}^n (P_4(t_k) - P_1(t_k))^2}{\sum_{k=1}^n (X_k - P_4(t_k))^2} = 14.26 > 2.65 = \mathcal{F}_{3,201:1-\alpha}$$

Weil Nullhypothese in $L_1$ liegt

Satz 117 (ii) $\Rightarrow$ Nullhypothese der Linearität wird abgelehnt!

<u>Varianzanalyse (ANOVA:Analysis Of VAriance.)</u>

**Beispiel 119.** (Einfluss Düngemittel auf Ernteertrag)

- Seien $G = \{1, \ldots, s\}$ eine endliche Menge von Düngemitteln
- Jedes Düngemittel $i \in G$ wird auf $n_i \geqslant 2$ verschiedene Flächen $F_{i,1}, \ldots F_{i,n_i}$ auftragen.

<u>Ansatz:</u>

Ernteertrag $X_{i,k}$ auf Fläche $F_{i,k}$

$$X_{i,k} = m_i + \sqrt{v}\zeta_{i,k}$$

$i \in G, k = 1, \ldots, n_i$

| Gruppe | Beobachtungen | Erwartungswert |
|---|---|---|
| 1 | $X_{1,1}, \ldots, X_{1,n_1}$ | $m_1$ |
| 2 | $X_{2,1}, \ldots, X_{2,n_2}$ | $m_2$ |
| $\vdots$ | $\vdots$ | $\vdots$ |
| s | $X_{s,1}, \ldots, X_{s,n_s}$ | $m_s$ |

**Tabelle 8.**

Als Vektor:

X$=(X_{1,1},\dots,X_{1,n_1},\dots,X_{s,1},\dots,X_{s,n_s})^T$

Unbekannte Parameter

$\gamma=(m_1,\dots,m_s)\in\mathbb{R}^s$

$v>0$

$\mathbb{E}[X]=\left(\underbrace{m_1,\dots,m_1}_{n_1},\underbrace{m_2,\dots,m_1}_{n_2},\dots,\underbrace{m_s,\dots,m_s}_{n_s}\right)$

Die Designmatrix ist

$$A=\underbrace{\begin{pmatrix} 1 & & & \\ \vdots & & 0 & \\ 1 & & & \\ 0 & 1 & & \\ \vdots & \vdots & & \\ \vdots & 1 & & \\ \vdots & 0 & & 1 \\ 0 & \vdots & & \vdots \\ 0 & 0 & & 1 \end{pmatrix}}_{s}$$

$A\gamma=\mathbb{E}[X]$

Ende Vorlesung 23

---

(Bonus: Finanzmathematik nicht relevant für die Klausur)

ANoVa:

$X_{i,k}=m_i+\sqrt{v}\zeta_{i,k}$: $i\in G, k=1,\dots,n_i$

X$=(X_{1,1},\dots,X_{1,n_1},\dots,X_{s,1},\dots,X_{s,n_s})^T$

Unbekannte Parameter

$\gamma=(m_1,\dots,m_s)\in\mathbb{R}^s$

$v>0$

Die Designmatrix ist

$$A=\underbrace{\begin{pmatrix} 1 & & & \\ \vdots & & 0 & \\ 1 & & & \\ 0 & 1 & & \\ \vdots & \vdots & & \\ \vdots & 1 & & \\ \vdots & 0 & & 1 \\ 0 & \vdots & & \vdots \\ 0 & 0 & & 1 \end{pmatrix}}_{s}$$

$A\gamma = \mathbb{E}[X]$

**Definition 120.** *Das Modell der Varianzanalyse besteht aus dem folgendem*

i. *Einer endlichen Menge $G$ von Beobachtungsgruppen mit $|G| = s$*

ii. *Anzahl $n_1, \ldots, n_s$ der Beobachtungen mit $n = \sum_{i=1}^{s} n_i$.*

iii. *Ein unbekannter Vektor $\gamma = (m_1, \ldots, m_s)$ von Beobachtungsmittelwerten $m_i$ der Gruppe $i \in G$, ein unbekannter Skalenparameter $v > 0$, paarweise unkorrelierte Störgrößen $\zeta_{i,k}$ mit $i \in G, k \in \{1, \ldots, n_i\}$.*

*$\Rightarrow$ Die Varianzanalyse ist durch das lineare Modell mit der $n \times s$ Designmatrix $A$ gegeben.*

**Lemma 121.**

i. *$L := \{A\gamma \colon \gamma \in \mathbb{R}^s\}$ ist gegeben durch $\{x \in \mathbb{R}^n \colon x_1 = \cdots = x_{n_1}, x_{n_1+1} = \cdots = x_{n_1+n_2}, \ldots, x_{n_1+\cdots+n_{s-1}+1} = \cdots = x_n\}$*

ii. $A^T A = \begin{pmatrix} n_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & n_s \end{pmatrix}$

iii. $A^T X = (n_1 M_1, \ldots, n_s M_s)^T$ wobei

$$M_i = \frac{1}{n_i} \sum_{k=1}^{n_i} X_{i,k}$$

*der Beobachtungsmittelwert innerhalb der Gruppe $i$, denn $A^T X = (\sum_{k=1}^{n_1} X_{1,k}, \ldots, \sum_{k=1}^{n_s} X_{s,k})^T$.*

iv. *Erwartungstreuer Schätzer $\hat{\gamma}$ für $\gamma$:*

$$\hat{\gamma} = (A^T A)^{-1} A^T X = \begin{pmatrix} M_1 \\ \vdots \\ M_s \end{pmatrix}, \text{denn}$$

$$\begin{pmatrix} \frac{1}{n_1} & & \\ & \ddots & \\ & & \frac{1}{n_s} \end{pmatrix} \begin{pmatrix} n_1 M_1 \\ \vdots \\ n_s M_s \end{pmatrix} = \begin{pmatrix} M_1 \\ \vdots \\ M_s \end{pmatrix}.$$

v. *Erwartungstreuer Schätzer $V^*$ für $v$.*

$$V^* := V_{\mathrm{iG}}^* := \frac{1}{n-s} \sum_{i \in G} (n_i - 1) V_i^* \text{ mit}$$

$$V_i^* := \frac{1}{n_i - 1} \sum_{k=1}^{n_i} (X_{i,k} - M_i)^2$$

*$V_i^*$: erwartungstreuer Schätzer für $v$ innerhalb der Gruppe $i$.*

*$V_{\mathrm{iG}}^*$: Stichprobenvarianz innerhalb der Gruppen.*

*Denn:*

$$V^* = \frac{1}{n-s}|X - \Pi_L X|^2 = \frac{1}{n-s}|X - A\hat{\gamma}|^2 = \frac{1}{n-s}\left|X - \left(\underbrace{M_1,\ldots,M_1}_{n_1},\ldots,\underbrace{M_s,\ldots,M_s}_{n_s}\right)\right|^2$$

$$\frac{1}{n-s}\sum_{i\in G}\sum_{k=1}^{n_i}(X_{i,k} - M_i)^2 = V^*_{\text{iG}}$$

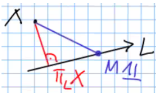
vi. *Totale empirische Varianz*

$$V^*_{\text{tot}} = \frac{1}{n-1}\sum_{i\in G}\sum_{k=1}^{n_i}(X_{i,k} - M)^2 = \frac{|X - M\mathbb{1}|^2}{n-1}, \text{wobei}$$

$M := \frac{1}{n}\sum_{i\in G}\sum_{k=1}^{n_i} X_{i,k}$ *totale empirische Mittelwert.*

*Benutzte* $M\mathbb{1}\in L$*:*

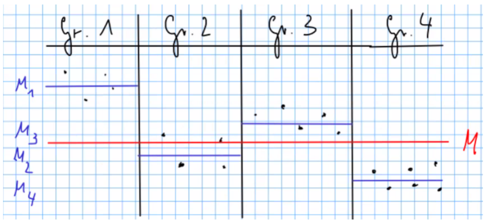
*Pythagoras:* $|X - M\mathbb{1}|^2 = |X - \Pi_L X|^2 + |\Pi_L X - M\mathbb{1}|^2$




$\Rightarrow (n-1)V^*_{\text{tot}} = (n-s)V^*_{\text{iG}} + (s-1)V^*_{\text{zG}},$ *wobei*

$$V^*_{\text{zG}} = \frac{1}{s-1}\sum_{i\in G} n_i(M_i - M)^2 = \frac{|\Pi_L X - M\mathbb{1}|^2}{s-1}$$

*Stichprobenvarianz zwischen den Gruppen.*




vii. $\mathbb{E}_{\gamma,v}[V^*_{\text{tot}}] = v + \frac{1}{n-1}\sum_{i\in G} n_i(m_i - \bar{m})^2$: $(*)$

*wobei* $\bar{m} = \frac{1}{n}\sum_{i\in G} n_i m_i$.

$\Rightarrow \mathbb{E}[V^*_{\text{tot}}] = v \Rightarrow m_i = \bar{m}\forall_i$

*Beweis von* $*$*:*

*In der Tat, sei* $H = \{x\in\mathbb{R}^n : x_1 = \cdots = x_n\} = \operatorname{span}\{\left(\begin{smallmatrix}1 & \ldots & 1\end{smallmatrix}\right)^T\}$

$\Rightarrow \underbrace{M\mathbb{1}}_{\in H} = \Pi_H X$

$M\mathbb{1} - X = \left(\frac{1}{n}\sum X_{i,k} - X_{1,1},\ldots,\frac{1}{n}\sum X_i - X_{s,n_s}\right)$

$\Rightarrow (n-1)V^*_{\text{tot}} = |X - \Pi_H X|^2 = |\Pi_H \perp X|^2$

$$\overset{X = A\gamma + \sqrt{v}\zeta}{=\!=\!=\!=} |\Pi_H \perp A\gamma|^2 + v|\Pi_H \perp \zeta|^2 + 2\sqrt{v} < A\gamma, \Pi_H \perp \zeta >$$

$\Rightarrow \mathbb{E}_{\gamma,v}[(n-1)V^*_{\text{tot}}] \overset{\mathbb{E}[\zeta]=0, \mathbb{E}[|\Pi_H\perp\zeta|^2]}{=\!=\!=\!=} |\Pi_H \perp A\gamma|^2 + v(n-1)$

*wobei* $\mathbb{E}[|\Pi_H \perp \zeta|^2] = \mathbb{E}\left[\left|E_{n-1}\underbrace{O^T\zeta}_{\eta}\right|^2\right] = \mathbb{E}[\sum_{i=1}^{n-1} \eta_i^2] = \sum_{i=1}^{n-1} \underbrace{\mathbb{E}[\eta_i^2]}_{\mathbb{E}[\sum_{i,j} O_{i,k}\zeta_i O_{j,k}\zeta_j] = \mathbb{E}[\sum_{i=1}^n O_{i,k}^2]=1} = (n-1)\cdot 1$

$A\gamma = \mathbb{E}[X] = (m_1, \ldots, m_s)^T$

$|\Pi_H \perp A\gamma|^2 = |A\gamma - \Pi_H A\gamma|^2 = \left|(m_1, \ldots, m_s)^T - \frac{1}{n}\sum_{i\in G} n_i m_i \cdot (1, \ldots, 1)^T\right|^2$

$$\sum_{i\in G} n_i\left(m_i - \frac{1}{n}\sum_{j\in G} n_j m_j\right)^2 = \sum_{i\in G} n_i (m_i - \bar{m})^2$$

$\Rightarrow \mathbb{E}_{\gamma,v}[V_{\text{tot}}^*] = v + \frac{1}{n-1}\sum_{i\in G} n_i (m_i - \bar{m})^2 \Rightarrow (*)$ □

Annahme: $\zeta_{i,k} \sim \mathcal{N}_{0,1}$ u.i.v.

$\Rightarrow$ lineares Gauß. Modell

$\Rightarrow$ Konfidenzbereiche und Tests vorhanden.

Wir beschränken uns auf zwei Beispiele.

**Beispiel 122.** (t-Test im 2-Stichprobenproblem)

$s = 2$ „Test auf Gleichwertigkeit"

$H_0$: $m_1 = m_2$ gegen $H_1$: $m_1 \neq m_2$.

$\Rightarrow c = (1, -1)^T \Rightarrow H_0$: $c\gamma = 0$ mit $\gamma = (m_1, m_2)$.

Nach Satz 117 (i) hat der Test mit Ablehnungsbereich

$$\left\{\left|\underbrace{M_1 - M_2}_{c\hat{\gamma} - 0}\right| > t_{n-2:1-\alpha/2}\right\} \cdot \sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right) V_{\text{iG}}^*}$$

Weil

$$(1, -1)\begin{pmatrix} \frac{1}{n_1} & 0 \\ 0 & \frac{1}{n_2} \end{pmatrix}(1, -1)^T V^*$$

**Beispiel 123.** (F-Test Mehrstichprobenproblem)

Sei $s \geq 2$: „Test auf Gleichwertigkeit"

$H_0$: $m_1 = m_2 = \cdots = m_s$ gegen $H_1$: $\exists i \neq j \quad s.d.$: $m_i \neq m_j$.

Sei $H = \{m\mathbb{1} : m \in \mathbb{R}\}$. Dann ist $H$ ein Unterraum von $L = \{A\gamma : \gamma \in \mathbb{R}^s\}$.

Aus Satz 117 (ii) betrachte

$$F_{H,L} := \frac{n-s}{s-1}\frac{|\Pi_L X - \Pi_H X|^2}{|X - \Pi_L X|^2} = \frac{n-s}{s-1}\frac{|\Pi_L X - M\mathbb{1}|^2}{|X - \Pi_L X|^2} = V_{\text{zG}}^* / V_{\text{iG}}^*$$

wobeo $\frac{n-s}{|X-\Pi_L X|^2}=V_{\text{iG}}^*$ und $\frac{|\Pi_L X - M\mathbb{1}|^2}{s-1}=: V_{\text{zG}}^*$.

Aus Satz 117 (ii) folgt der Test mit Ablehnungsbereich

$$\{V_{\text{zG}}^* > \mathcal{F}_{s-1,n-s:1-\alpha} V_{\text{iG}}^*\}$$

hat Niveau $\alpha$.

Ende Vorlesung 24

---

Klausurrelevant:

Kapitel 1, d.h.:

- Parameterschätzung
- Konfidenzbereichen
- Hypothesentests
- Regression
- Varianzanalyse

$\Rightarrow$ Alles bis auf Finanzmathematik

# 9 Einf. in die Finanzmathematik

## 9.1 Einleitung

Wir fangen mit einem Beispiel an.

Ein Feinkostladen (FK) hat heute ($t=0$) schweizer Käse für 100.000 CHF bestellt. Die Lieferung erfolgt 3 Monate später $t=T=3$ und muss am Liefertag bezahlt werden. Der wechselkurs heute 1CHF = 0.65€.

Problem:

Currency risk: FK weiß nicht was der Kurs sein wird in 3 Monaten. Was soll FK tun um sich gegen dieses Risiko zu schützen?

Einige Möglichkeiten:

1. FK kauft heute 100.000 CHF zum Preis von 65.000€ und legt das Geld auf ein Konto bis $t=T$.

   Vorteil: Kein Risiko.

   Nachteil: Für drei Monate sind die 65.000€ nicht benutzbar.

   Außerdem: Vielleicht hat FK heute gar keine 65.000€ zur Verfügung.

2. FK kauft eine forward contract für 100.000 CHF mit Ablieferung in 3 Monaten bei einer Bank.

   - die Bank wird 100.000 CHF zur Zeit $t=T$ bezahlen.

- FK bezahlt später die Bank mit Kurs $K$ €/CHF, wobei $K$ zur Zeit $t=0$ festgelegt wird.

Vorteil: Kein Risiko, muss nicht schon heute bezahlt werden.

Nachteil: Sei z.B. $K=0.67$€/CHF.

Da der Vertrag rechstverbindlich ist, können folgende zwei Situationen eintreten:

i. Zur Zeit $t=T$ ist der Wechselkurs 0.7€/CHF.

   Dann ist FK zufrieden, weil FK kauft 100.000 CHF für 67.000€, d.h. 3000€ indirekter Gewinn.

ii. Zur Zeit $t=T$, ist der Wechselkurs 0.65€/CHF. Dann muss FK die 100.000 CHF für einen Preis von 67.000€ kaufen und ärgert sich über den indirekten Verlust von 2000€.

3. Was FK eigentlich will, ist ein Vertrag, der gegen zu hohe Kosten schützt, aber unverbindlich ist, d.h. wen nder Kurs fällt, muss man nicht kaufen.

   $\rightarrow$ Diese Verträge heißen European Call Option.

**Definition 124.** *Eine European Call Option von $X$ CHF mit Ausübungspreis (strike price) $K$€/CHF und Ausübungstag (exercise date) $T$ ist ein Kontrakt der am Zeitpunkt $t=0$ geschrieben ist.*

*i. Der Kontraktbesitzer hat, zur Zeit $t=T$, das Recht $X$ CHF zu kaufen für den Preis $K$€/CHF.*

*ii. Der Kontraktbesitzer hat die Möglichkeit, aber nicht die Verpflichtung zu kaufen.*

**Bemerkung 125.** „Option“: Weil man die Möglichkeit hat zur Zeit $t=T$ zu entscheiden.

„Call“: Recht zu kaufen

vs.

„Put“: Recht zu verkaufen

„European“: Nur zur Zeit $t=T$

vs.

„American“: Zu jedem Zeitpunkt $t\leqslant T$

Eine Möglichkeit für FK ist also eine European Call Option, *z.B.* $K=0.67$€/CHF und wenn der Kurs zur Zeit $t=T$ größer ist als $K$, dann benutz man die Option, sonst nicht.

Vorteil: Kein Risiko, Flexibilität

Nachteil: Solche Optionen kosten etwas.

Probleme: Welche Option sollte man kaufen?

Fragen:

- Was ist ein „fairer Preis“ einer solchen Option?
- Wenn wir eine Option verkauft haben, haben wir ein Risiko un den anderen versichert gegen dieses Risiko.
- Wie sollen wir uns gegen das Risiko schützen (hedging)?

## 9.2 Binomialmodell

### 9.2.1 Eine Periode

Sei $t$ die Zeit und wir betrachten nur einen Schritt, von heute ($t=0$) nach morgen ($t=1$).

Betrachten wir zwei Assets (Wertpapiere)

- ein Bond (Anleihe) gibt feste Zinsen $R$.
- ein Stock (Aktien) variabel, Marktabhängig.

Sei $B_t$ der Bondpreis zur Zeit $t$ und $S_t$ der Stockpreis zur Zeit $t$.

$B_0=1 \Rightarrow B_1=1+R$: deterministisch

$S_0=s \Rightarrow S_1=\begin{cases} s\cdot u & \text{mit } W'\text{keit } p_u, u>1 \\ s\cdot d & \text{mit } W'\text{keit } p_d, d<1 \end{cases}$

mit $p_u+p_d=1$

Sei $Z$ die ZV mit

$$\mathbb{P}[Z=u]=p_u$$

$$\mathbb{P}[Z=d]=p_d$$

$S_0=s \Rightarrow S_1=s\cdot Z$

Wir nehmen an, dass $R, s, u, d, p_u, p_d$ bekannt sind mit $u>1>d$ und $p_u+p_d=1$.

Bond: $1 \rightarrow 1\cdot(R+1)$

Stock:

Entweder

$s \rightarrow s\cdot u$

oder

$s \rightarrow s\cdot d$

### 9.2.2 Portfolios und Arbitrage

Wir analysieren jetzt verschiedene Zusammensetzungen von Bonds und Stocks im $(B,S)-$Markt.

Ein Portfolio ist dann ein Vektor $h=(x,y)\in\mathbb{R}^2$ mit $x=\#\text{Bonds in } h$ und $y=\#\text{Stocks in } h$.

**Bemerkung 126.** $x$ und $y$ können positiv aber auch negativ sein.

**Beispiel 127.** $x=3$ bedeutet, dass wir 3 Bonds haben zur Zeit $t=0$

$y=-2$ bedeutet, dass wir 2 Stocks zur Zeit $t=0$ verkauft haben.

Annahmen:

1) Jedes $h\in\mathbb{R}^2$ ist zulässig (man kann auch Teile von Bonds/Stocks besitzen)

2) Verkaufspreis = Kaufspreis

3) Keine Transaktionskosten

4) Markt ist liquide, d.h. es ist immer möglich zu kaufen/verkaufen, so viel wie gewünscht.

- Sei $h=(x,y)$ ein Portfolio zur Zeit $t=0$.

Wir schauen uns an wie der Portfoliowert sich entwickelt.

- Der Wert zur Zeit $t = 0$ ist deterministisch:

  Markwert.

- Der Wert Zur Zeit $t = 1$ ist stochastisch.

Ende Vorlesung 25

---

**Definition 128.** *Der Value-process von einem Portfolio $h = (x, y)$ ist definiert durch*

$$V_t^h := \mathrm{xB}_t + \mathrm{yS}_t, t = 0, 1, d.h.$$

$$V_0^h = x + \mathrm{ys}$$

$$V_1^h = x(1 + R) + \mathrm{ysZ}.$$

**Definition 129.** *Ein Arbitrage Portfolio ist ein Portfolio $h$ mit den folgenden Eigenschaften:*

$$V_0^h = 0$$

$$V_1^h > 0 \,\mathrm{mit}\, W'\mathrm{keit}\, 1.$$

*Was bedeutet das? Mit Null-Kapital kann man sicher Geld erzeugen.*

*Wenn das der Fall ist, dann gibt es ein mispricing im Markt.*

*Wir gehen davon aus, dass das in effizienten Märkten nicht möglich ist.*

$\Rightarrow$ *Arbitrage-frei.*

**Satz 130.** *Das Binomialmodell ist frei von Arbitrage genau dann wenn*

$$d \leqslant 1 + R \leqslant u \, (*)$$

**Bemerkung 131.** Die Bedingung $(*)$ bedeutet, dass ein Gewinn im Stockwert nicht von dem Gewinn im Bondwert domoniert seien kann und umgekehrt der Gewinn im Bondwert den Verlust im Stockwert dominiert.

**Beweis.** (Satz 125)

$\Rightarrow$:

Nehmen wir an, dass $(*)$ nicht gilt.

$$\Rightarrow s(1 + R) > \mathrm{su}$$

Dies gilt, weil $d < 1$ und damit immer $\leqslant 1 + R$.

Dann haben wir auch $s(1+R)>sd$

$\Rightarrow$ Es ist besser in Bond zu investieren als in Stock.

Sei $h=(s,-1)$

$\Rightarrow V_0^h = s - s = 0$ und

$$V_1^h = s(R+1) - \underbrace{\mathrm{sZ}}_{<s(R+1)} > 0$$

$\Rightarrow \exists$ Arbitrage Portfolio.

$\Leftarrow$ Nehmen wir an, dass $(*)$ gilt. Sei $h$ ein Portfolio mit $V_0^h = 0$.

$$x + \mathrm{ys} = 0, d.h. x = -\mathrm{ys}$$

$$\Rightarrow V_1^h = \begin{cases} \mathrm{ys}(u-(R+1)) & Z=u \\ \mathrm{ys}(d-(R+1)) & Z=d \end{cases}.$$

Wenn $y>0 \Rightarrow h$ h ist ein Arbitrage Portfolio genau dann wenn

$$\begin{cases} u > 1+R & \\ d > 1+R & \text{Widerspruch zu } (*) \end{cases}$$

Wenn $y<0 \Rightarrow$ h ist ein Arbitrage Portfolio genau dann wenn

$$\begin{cases} u < 1+R & \text{Widerspruch zu } (*) \\ d < 1+R & \end{cases}$$

$\square$

Das Ergebnis von Satz 2.3 kann wie folgt umgeschrieben werden:

$$1+R = q_u \cdot u + q_d \cdot d$$

Wobei $q_u, q_d \geqslant 0$ und $q_u + q_d = 1$, d.h.

$$1+R = q_u \cdot u + (1-q_u) \cdot d$$

$$\Rightarrow q_u = \frac{1+R-d}{u-d}$$

$$q_d = \frac{u-(1+R)}{u-d}$$

Wir können $q_u$ und $q_d$ als W'keiten eines W'maßes interpretieren, wobei

$$\frac{1}{R+1} Q[Z=u] = q_u, Q[Z=d] = q_d$$

Es gilt:

$$\mathbb{E}_Q[S_1] = \frac{1}{R+1}$$

$$= \frac{1}{R+1}(q_u \cdot s \cdot u + q_d \cdot s \cdot d) = \frac{1}{R+1} \cdot s \cdot \underbrace{(q_u \cdot u + q_d \cdot d)}_{R+1} = s$$

Bedeutung: Der Stockpreis verhält sich Risiko-neutral.

heutiger Preis=erwarteter Preis von morgen (diskontiert, d.h. in der Währung Bond).

**Definition 132.** *Ein W'keitmaß $Q$ heißt Martingalmaß, wenn $S_0 = \frac{1}{1+R}\mathbb{E}_Q[S_1]$.*

Literatur: Stochastic Finance von Hans Föllner und Alexander Schied.

**Satz 133.** *(„Fundamental Theortem of Asset Pricing")*
*Das Markmodell ist frei von Arbitrage genau dann wenn ein Martingalmaß existiert.*

**Beweis.** $S_0 = s$

$\Leftarrow$: Angenommen $\exists\ h = (x, y)$ s.d.

$$V_0^h = x + \mathrm{sy} = 0 \Rightarrow x = -\mathrm{sy}$$

und $V_1^h = x(R+1) + \mathrm{syZ} > 0$ f.s..

$\Rightarrow V_1^h = -\mathrm{ys}(R+1-Z) > 0$

Wenn $y > 0 \Rightarrow \mathbb{E}_Q[V_1^h] > 0$

$$\Leftrightarrow \mathbb{E}_Q[Z] < R + 1 \forall W - \text{Maße} Q$$

Wenn $y < 0 \Rightarrow \mathbb{E}_Q[V_1^h > 0]$

$$\Leftrightarrow \mathbb{E}_Q[Z] > R + 1 \forall W - \text{Maße} Q$$

$\mathbb{E}_Q[S_1]/(1+R) = S_Q\mathbb{E}[Z]/(1+R) \neq s \Rightarrow$ Es existiert kein Martingalmaß.

$\Rightarrow$ Wenn $\forall Q, \mathbb{E}_Q[Z] \neq R+1$.

Dann gilt auch das für die folgende Klasse von W'maßen:

$$u = \max\{Z\}, d = \min\{Z\}$$

$\Rightarrow Q = q_u\delta_u + q_d\delta_d$, $p_u + p_d = 1$

$\Rightarrow 1 + R \neq q_u u + (1 - q_u)d \Rightarrow 1 + R \notin [d, u]$

Aus Satz 125 folgt die Behauptung. □

**Folgerung 134.** *Für das Binomialmodell sind die Martingalmaß W'keiten gegeben durch*

$$q_u = \frac{1 + R - d}{u - d}$$

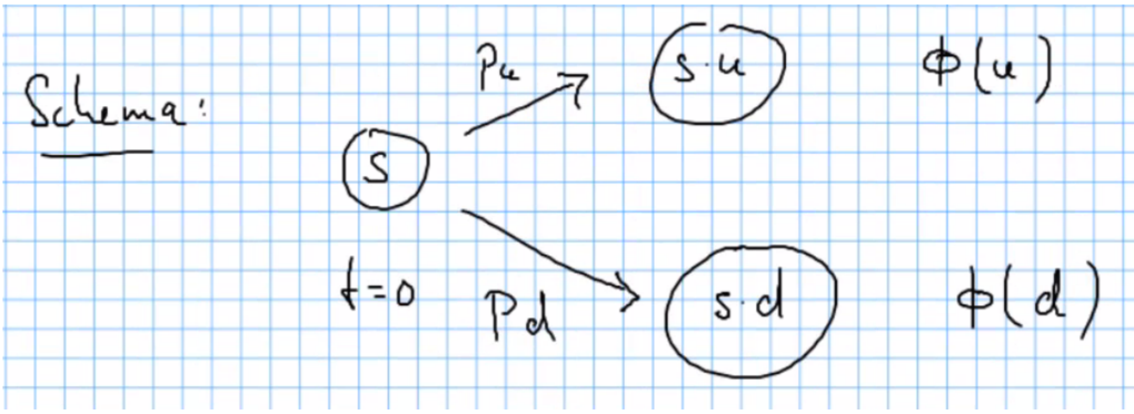
$$q_d = \frac{u - (R+1)}{u - d}$$

*Schon bewiesen!*

## 9.3 Financial derivative

**Definition 135.** *Ein Financial derivative (Derivat) ist jede zufällige Variable $X$ der Form $\Phi(Z)$, wobei $Z$ die zufällige Variable aus dem Stockpreis ist. $S_1 = \mathrm{sZ}$.*

*Wir interpretiere ein Derivat als einen Kontrakt der $X$€ an den Besitzer Zahlt zur Zeit $t = 1$.*




**Beispiel 136.** Eine European Call option an einem Stock mit Strike K.

$\Rightarrow$ Falls $s \cdot d < K < s \cdot u$

Falls $S_1 > K \Rightarrow$ Wir benutzen die Option den Stock zu kaufen und verkaufen es auf dem Markt für $S_1 = s \cdot u \Rightarrow$ Profit $s \cdot u - K$>0.

Falls $S_1 < K \Rightarrow$ Wir benutzen die Option nicht

$\Rightarrow$

$$X = \begin{cases} \text{su} - K & Z = u \\ 0 & Z = d \end{cases}$$

d.h. $\Phi(u) = \text{su} - K$ und $\Phi(d) = 0 \Rightarrow X = \Phi(Z)$.

Das Problem ist an diesem Punkt wie wir einen fairen Preis erhalten, wenn so etwas existiert:

$$\Pi(t, X) = \text{Preis von } X \text{ zur Zeit } t.$$

$$\Pi(1, X) = X \text{ um Arbitrage zu vermeiden}$$

$$\Pi(0, X) = ?$$

**Definition 137.** *Ein gegebenes Derivat $X = \Phi(Z)$ ist erreichbar, wenn ein Portfolio existiert s.d.*

$$V_1^h = X \text{ mit } W'\text{keit } 1.$$

*In diesem Fall ist $h$ ein hedging Portfolio. Wenn alle Derivate erreichbar sind, dann heißt der Markt komplett.*

**Bemerkung 138.** Wenn ein Derivat X erreichbar ist mit einem hedging Portfolio, dann macht es keinen Unterschied das Portfolio oder das Derivat zu besitzen.

First pricing principle: Wenn ein Derivat $X$ erreichbar ist mit einem hedging Portfolio $h$, dann ist der einzig „sinnvolle“ Preis für X gegeben durch

$$\Pi(t, X) = V_t^h, t = 0, 1.$$

Ende Vorlesung 26

---

**Beispiel 139.**

European Call Option:

Strike Price K $\rightarrow \Phi(Z) = \max\{0, S_1 - K\}$

**Satz 140.** *Nehmen wir an, dass X durch ein hedging Portfolio h erreichbar ist. Dann gibt jeder Preis zur Zeit $t=0$ für X außer $V_0^h$ eine Arbitrage Möglichkeit.*

**Beweis.** Sei $\Pi(0,X) < V_0^h$.

Dann kauft man zur Zeit $t=0$ 1mal $X$ und verkaufen $h$.

Zur Zeit $t=1$: Portfolio $h$ ist so viel wert wie Derivat $X$. D.h. wir haben ohne Risiko Gewinn $V_0^h - \Pi(0,X)$ gemacht. $\Rightarrow$ Haben Arbitrage Möglichkeit.:

| | Der | Port | Cash | Value |
|---|---|---|---|---|
| t=0 | 1 | $-1$ | $V_0^h - \Pi(0,X)$ | 0 |
| t=1 | 0 | 0 | $V_0^h - \Pi(0,X)$ | $V_0^h - \Pi(0,X)$ |

**Tabelle 9.**

Falls $\Pi(0,X) > V_0^h \Rightarrow$ statt kaufen, verkaufen- □

**Satz 141.** *Nehmen wir an, $d \leqslant R+1 \leqslant u$, d.h.*

*das Binomialmodell ist frei von Arbitrage. Dann ist es auch komplett.*

**Beweis.** Sei $X$ ein bel. Derivat mit Kontrakt Funktion $\Phi$.

Zu zeigen: $\exists h = (x,y)$ s.d.

$$V_1^h = \begin{cases} \Phi(u) & Z = u \\ \Phi(d) & Z = d \end{cases}$$

Explizit: $(1+R)x + \text{syu} = \Phi(u)$

oder $(1+R)x + \text{syd} = \Phi(d)$

Da $u > d$ und $\exists$ Lösung: $(*)$

$$x = \frac{1}{R+1} \cdot \frac{u\Phi(d) + d\Phi(u)}{u-d}$$

$$y = \frac{1}{s} \cdot \frac{\Phi(u) - \Phi(d)}{u-d}$$

□

Da unser arbitrage-freies Binomialmodell komplett ist, können wir jedes Derivat bewerten. Aus Satz 141 folgt:

$$\Pi(0,X)=V_0^h$$

mit $h=(x,y)$, wobei $x$ und $y$ lösen $(*)$.

D.h.

$$\Pi(0,X)=\frac{1}{R+1}\cdot\left\{\frac{1+R-d}{u-d}\Phi(u)+\frac{u-(R+1)}{u-d}\Phi(d)\right\}$$

Im Korollar ? hatten wir die W'keiten

$$q_u=\frac{1+R-d}{u-d},\, q_d=\frac{u-(R+1)}{u-d}$$

$\Rightarrow\Pi(0,X)=\frac{1}{R+1}(q_u\Phi(u)+q_d\Phi(d))=\frac{1}{R+1}\mathbb{E}_Q[X]$, wobei $Q$ das Martingalmaß aus dem Korollar ?.

Was haben wir erreich?

**Satz 142.** *Für das Arbitrage-freie Binomialmodell $(d\leqslant R+1\leqslant u)$ ist das Arbitrage-freie Preis einer Derivate $X=\Phi(Z)$*

$$\Pi(0,X)=\frac{1}{R+1}\mathbb{E}_Q[X]$$

*„Risiko Neutrale Bewertung“,*

*wobei das Martingalmaß $Q$ definiert ist durch*

$$S_0=\frac{1}{R+1}\mathbb{E}_Q\left[S_1\right]$$

*und das hedging Portfolio $h=(x,y)$ ist gegeben durch*

$$x=\frac{1}{R+1}\cdot\frac{u\Phi(d)+d\Phi(u)}{u-d}$$

$$y=\frac{1}{s}\cdot\frac{\Phi(u)-\Phi(d)}{u-d}.$$

**Bemerkung 143.** Die W'keiten $p_u,p_d$ tauchen hier nicht auf sondern nur $q_u$, $q_d$-

$\rightarrow$ Die Preise sind so gestaltet, als würden wir in einer Risiko neutraen Welt leben.

**Beispiel 144.** Sei $s=100, R=0, u=1.2, d=0.8$

$$p_u=0.6, p_d=0.4$$

Preisprozess:

$$S_0=100$$

$$S_1=\begin{cases}120 & \text{Mit } W'\text{keit } 0.6\\ 80 & \text{Mit } W'\text{keit } 0.4\end{cases}.$$

Diskontierter erwarteter Wert vom Stock

$$\frac{1}{R+1}\mathbb{E}_p[S_1]=1\cdot\{120\cdot 0.6+80\cdot 0.4\}=104$$

$\Rightarrow$ größer als $S_0 = 100$, d.h. der Markt ist Risiko-Scheu.

Unser Martk ist Arbitrage-frei: $u \geqslant 1 + R \geqslant d$.

Betrachten wir eine European Call Option mit Strike Price $K = 110$, so ist das Derivat

$$X = \begin{cases} 10 & \text{falls } S_1 = 120 \\ 0 & \text{falls } S_1 = 80 \end{cases}.$$

Wenn wir den Preis von $X$ als diskontierten erwarteten Wert unter den W'keiten:

$$\frac{1}{1+0}\{10 \cdot 0.6 + 0 \cdot 0.4\} = 6$$

Wenn wir aber die Martigalwahrscheinlichkeiten benutzen:

$$q_u = q_d = 0.5$$

$\Rightarrow \Pi(0, X) = \frac{1}{1+0} = (10 \cdot 0.5 + 0 \cdot 0.5) = 5 \neq 6$

Wie sieht das hedging Portfolio $h = (x, y)$ aus?

$$x = \frac{1.2 \cdot 0 + 0.8 \cdot 10}{1.2 - 0.8} = -20$$

$$y = \frac{1}{100} \cdot \frac{10 - 0}{1.2 - 0.8} = \frac{1}{4}$$

d.h. wir leihen 20€ und investieren in $\frac{1}{4}$Stock.

$$\Rightarrow V_0^h = \frac{1}{4} \cdot 100 - 20 = 5$$

$$V_1^h = \begin{cases} \frac{1}{4}120 - 20 = 10 & \text{wenn } S_1 = 120 \\ \frac{1}{4}80 - 20 = 0 & \text{wenn } S_1 = 80 \end{cases} = X$$

Wenn jemand bereit ist, die Option für 6€ zu kaufen $\Rightarrow$

Verkaufen: 6€

Investieren: 5€ in hedging Portfolio h.

Zur Zeit $t = 1$ ist das Derivat und das hedging Portfolio gleichwertig, und wir haben 1€ übrig.

# 10 Das Multiperioden Modell (nur kurz)

Zeit $t \in \{0, 1, \ldots, T\}$

Zwei Assets: $\begin{cases} \text{Bond} & B_t \\ \text{Stock} & S_t \end{cases}$

Bond Dynamik: $B_{t+1} = (1+R)B_t, B_0 = 1$

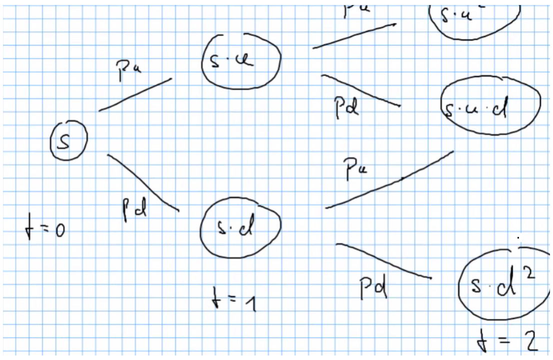
Stock Dynamik: $S_{t+1} = S_t Z_t, S_0 = s,$

Wobei $Z_0, \ldots, Z_T$ u.i.v. mit

$$\mathbb{P}[Z = u] = p_u$$

$$\mathbb{P}[Z = d] = p_d$$

mit $d \leqslant R+1 \leqslant u$.




Portfolio-Strategie: $h_t = (x_t + y_t)_{t=0,1,\ldots,T}$

Value Process: $V_t^h = x_t(R+1) + y_t S_t$

Derivat: $X = \Phi(S_T)$:

Beispiel European Call Option im Gegensatz dazu American Call Option

$$X = \Phi(S_0, \ldots, S_T)$$

Preis Prozess: $\{\Pi(t, X), t = 0, \ldots, T\}$?

$X$ ist erreichbar wenn ein hedging Portfolio existiert mit $V_T^h = X$ mit Wkeit 1.

**Satz 145.** *Nehmen wir an, dass $X$ durch $h$ erreichbar ist. Außerdem ist es möglich zu einer Zeit $t$ $X$ zu kaufen für einen billigeren Preis als $V_t^h$. Dann ist es möglich einen Arbitrage Profit zu erzielen.*

$\Rightarrow$ *Vernünftiger Preis Prozess* $\Pi(t, X) = V_t^h, t = 0, \ldots, T$

**Bemerkung 146.** Im Binomialmodell (arbitrage-frei) ist jedes Derivat hedgebar.

$\rightarrow$Dieses Modell (welches des Wert von X beschreibt) heißt auch Cox – Ross – Rubinstein:

$$\Pi(0, X) = \frac{1}{(R+1)^T} \sum_{k=0}^{T} \binom{T}{k} q_u^k q_d^{T-k} \Phi(s \cdot u^k d^{T-k})$$

$\Rightarrow$ Für European Call Option: $\Phi = \max(0, S_T - K)$

Und weiter?

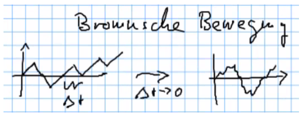
„Limes nehmen zur stetigen Zeit“

$\rightarrow$Black – Scholes – Modell

Bond: $\mathrm{dB}_t = \mathrm{rB}_t \mathrm{dt} \Rightarrow B_t = B_0 e^{\mathrm{rt}}$

$r$ instantaneous interest rate (Zinssatz)

Stock: $\mathrm{dS}_t = \underbrace{\mu}_{\text{Drift}} S_t \mathrm{dt} + \underbrace{\sigma}_{\text{Volatilität}} S_t \underbrace{\mathrm{dW}_t}_{\text{Brownsche Bewegung}}$




$S_0 = s$

$\mu > 0 \Rightarrow$ Erwartungswert steigt

$\mu = 0 \Rightarrow$ Erwartungswert konstant $\Rightarrow$ Martingal

$\sigma = 0 \Rightarrow$ Exponentialfunktion

$\xrightarrow{\text{Itô}-\text{Formel}} S_t = \mathrm{s}e^{(\mu - \sigma^2/2)t + \sigma W_t}$

Preis einer Call Option $X$ mit Strike Price $K$ und exercise date $T$:

$$\Pi(0, X) = s\varphi(v) - \mathrm{Ke}^{-\mathrm{rT}}\varphi(v - \sigma\sqrt{T}) \colon \mathrm{Black} - \mathrm{Scholes} - \mathrm{Formel}$$

$v = \frac{\log(s/K) + \mathrm{rT}}{\sigma\sqrt{T}} + \frac{\sigma\sqrt{T}}{2}$

$\varphi(x) = \int_{-\infty}^{x} e^{-y^2/2} / \sqrt{2\pi}\,\mathrm{dy}$

Ende Vorlesung 27

---

# 11 Bonus: Interessantes Wissen aus den Blättern

## 11.1 Blatt 01

**Bemerkung 147.** (Momentenmethode):

- Gegeben seien $X_1, \ldots, X_n$ u.i.v. Zufallsvariablen.
- Wir kennen die Verteilung, gesucht sind Schätzungen von einem oder mehreren Parametern.
- Bestimmen der ersten k theoretischen Momente $\mathbb{E}(X^j)$ der Verteilung und ermitteln einen Zusammenhang zu den gesuchten Parametern.
- Ersetzen der theoretischen Momente durch $\frac{1}{n}\sum X_i^j$ und lösen nach den gesuchten Parametern auf.

## 11.2 Blatt 03

Es sei $(\mathcal{X}, \mathcal{F}, \mathbb{P}_\vartheta: \vartheta \in \theta)$ ein stat. Modell.

**Definition 148.** *Wir nennen eine Statistik $T: \mathcal{X} \to \Sigma$ mit abzählbarem Wertebereich $\Sigma$:*

a) *suffizient, falls für alle $s \in \Sigma$ eine Verteilung $Q_s$ auf $(\mathcal{X}, \mathcal{F})$ existiert, s.d.*

$$\mathbb{P}_\vartheta[\cdot|T = s] = Q_s$$

*für alle $\vartheta \in \theta$ mit $\mathbb{P}_\vartheta[T = s] > 0$ und*

b) *vollständig, falls keine nicht identisch verschwindende Funktion $g: \Sigma \to \mathbb{R}$ existiert, s.d. $\mathbb{E}_\vartheta[g(T)] = 0$ für alle $\vartheta \in \theta$.*

Für eine eine relle Schätzfunktion $\tau$:

**Satz 149.** *(Satz von Rao-Backwell)*

*Es sei $T$ suffizient und $S$ ein Erwartungstreuer Schätzer für $\tau$. Wir definieren $g_S: \Sigma \to \mathbb{R}$ durch $g_S(s) := \mathbb{E}_{Q_s}[S]$. Der Schätzer $g_S(T)$ ist erwartungstreu für $\tau$ und für alle $\vartheta \in \theta$ gilt:*

$$\mathrm{Var}_\vartheta[g_S(T)] \leqslant \mathrm{Var}_\vartheta[S].$$

**Satz 150.** *(Satz von Lehmann-Scheffè)*

*Es sei $g: \Sigma \to \mathbb{R}$ gegeben und $T$ sei suffizient und vollständig. $g(T)$ ist ein bester Schätzer für $\tau$, falls $g(T)$ erwartungstreu für $\tau$ ist.*

**Beispiel 151.** Dichten in Exponentiellen Familien:

- $f_\vartheta(x) = \frac{\vartheta^x}{x!}\exp(-\vartheta), x \in \mathbb{N}_0$ und $\vartheta > 0$.
- Inverse Gammaverteilung 1: $f_\alpha(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{(-\alpha+1)} \exp(-\beta/x) \mathbb{1}_{(0,\infty)}(x)$, wobei $\beta > 0$ bekannt, $x \in \mathbb{R}$ und $\alpha > 0$.
- Inverse Gammaverteilung 2: $f_\alpha(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{(-\alpha+1)} \exp(-\beta/x) \mathbb{1}_{(0,\infty)}(x)$, wobei $\alpha > 0$ bekannt, $x \in \mathbb{R}$ und $\beta > 0$.

Kein Element einer Exponentiellen Familie:

- $f_\vartheta(x) = \exp(-2\log(\vartheta) + \log(2x)) \mathbb{1}_{(0,\vartheta)}(x)$, wobei $x \in \mathbb{R}, \vartheta > 0$.
- Laplace-Verteilung: $f_\vartheta(x) = \frac{1}{2}\exp(-|x - \vartheta|)$, wobei $x, \vartheta \in \mathbb{R}$.

## 11.3 Blatt 05

## 11.4 Blatt 06: Zusammenhang zwischen Konfidenzbereichen und Tests

Für ein stat. Modell $(\mathcal{X}, \mathcal{F}, \mathbb{P}_\vartheta: \vartheta \in \theta)$ gilt

- Ist $C: \mathcal{X} \to \mathcal{P}(\theta)$ ein Konfidenzbereich für $\vartheta$ zum Niveau $1 - \alpha$ und $\vartheta_0 \in \theta$, so ist $K := \{x \in \mathcal{X}: \vartheta_0 \notin C(x)\}$ der kritische Bereich eines Tests von der Hypothese $\theta_0 = \{\vartheta\}$ gegen die Alternative $\theta_1 = \{\vartheta: \vartheta \neq \vartheta_0\}$ zum Niveau $\alpha$.
- Es sei für jedes $\vartheta \in \theta$ $K_\vartheta$ der kritische Bereich eines Tests von der Hypothese $H_0$ mit $\theta_0 = \{\vartheta\}$ gegen die Altenative $H_0$ mit $\{\tilde{\vartheta}: \tilde{\vartheta} \neq \vartheta_0\}$ zum Niveau $\alpha$. Dann ist $C = \{(x, \vartheta) \in \mathcal{X} \times \vartheta: x \in K_\vartheta^c\}$. ein Konfidenzbereich zum Niveau $1 - \alpha$ für $\vartheta$.