

1 Einleitung

Hier fehlt das erste Kapitel, was allerdings nur motivierend sein soll. Stattdessen ein Kommentar:

Diese Mitschrift enthält weniger / schlechtere Bilder als die Mitschrift von Dr. Kopfer, allerdings befindet sich in Kapitel 7 interessantes Wissen / Beispiele aus den Übungsblättern. Das hier ist weitestgehend live während der VL mitgeschrieben worden. D.h. Fehler sind zu erwarten. Wer solche findet kann mir diese gerne an mh@mssh.dev schicken.

Viele Grüße, Manuel

2 Statistische Modelle

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

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

Wieso \mathcal{X} und nicht Ω ? \mathcal{X} ist das Bild eines Zufallsexperiments $\mathcal{X}: \Omega \rightarrow \mathcal{X}$.

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

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

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

Bemerkung 3. Wenn man $(\mathbb{P}_\vartheta: \vartheta \in \theta)$ 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}_ϑ eine Zähl-dichte: $\zeta_\vartheta: x \mapsto \mathbb{P}_\vartheta[\{x\}]$.

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

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

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

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

ist das zugehörige *n-fache Produktmodell*.

$X_k: \mathcal{X} \rightarrow E$ ist die k -te Koordinate und beschreibt den Ausgang des k -ten Experiments. Insbesondere sind X_1, \dots, X_n u.i.v. (unabhängig und identisch verteilt) bzgl. \mathbb{P}_ϑ mit Verteilung \mathbb{Q}_ϑ .

3 Schätzer

Sei $(\mathcal{X}, \mathcal{F}, (\mathbb{P}_\vartheta: \vartheta \in \theta))$ stat. Modell und (Σ, ζ) ein Ereignisraum.

Eine bel. Zufallsvariable

$$\delta: (\mathcal{X}, \mathcal{F}) \rightarrow (\Sigma, \zeta)$$

heißt *Statistik*.

Sei $\tau: \theta \rightarrow \Sigma$ eine Abbildung. $\tau(\vartheta) \in \Sigma$ heißt **Kenngroße**. Eine Statistik $T: \mathcal{X} \rightarrow \Sigma$ 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 τ zugeschnitten ist.
- iii. Was hat T mit τ zu tun? Es gibt nicht nur einen Schätzer T für $\tau(\vartheta)$. Daher ist es nicht formalisiert um nicht zu restriktiv zu sein.
- iv. Man spricht auch von **Punktschätzern** um von Bereichsschätzern abzugrenzen. (Kapitel: Konfidenzbereiche)

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

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

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

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

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

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

Aus dem Gesetz der großen Zahlen

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

aber

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

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

- Was sind Qualitätskriterien?
- Wie Schätzer finden?

3.1 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. Standardmodell. Die **Likelihoodfunktion** ist

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

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

wobei ρ_ϑ die Dichtefunktion von \mathbb{P}_ϑ ist.

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

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

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

Definition 8. Ein Schätzer $T: \mathcal{X} \rightarrow \theta$ für ϑ 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 ρ_x auf θ .

Beispiel 9. (Schätzung von Erfolgswahrscheinlichkeit)

Sei ϑ der Wirkungsgrad eines Medikaments.

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

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

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

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

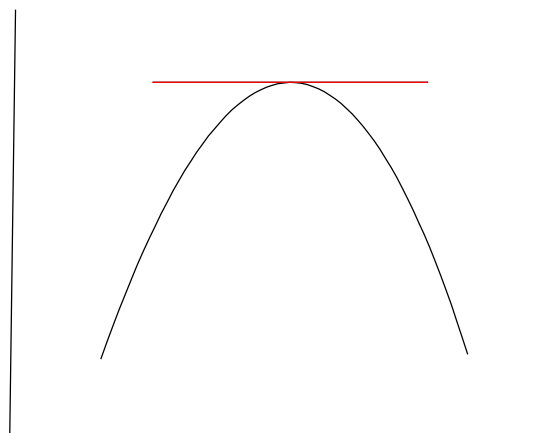
Was ist der M-L-Schätzer ?

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

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

Also $x = n\vartheta$.

Maximum ? Ja weil für $\vartheta \leq \frac{x}{n}$ ist ρ_x wachsend, für $\vartheta \geq \frac{x}{n}$ ist ρ_x fallend.



„unimodal“

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

Beispiel 10. (Physikalische Messungen)

In jeder physikalischen Messung gibt es Messfehler.

Annahme:

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

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

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

Der M-L-Schätzer für (m, σ^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)

3.2 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 $\tau: \theta \rightarrow \mathbb{R}$ eine reelle Kenngröße.

Ein Schätzer $T: \mathcal{X} \rightarrow \mathbb{R}$ für τ 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 reellen Produktmodellen)

Sei $n \geq 2$ und $\mathcal{M} = (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n)) \dots$

Sei $m(\vartheta) := \mathbb{E}_{\vartheta}[X]$ und $\nu(\vartheta) := \text{Var}_{\vartheta}[X]$ 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, ν) .

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

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

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

$$\begin{aligned} \mathbb{E}_{\vartheta}[V] &\stackrel{\text{lin.}}{=} \frac{1}{n} \sum_{k=1}^n \mathbb{E}[X_k - M]^2 \stackrel{\mathbb{E}_{\vartheta}[X_k - M] = 0}{=} \frac{1}{n} \sum_{k=1}^n \text{Var}[X_k - M] \\ &\stackrel{X_{ki.i.d.}}{=} \text{Var}_{\vartheta} \left[X_1 - \frac{1}{n} \sum_{k=1}^n X_k \right] = \text{Var}_{\vartheta} \left[X_1 \cdot \frac{n-1}{n} - \frac{1}{n} \sum_{k=2}^n X_k \right] \\ &\stackrel{X_{ki.i.d.}}{=} \left(\frac{n-1}{n} \right)^2 \text{Var}_{\vartheta}[X_1] + \left(\frac{1}{n} \right)^2 \text{Var}_{\vartheta}[X_1] = \left(\frac{n-1}{n} \right)^2 \nu(\vartheta) + \frac{n-1}{n^2} \nu(\vartheta) = \frac{n-1}{n} \nu(\vartheta) \end{aligned}$$

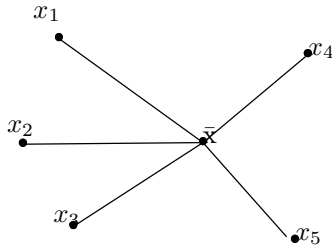
$$\Rightarrow \mathbb{E}_{\vartheta}[V^*] = \nu(\vartheta).$$

□

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!!

Erwartungstreue ist wünschenswert, aber nicht immer „besser“.

Beispiel 14. (Vorsetzung Binomialmodell)

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

$T(x) = \frac{x}{n}$ ist ML-Schätzer für θ .

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

Anderer Schätzer

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

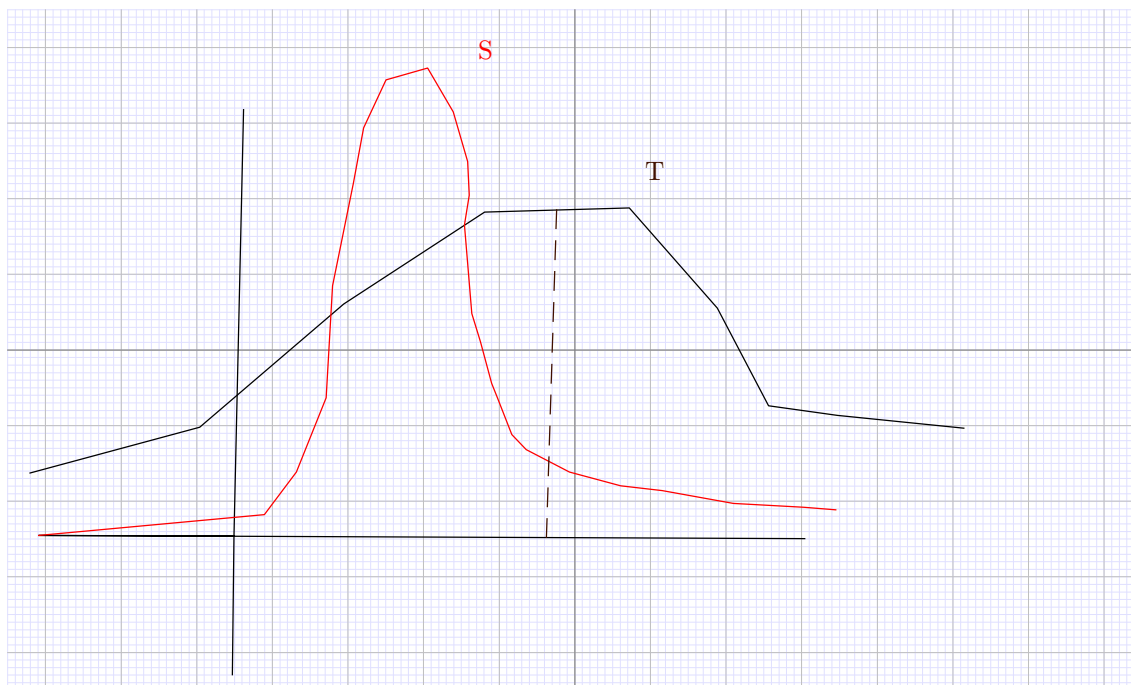
$$\mathbb{B}_\vartheta[S] = \frac{n\vartheta + 1}{n+2} - \vartheta = \frac{1-2\vartheta}{n+2} > 0$$

Aber was ist mit der **mittleren quadratischen Abweichung?**

Definition 15. Der **mittlere quadratische Fehler** eines Schätzers T für τ ist

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

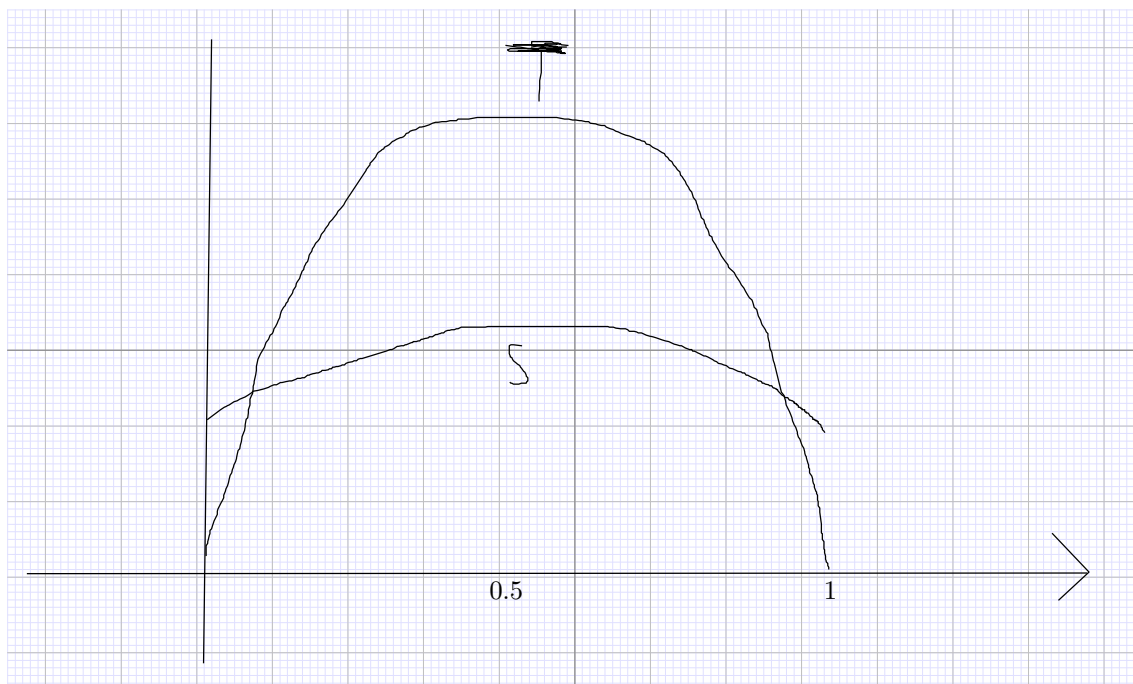
Wir wollen beide Terme gleichzeitig minimieren.



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

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

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



Für Zentrale Werte von ϑ : S ist besser als T.

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

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

3.3 Konsistenz von Schätzern

Ein weiteres Qualitätskriterium ist die **Konsistenz**.

- Sei $\mathcal{M}=(\mathcal{X}, \mathcal{F}, (\mathbb{P}_{\vartheta}))$ ein stat. Modell und $\tau: \theta \rightarrow \mathbb{R}$ eine reelle Kenngröße.
- Wiederholung der Messung: Sei $(X_n)_{n \geq 1}$ 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 $n \geq 1$ $T_n: \mathcal{X} \rightarrow \mathbb{R}$ ein Schätzer für τ

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

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

oder:

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

„Konvergenz im Maß (Stochastische Konvergenz)“

Im folgenden:

Standardfall 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:

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

die Erwartungstreuen Schätzer für m bzw. v .

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

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

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

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

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

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

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

Auch M-L-Schätzer sind konsistent:

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

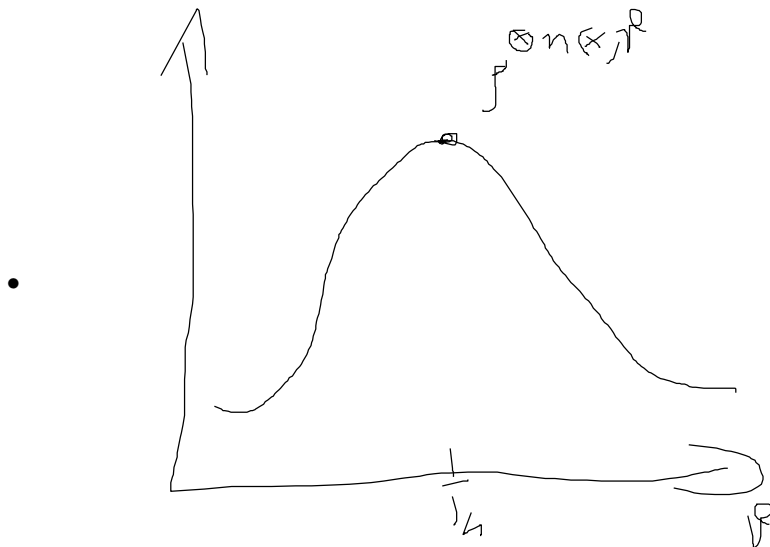
Sei $(E, \mathcal{E}, \mathbb{Q}_\vartheta)$ eine **inparametrisches** Standardmodell (d.h. $\theta \subseteq \mathbb{R}$), mit Likelihood-Funktion ρ .

Es gelte:

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

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

unimodal, d.h. \exists ML-Schätzer $T_n: E^n \rightarrow \mathbb{R}$ 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 ϑ .

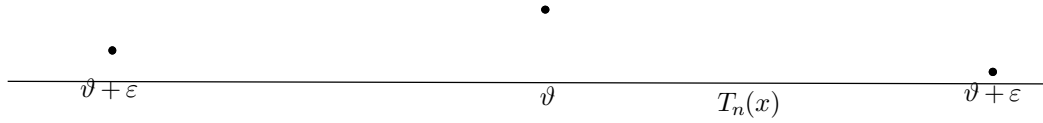
Beweis. (grob)

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

$$\mathbb{P}_\vartheta[\vartheta - \epsilon \leq T_n \leq \vartheta + \epsilon] \xrightarrow{n \rightarrow \infty} 1$$

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

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



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

„+“-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 $\frac{1}{n} \sum_{i=1}^n X_i \rightarrow \mathbb{E}[X_1]$**)

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

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

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.

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

„-“ Fall genau so...

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

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

$$\subseteq \left\{ x: \log \left(\frac{\rho_{\vartheta}^{\otimes n}(x)}{\rho_{\vartheta \pm \varepsilon}^{\otimes n}(x)} \right) > 0 \right\} \subseteq \{x: \vartheta - \varepsilon \leq T_n(x) \leq \vartheta + \varepsilon\}$$

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

□

3.4 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

$$\text{Var}_\vartheta[T] \leq \text{Var}_\vartheta[S] \forall \vartheta \in \theta.$$

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

- i. θ ist ein offenes Intervall in \mathbb{R} .
- ii. Die Likelihoodfunktion ρ ist auf $\mathcal{X} \times \theta$ strikt positiv und nach ϑ stetig differenzierbar.
- iii. Für jedes $\vartheta \in \theta$ ex. die Varianz:

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

und ist nicht 0. Außerdem gilt die Vertauschungsregel:

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

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**. $I(\vartheta) = \text{Var}_\vartheta[U_\vartheta]$

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

$$\begin{aligned} \mathbb{E}[U_\vartheta] &= \int_{\mathcal{X}} \frac{d}{d\vartheta} \log \rho(x, \vartheta) \rho(x, \vartheta) dx \\ &= \int_{\mathcal{X}} \frac{d}{d\vartheta} \rho(x, \vartheta) dx = \frac{d}{d\vartheta} \underbrace{\int \rho(x, \vartheta) dx}_1 = 0 \end{aligned}$$

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

- 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 θ_0 . Also kann keine Beobachtung die Parameter in θ_0 unterscheiden.

I ist additiv für unabhängige Beobachtungen

Satz 22. Sei $(\mathcal{X}, \mathcal{F}, (\mathbb{P}_\vartheta: \vartheta \in \theta))$ ein reguläres 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

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

und

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

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

□

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

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

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

i. Es gilt

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

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

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

$$\text{Beweis. (i)} \quad \text{Cov}_{\vartheta}[T, U_{\vartheta}] := \mathbb{E}[T \cdot U_{\vartheta}] - \mathbb{E}[T]\mathbb{E}[U_{\vartheta}] \stackrel{\mathbb{E}[U_{\vartheta}]=0}{=} \mathbb{E}[T \cdot U_{\vartheta}]$$

$$\begin{aligned} &= \int_{\mathcal{X}} T(x) U_{\vartheta}(x) \rho(x, \vartheta) dx = \int_{\mathcal{X}} T(x) \frac{d}{d\vartheta} \rho(x, \vartheta) dx \\ &= \frac{d}{d\vartheta} \mathbb{E}_{\vartheta}[T] \stackrel{T \text{ erwartungstreu}}{=} \tau'(\vartheta) \end{aligned}$$

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

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

(ii)

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

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

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

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

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

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

Also

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

Unbestimmte Integration in ϑ liefert

$$\log \rho(x, \vartheta) - \underbrace{h(x)}_{\text{Integrationskonstante}} = a(\vartheta)T(x) - \underbrace{b(\vartheta)}_{= \int \frac{I(\tilde{\vartheta})}{\tau'(\tilde{\vartheta})} \tau(\tilde{\vartheta}) d\tilde{\vartheta}}$$

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

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

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

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

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

Warum gilt $(*)$?

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

$$\begin{aligned} b'(\vartheta) &= \frac{a'(\vartheta) \int T(x) e^{a(\vartheta)T(x)} h(x) dx}{\int e^{a(\vartheta)T(x)} h(x) dx} = \frac{a'(\vartheta) \int T(x) e^{a(\vartheta)T(x) - b(\vartheta)} h(x) dx}{\underbrace{\int e^{a(\vartheta)T(x) - b(\vartheta)} h(x) dx}_{=1}} = a'(\vartheta) \mathbb{E}_{\vartheta}[T] = a'(\vartheta) \tau(\vartheta) \\ &= \frac{I(\vartheta)}{\tau'(\vartheta)} \tau(\vartheta) \end{aligned}$$

ad(**) Wann gilt Gleichheit ?

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

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

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

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

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

□

Bemerkung 24. Wenn T erwartungstreu regulärer Schätzer, s.d. Gleichheit in Cramér-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 einparametrisches Standardmodell mit θ offen. Wenn die Likelihoodfkt. der Form

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

mit Funktionen $a: \theta \rightarrow \mathbb{R}, a' \neq 0$

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

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

Beispiel 26. (Poisson-Verteilung)

\mathbb{P}_ϑ hat die Dichte ...

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

Da T erwartungstreu ist, ist T ein bester Schätzer für ϑ .

Ende Vorlesung 4

KEINE OFFIZIELLEN MUSTERLÖSUNGEN ZU DEN ÜBUNGSBLÄTTERN!

Proposition 27. (*Eigenschaften von exponentiellen Modellen*)

- a) $b(\vartheta)$ ist auf θ stetig diff'bar mit $b'(\vartheta) = a'(\vartheta) \mathbb{E}_\vartheta[T]$ (Insbesondere existiert der Erwartungswert von T).
- b) Jede Statistik $S: \mathcal{X} \rightarrow \mathbb{R}$ mit existierendem $\mathbb{E}_\vartheta[S]$ ist regulär. Insbesondere sind M und T regulär und $\tau(\vartheta) := \mathbb{E}_\vartheta[T]$ ist stetig diff'bar mit $\tau'(\vartheta) = a'(\vartheta) \cdot \text{Var}_\vartheta[T] \neq 0 \forall \vartheta \in \theta$.
- 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 $I(\vartheta) = a'(\vartheta) \tau'(\vartheta)$

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

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

Beweis. (Von Prop. 27)

Wir nehmen an, dass $a(\vartheta) = \vartheta$ (Da $a'(\vartheta) \neq 0$ folgt die allg. Aussage mit Kettenregel).

(Sonst $\left(\tilde{\vartheta} = a(\vartheta) \Rightarrow \frac{d}{d\tilde{\vartheta}}(\dots) = a'(\vartheta) \frac{d}{d\vartheta}(\dots) \right)$)

Sei S in \mathcal{L}^1

Sei $u_S(\vartheta) := e^{b(\vartheta)} \mathbb{E}_\vartheta[S] = \int_{\mathcal{X}} S(x) h(x) e^{\vartheta T(x)} dx$.

$u_S(\vartheta)$ ist (reell)-analytisch in ϑ , denn für $\vartheta + t \in \theta$ und $a_k = \int_{\mathcal{X}} \frac{S(x) h(x) T(x)^k}{k!} e^{\vartheta T(x)} dx$

gilt

$$\begin{aligned}
(*) \sum_{k=0}^{\infty} |a_k| |t|^k &\leq \sum_{k=0}^{\infty} \frac{|t|^k}{k!} \int_{\mathcal{X}} |S(x)| h(x) |T(x)|^k e^{\vartheta T(x)} dx \\
&\stackrel{\text{mon. Konv.}}{=} \int_{\mathcal{X}} |S(x)| h(x) e^{\vartheta T(x)} \underbrace{\sum_{k=0}^{\infty} \frac{|t|^k}{k!} |T(x)|^k}_{\exp(|tT(x)|)} dx \\
&= \int_{\mathcal{X}} |S(x)| h(x) \underbrace{e^{\vartheta T(x) + |tT(x)|}}_{\substack{e^{\vartheta T(x) + |tT(x)|}, & tT(x) > 0 \\ e^{\vartheta T(x) - |tT(x)|}, & tT(x) < 0}} dx \\
&\leq \int_{\mathcal{X}} |S(x)| h(x) \underbrace{e^{(\vartheta+t)T(x)}}_{e^{b(\vartheta)\rho_{\vartheta} + t(x)}} dx + \int_{\mathcal{X}} |S(x)| h(x) \underbrace{e^{(\vartheta-t)T(x)}}_{e^{b(\vartheta)\rho_{\vartheta} - t(x)}} dx < \infty
\end{aligned}$$

Da $S \in \mathcal{L}^1(\mathbb{P}_{\vartheta \pm 1})$.

$$\Rightarrow u_S(\vartheta + t) = \int_{\mathcal{X}} S(x) h(x) e^{(\vartheta+t)T(x)} dx \stackrel{(**)}{=} \sum_{k=0}^{\infty} t^k a_k.$$

$$(**) \text{ gilt, da } e^{tT(x)} = \sum_{k=0}^{\infty} t^k \frac{T(x)^k}{k!}.$$

und Summe und Integral vertauscht werden dürfen wegen (*).

Also ist $u_S(\vartheta)$ analytisch und ist seine Taylorreihe, d.h.

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

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

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)

$$\begin{aligned}
\tau'(\vartheta) &= b''(\vartheta) = \frac{u''_1(\vartheta)}{u_1(\vartheta)} - \left(\frac{u'_1(\vartheta)}{u_1(\vartheta)} \right)^2 \\
&= \mathbb{E}_{\vartheta}[T^2] - (\mathbb{E}_{\vartheta}[T])^2 = \text{Var}[T]
\end{aligned}$$

Allgemeine S (Regularität)

$$\begin{aligned}
\frac{d}{d\vartheta} \int_{\mathcal{X}} S(x) \rho(x, \vartheta) dx &= \frac{d}{d\vartheta} \mathbb{E}_{\vartheta}[S] = \frac{d}{d\vartheta} [e^{-b(\vartheta)} u_S(\vartheta)] \\
&= (u'_S(\vartheta) - u_S(\vartheta) b'(\vartheta)) e^{-b(\vartheta)} = \mathbb{E}_{\vartheta}[\text{ST}] - \mathbb{E}_{\vartheta}[S] \underbrace{\mathbb{E}_{\vartheta}[T]}_{=: \tau(\vartheta)} \\
&= \mathbb{E}_{\vartheta} \left[S \left(\underbrace{T - \tau(\vartheta)}_{=(a'(\vartheta)) U_{\vartheta}: \text{Satz 23 (ii)}} \right) \right] = \mathbb{E}_{\vartheta}[\text{SU}_{\vartheta}] \\
&= \int_{\mathcal{X}} S(x) \frac{d}{d\vartheta} \rho(x, \vartheta) dx
\end{aligned}$$

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

$\Rightarrow T$ regulär.

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

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

(*) \mathcal{M} regulär:

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

□

Beispiel 29. (Binomialverteilung)

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

$$T(x) = \frac{x}{n} \text{ (ML-Schätzer)}$$

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

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

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

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

Bemerkung 30. (Produktmodelle)

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

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

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

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

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

□

3.5 Bayes-Schätzer

\mathcal{M} Standardmodell.

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

$L(\vartheta, T)$ eine "Verlustfunktion"

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

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 ϑ nicht unbed. gleichhäufig sind, aber haben eine Verteilungsdichte $a(\vartheta)$ (a priori Verteilung).

Definition 31.

i. Das Bayesrisiko des Schätzers T bzgl. α und L ist gegeben durch

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

ii. Ein Schätzer T heißt Bayes-Schätzer von $\tau(\vartheta)$ bzgl. α 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 ϑ gemäß der Dichte α und dann zieht man x gemäß der Likelihoodfunktion $\rho(x, \vartheta)$.
- b) Wenn x gezogen ist, verändert dies die Information über α . 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. α ist gegeben durch

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

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

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

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

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

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

$\Rightarrow T$ ist Bayes Schätzer. Gleichheit gilt genau dann wenn $S(x) = T(x)$ ρ_{α} f.s.

□

Beispiel 33. (Auto Versicherung)

ϑ = Schadenshäufigkeit pro Jahr.

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

Nach n Jahren hat der Kunde x Schaden produziert.

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

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 ...)

4 Konfidenzbereiche

Beispiel 34. Betrachten wir das Binomialmodell

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

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

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

\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 Erfolg \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, Σ eine bel. Menge, $\tau: \theta \rightarrow \Sigma$ eine unbekannte Größe und $0 < \alpha < 1$.

Eine Abbildung

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

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

heißt Konfidenzbereich für τ zum Irrtumsniveau α , 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 α möglichst klein. Diese zwei Wünsche konkurrieren. 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 ϑ in 90% der Fälle in dem Intervall liegt: ϑ 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_ϑ eine Untermenge s.d.

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

und C_ϑ so klein wie möglich. (z.B. Standardmodell)

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

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 τ zum Irrtumsniveau α .

Beispiel 37. (Schätzung mittlere Lebenszeit von radioaktiven Zerfall)

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

Sei $x > 0$ Messung.

Für geg. ϑ suchen wir C_ϑ s.d.

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

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

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

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

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

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

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

Ende Vorlesung 6

Wiederholung des Beispiels 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**)

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

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

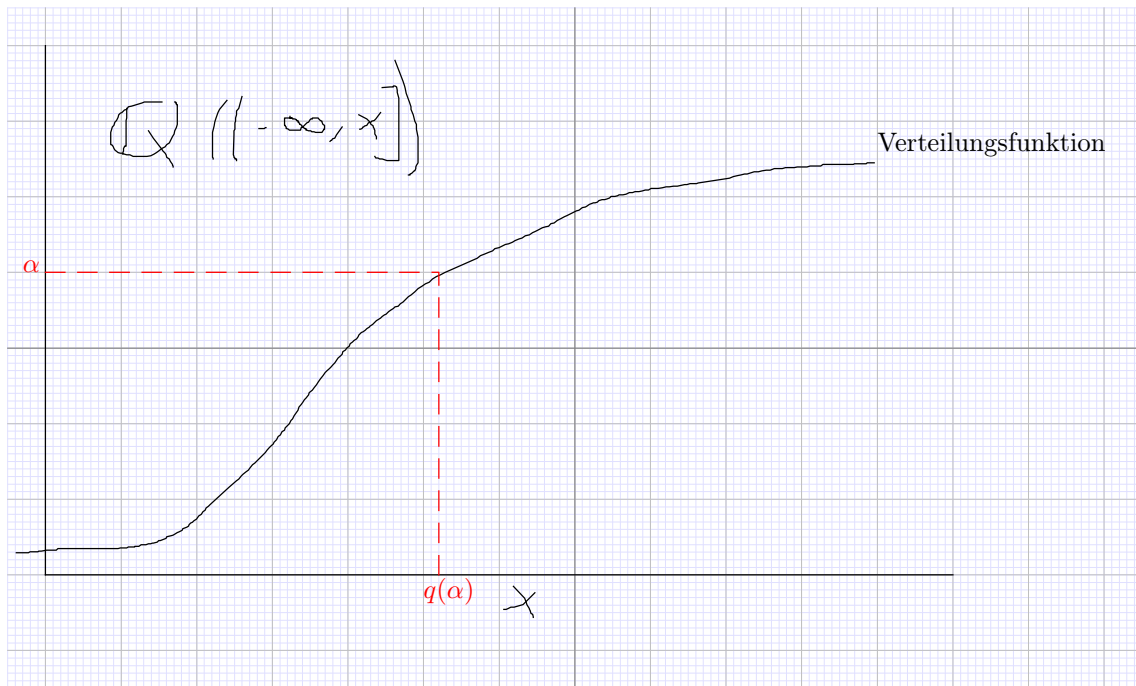
$$\rho_\vartheta^{(n)}(X) = \gamma_{\frac{1}{\vartheta}, n}(x) = \left(\frac{x}{\vartheta}\right)^{n-1} \frac{e^{-x/\vartheta}}{(n-1)! \vartheta}, x > 0.$$

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

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

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

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



Im abs. stetigen Fall $\mathbb{Q}((-\infty, q]) = \alpha = \int_{-\infty}^q \rho(x) dx$. 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 $\mathcal{S}_{m,v} := \left(\frac{X_k - m}{\sqrt{v}} \right)_{k=1, \dots, m}$. Dann, da $X_k \sim m + \sqrt{v} Z_k, Z_k \sim \mathcal{N}(0, 1)$

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

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

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

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

\Rightarrow Deshalb

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

Welche Verteilung hat Q ?

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

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

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

□

Ende Vorlesung 7

Für gegebenes α , 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)

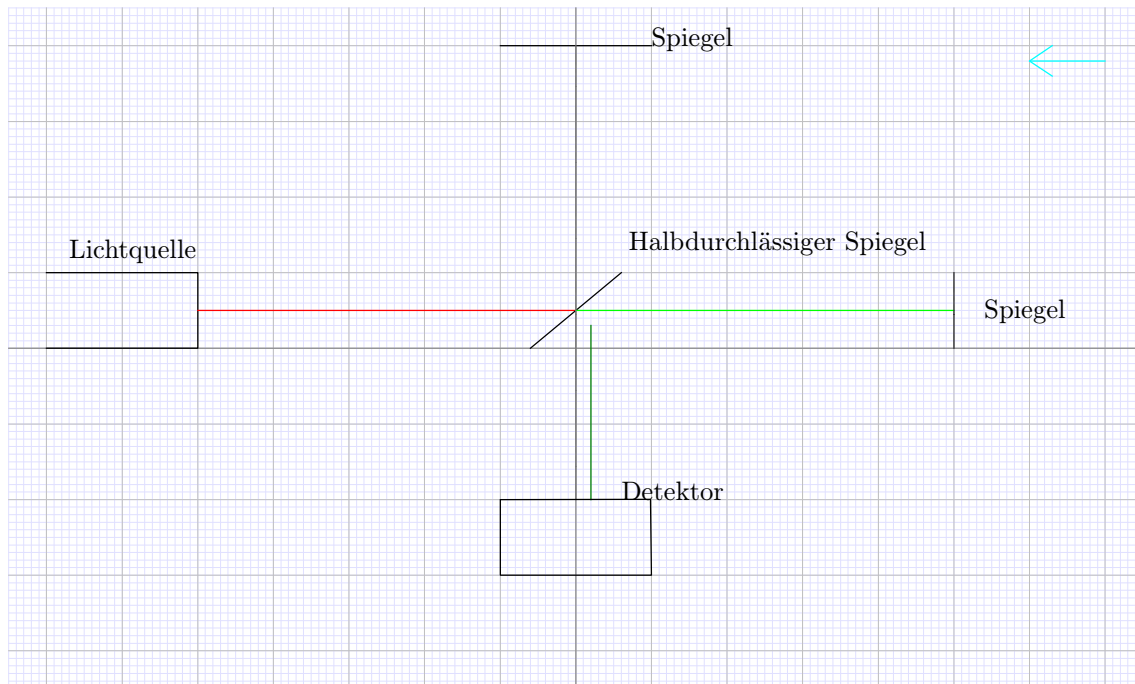
Sei das n -Produktmodell mit unbekanntem Erwartungswert m und unbekannter Varianz v . Sei

$x^* := F_Q^{-1}(1 - \alpha/2)$, mit $Q \sim t_{n-1}$ ($F_Q(x) = [(-\infty, x])$).

$\Rightarrow C(x) = (M - x^* \sqrt{V^*/n}, M + x^* \sqrt{V^*/n})$ ein Konfidenzintervall für m zum Irrtumsniveau α .

Anwendung: (Bestimmung der Lichtgeschwindigkeit von Michelson und Morely)

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



Annahme: Messungen sind Normalverteilt mit unbekannten m, v .

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

Lösung: Anwenden des Satzes:

1. $x^* = (1 - \alpha/2) - \text{Quantil von } t_{n-1} \text{ mit } n = 20.$
 ≈ 2.54

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

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

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

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

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

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

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

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

Tatsächlich: 299792

Konfidenzintervalle im Binomialmodell

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

Ges.: Konfidenzintervall für ϑ

3 Methoden:

1. Tchebyschev:

Der beste Schätzer für ϑ 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

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

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

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

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

2. Normalapproximation

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

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

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

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

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

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

$$\text{Für } \alpha = 0.05, n = 1000 \Rightarrow \frac{1}{\sqrt{4n}} \Phi^{-1}(1-\alpha/2) = 1.96$$

$$\Rightarrow \varepsilon \geq 0.03$$

3. Verwendung von Quantilen

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

a) $\forall \vartheta \in (0, 1): x \mapsto \text{Bin}(n, \vartheta)(\{x\})$ strikt wachsend auf $\{0, \dots, (n+1)\vartheta - 1\}$

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

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

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

Das benutzen wir nun als 3. Methode.

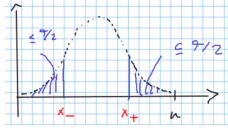


Abbildung 1.

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

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

$$\text{wobei } x_-(\vartheta) = \max \{x \in \mathcal{X} : \text{Bin}_{n,\vartheta}(\{0, \dots, x-1\}) \leq \alpha/2\}$$

$$x_+(\vartheta) = \min \{x \in \mathcal{X} : \text{Bin}_{n,\vartheta}(\{x+1, \dots, n\}) \leq \alpha/2\}$$

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

$$x \leq x_+(\vartheta) \Leftrightarrow \text{Bin}_{n,\vartheta}(\{x, \dots, n\}) \geq \alpha/2$$

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

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

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

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

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

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

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

$$\geq 1 - \alpha.$$

Ende Vorlesung 8

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

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

wobei

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

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

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

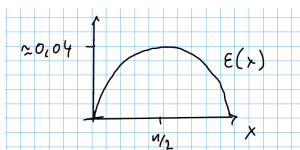


Abbildung 2.

Beweis. (voherigem Lemma)

a)

Sei $x \geq 1$

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

b)

Seien U_1, \dots, 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 $\text{Bin}(n, \vartheta)$.

$$\text{Bin}(n, \vartheta)(\{x, \dots, n\}) = \underbrace{\mathbb{P}[S_\vartheta \geq x]}_{\substack{\text{mind. } x \\ \text{der } U_i \text{ liegen in } [0, \vartheta]}}$$

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

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

$$\begin{aligned} &= n! \int_0^1 \dots \int_0^1 \mathbb{1}_{[0,\vartheta]}(t_x) \mathbb{1}_{\{t_1 < t_2 < \dots < t_n\}} dt_1 \dots dt_n \\ &= n! \int_0^\vartheta \left(\int_0^{t_x} \dots \int_0^{t_x} \mathbb{1}_{\{t_1 < t_2 < \dots < t_x\}} dt_1 \dots dt_{x-1} \right) \left(\int_{t_x}^1 \dots \int_{t_x}^1 \mathbb{1}_{\{t_x < \dots < t_n\}} dt_{x+1} \dots dt_n \right) dt_x \end{aligned}$$

□

Ende Vorlesung 9

5 Normalverteilung, χ^2 , t, F-Verteilung

Lemma 43. Sei $X \subseteq \mathbb{R}^n$ offen, \mathbb{P} ein W'maß auf $(X, B(X))$ mit Dichte ρ bzgl. \mathcal{L} .

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

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

$$\rho_T(y) = \rho(T^{-1}(y)) |\det DT^{-1}(y)| \quad \forall y \in Y$$

$$\textbf{Beweis.} \quad \mathbb{P} \circ T^{-1}(A) = \int_{T^{-1}(A)} \rho(x) dx \stackrel{\text{Trafo.}}{=} \int_A \rho(T^{-1}(y)) |\det DT^{-1}(y)| dy$$

□

Satz 44. Seien X_1, \dots, X_n u.i.v. $\sim \mathcal{N}(0, 1)$ und $X = (X_1, \dots, X_n)^t$.

Sei $B \in \mathbb{R}^{n \times n}$ eine reguläre ($\det B \neq 0$) Matrix und $m \in \mathbb{R}^n$. Dann $Y = BX + m$ 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 $\mathbb{E}[Y_k] = m_k, \text{Cov}[Y_k, Y_l] = C_{k,l} \quad 1 \leq k, l \leq n$.

Bemerkung 45. 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

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

Lemma 4.3 $\Rightarrow Y$ hat Dichte

$$\begin{aligned} & \Phi_{0,E}(B^{-1}(y-m)) |\det B^{-1}| \\ &= \frac{1}{\sqrt{2\pi}^n} \frac{1}{\sqrt{|\det C|}} \exp\left(-\frac{1}{2}(y-m)^T C^{-1}(y-m)\right) \\ \mathbb{E}[Y_i] &= \mathbb{E}\left[\sum_{k=1}^n B_{i,k} X_k + m_i\right] = \sum B_{i,k} \underbrace{\mathbb{E}[X_k]}_{=0} + m_i = m_i \\ \text{Cov}[Y_k, Y_l] &= \sum_{i,j=1}^n B_{k,i} B_{l,j} \underbrace{\text{Cov}[X_i, X_j]}_{=\delta_{i,j}} = \sum_{i=1}^n B_{k,i} B_{l,i} = C_{k,l} \end{aligned}$$

□

Ein paar Eigenschaften:

1. $\mathcal{N}(0, E) \circ R^{-1} = \mathcal{N}(0, E)$ ($Y = RX$) 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 $a \in \mathbb{R}^k$
 $\Rightarrow Y = Ax + a \sim \mathcal{N}_k(Am + a, ACA^t)$

Besondere Verteilungen:

Γ -Verteilung

β -Verteilung

Chi-Quadrat ...

Satz 46.

$$X \sim N(0, 1) \Rightarrow X^2 \sim \Gamma_{\frac{1}{2}, \frac{1}{2}}$$

Beweis: Siehe Blatt 4 Aufgabe 1.

Satz 47. 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 48. $\Gamma_{\alpha,r} * \Gamma_{\alpha,s} = \Gamma_{\alpha,r+s}$ (*Faltungshalbgruppe*)

Aus Satz 46 + Korollar 48 folgt sofort

Satz 49. (*Chi-Quadrat Verteilung*)

Seien X_1, \dots, 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 50. (*Fisher-Verteilung*)

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

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

Beweis. $X = \sum_{k=1}^m X_k^2 \sim \Gamma_{1/2, m/2}$ Satz 49

$$Y = \sum_{k=1}^n Y_k^2 \sim \Gamma_{1/2, n/2} \quad \text{Satz 49}$$

$$\Rightarrow \frac{X}{X+Y} \sim \beta_{m/2, n/2} \quad \text{Satz 47}$$

$$\text{Au\ss} \text{erdem } F: m, n = \frac{n}{m} \frac{X}{Y} = \frac{n}{m} \frac{Z}{1-Z} = T(Z)$$

$$T(x) = \frac{n}{m} \frac{x}{1-x}: (0,1) \rightarrow (0,\infty)$$

$$T^{-1}(y) = \frac{my}{n+my}$$

Lemma 43

$$\Rightarrow f_{m,n} = \beta_{m/2, n/2} \left(\frac{my}{n+my} \right) \cdot \frac{mn}{(n+my)^2}$$

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

□

Definition 51. Die Verteilung $F_{m,n}$ auf $(0, \infty)$ mit Dichtefunktion $f_{m,n}$ hei\ss t Fisher-Verteilung mit m und n Freiheitsgraden.

Bemerkung 52. $\forall m, n \in \mathbb{N} \quad F_{m,n} = \beta_{m/2, n/2} \circ T^{-1}$

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

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

Satz 53. (*Student Verteilung*)

Seien X, Y_1, \dots, 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

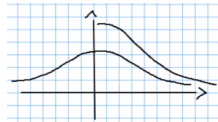
$$\tau_n(x) = \frac{1}{\left(1 + \frac{x^2}{n}\right)^{\frac{n+1}{2}}} \frac{1}{\sqrt{n} B(1/2, n/2)}$$

Beweis. Blatt 4, Aufgabe 1, oder

$T^2 \sim F_{1,n} \xrightarrow{\text{Lemma 43}} |T| = \sqrt{|T|}$ 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.**

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

□

Definition 54. Das W -Maß t_n mit Dichte τ_n heißt Student-t-Verteilung mit n Freiheitsgraden.

Satz 55. (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. $M \sim \mathcal{N}_{m,v/n}, \frac{n-1}{n} V^* \sim \chi_{n-1}^2$
- iii. $\sqrt{n}(M - m) / \sqrt{V^*} \sim t_{n-1}$

Beweis. (i)

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

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

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

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

$$\begin{aligned} \Rightarrow (n-1)V^* &= \sum_{k=1}^n (X_k - M)^2 = \sum_{k=1}^n X_k^2 - nM^2 \\ &= |Y|^2 - Y_1^2 = \sum_{k=1}^n Y_k^2 \end{aligned}$$

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

□

Ende Vorlesung 10

6 Testtheorie

Beispiel 56. (Test faire Münze)

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

Werfen n -mal die Münze $X = (X_1, \dots, X_n)$ 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 $|T(X) - \frac{1}{2}| > \frac{1}{\sqrt{n}}$ nehmen wir H_1 an, sonst H_0 .

Was ist ein Test? Wie sollte man Entscheidungsverfahren 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 θ in θ_0 und θ_1 s.d.

$\vartheta \in \theta_0 \Leftrightarrow \vartheta$ ist akzeptabel, gewünschter Normalfall

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

3. Wahl eines Irrtumsniveau

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

4. Wahl einer Entscheidungsregel

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

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

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

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

$\varphi(x) \Leftrightarrow$ Nehme die Nullhypothese mit W'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 57. Sei $(\mathcal{X}, \mathcal{F}, (\mathbb{P}_\vartheta: \vartheta \in \theta))$ ein stat. Modell und $\theta = \theta_0 \cup \theta_1$ eine Zerlegung von θ in Nullhypothese und Alternative.

- a) Jede Statistik $\varphi: \mathcal{X} \rightarrow [0, 1]$ heißt Test von der θ_0 gegen θ_1 .
- b) Ein Test φ heißt nicht randomisiert, falls $\varphi(x) \in \{0, 1\} \forall x \in \mathcal{X}$.
In diesem Fall heißt $\{x \in \mathcal{X}: \varphi(x) = 1\}$ Ablehnungsbereich oder kritischer Bereich.
- c) Falls $\varphi(x) \notin \{0, 1\}$ für ein $x \in \mathcal{X}$
 $\Rightarrow \varphi$ ist randomisiert.
- d) Effektives Niveau von φ 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 φ hat (Irrtums-)niveau α wenn

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

- f) Gütefunktion $G_\varphi: \theta \rightarrow [0, 1]$

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

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

Wie soll man φ wählen?

- a) $G_\varphi(\vartheta) \leq \alpha$ für alle $\vartheta \in \theta_0$. \Rightarrow Irrtums - W'keit für Fehler 1. Art $\leq \alpha$.
- b) $G_\varphi(\vartheta)$ so groß wie möglich $\forall \vartheta \in \theta_1 \Rightarrow$ Fehler 2. Art minimieren.

Definition 58. Ein Test φ von θ_0 gegen θ_1 heißt besten Test zum Niveau α , wenn $\text{Niveau}(\varphi) = \alpha$ und \forall Tests ψ mit $\text{Niveau}(\psi) = \alpha$ gilt $G_\varphi(\vartheta) \geq G_\psi(\vartheta) \forall \vartheta \in \theta_1$. Auf englisch: UMP-Test = uniform most powerful test.

Beispiel 59. (Faire Münze)

Wir werfen n mal eine Münze

- $\mathcal{X} = \{0, \dots, n\}, \mathbb{P}_p = \text{Bin}_{n,p}$ mit $p \in [0, 1]$.
- $\theta_0 = \{\frac{1}{2}\}$ und $\theta_1 = [0, 1] \setminus \{\frac{1}{2}\}$
- $\alpha \in (0, 1)$ fest z.B.: $\alpha = 0.05$ (5% Fehler 1. Art möglich)
- $\varphi(x) = \mathbb{1}_{\{|x - \frac{n}{2}| > c\}}$

Jetzt wollen wir c berechnen s.d. der Test Niveau α hat.

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

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

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

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

hat Niveau α

Beispiel 60. $\alpha = 0.05$

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

10	0	0.5987	0.3487	0.1969	0.1074	0.0563	0.0282	0.0135	0.0060	0.0025	0.0010
1	1	0.9139	0.7361	0.5443	0.3758	0.2440	0.1493	0.0860	0.0464	0.0233	0.0107
2	2	0.9885	0.9298	0.8202	0.6778	0.5256	0.3828	0.2616	0.1673	0.0996	0.0547
3	3	0.9990	0.9872	0.9500	0.8791	0.7759	0.6496	0.5138	0.3823	0.2660	0.1719
4	4	0.9999	0.9984	0.9901	0.9672	0.9219	0.8497	0.7515	0.6331	0.5044	0.3770
5	5	1.0000	0.9999	0.9986	0.9936	0.9803	0.9527	0.9051	0.8338	0.7384	0.6230
6	6		1.0000	0.9999	0.9991	0.9965	0.9894	0.9740	0.9452	0.8980	0.8281
7	7			1.0000	0.9999	0.9996	0.9984	0.9952	0.9877	0.9725	0.9453
8	8				1.0000	1.0000	0.9999	0.9995	0.9983	0.9955	0.9893
9	9						1.0000	1.0000	0.9999	0.9997	0.9990
10	10								1.0000	1.0000	1.0000

Abbildung 4.

Frage: Können wir nicht einfach einen Test anderer Form wählen

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

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

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

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

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

Um diese Absurdität zu vermeiden:

Definition 61. Ein Test φ heißt unverfälscht zum Niveau α wenn

$$G_{\varphi}(\vartheta_0) \leq \alpha \leq G_{\varphi}(\vartheta_1) \forall \vartheta_0 \in \theta_0, \vartheta_1 \in \theta_1,$$

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

Ende Vorlesung 11

6.1 Neyman-Pearson-Test

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

Annahme:

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

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

Seien ρ_0, ρ_1 die Dichte von $\mathbb{P}_0, \mathbb{P}_1$.

Gesucht: Ein bester Test φ von \mathbb{P}_0 gegen \mathbb{P}_1 .

Idee:

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

Deshalb definiert man den Likelihood-Quotienten

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

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

\Rightarrow Wir kriegen einen Test der Form

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



Abbildung 5.

Definition 62. Ein Test dieser Form heißt Neyman-Pearson-Test zum Schwellenwert c .

Satz 63. (Neyman-Pearson-Lemma 1932)

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

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

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

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

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

Bemerkung 64. Im Beweis werden wir sehen:

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

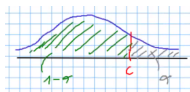


Abbildung 6.

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

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

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

1. Fall:

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

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

2. Fall:

$\mathbb{P}_0[R(x) = c] > 0$, 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 φ ein N-P-Test mit $\mathbb{E}_0[\varphi] = \alpha$ und Schwellenwert c und ψ ein Test zum Niveau α .

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

1. Fall:

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

2. Fall:

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

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

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

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

Ununterscheidbar: Sei ψ ein bester Test mit Niveau α .

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

$\Rightarrow f_1(x) = cf_0(x)$ 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. □

Beispiel 65. (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 $n = 10$ Würfeln entscheiden soll, was der Fall ist mit Irrtumsniveau $\alpha = 0.01$.

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

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

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

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 α -Fraktil von $\mathbb{P}_0 = \text{Bin}_{10, \frac{1}{2}}$

$$\mathbb{P}_0[X \geq a] \geq \alpha \quad \text{und} \quad \mathbb{P}_0[X > a] \leq \alpha$$

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

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

$\Rightarrow a = 9$

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

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

Seien $\mathbb{P}_0, \mathbb{P}_1$ zwei W-maße mit Dichten ρ_0, ρ_1 . Dann

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

heißt **relative Entropie** von \mathbb{P}_0 bzgl. \mathbb{P}_1 . Insbesondere $H(\mathbb{P}_0|\mathbb{P}_1) \geq 0$ 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 66. (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 φ_n ein $N - P$ - Test mit $\mathbb{E}_0[\varphi] = \alpha$, der nur von den ersten n Beobachtungen X_1, \dots, X_n abhängt. Dann

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

d.h.

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

7 Bonus: Interessantes Wissen aus den Blättern

7.1 Blatt 01

Bemerkung 67. (Momentenmethode):

- Gegeben seien X_1, \dots, 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.

7.2 Blatt 03

Es sei $(\mathcal{X}, \mathcal{F}, \mathbb{P}_\vartheta: \vartheta \in \theta)$ ein stat. Modell.

Definition 68. Wir nennen eine Statistik $T: \mathcal{X} \rightarrow \Sigma$ mit abzählbarem Wertebereich Σ :

- 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 $g: \Sigma \rightarrow \mathbb{R}$ existiert, s.d. $\mathbb{E}_\vartheta[g(T)] = 0$ für alle $\vartheta \in \theta$.

Für eine reelle Schätzfunktion τ :

Satz 69. (Satz von Rao-Blackwell)

Es sei T suffizient und S ein Erwartungstreuer Schätzer für τ . Wir definieren $g_S: \Sigma \rightarrow \mathbb{R}$ durch $g_S(s) := \mathbb{E}_{Q_s}[S]$. Der Schätzer $g_S(T)$ ist erwartungstreu für τ und für alle $\vartheta \in \theta$ gilt:

$$\text{Var}_\vartheta[g_S(T)] \leq \text{Var}_\vartheta[S].$$

Satz 70. (Satz von Lehmann-Scheffé)

Es sei $g: \Sigma \rightarrow \mathbb{R}$ gegeben und T sei suffizient und vollständig. $g(T)$ ist ein bester Schätzer für τ , falls $g(T)$ erwartungstreu für τ ist.

Beispiel 71. Dichten in Exponentiellen Familien:

- $f_\vartheta(x) = \frac{\vartheta^x}{x!} \exp(-\vartheta)$, $x \in \mathbb{N}_0$ und $\vartheta > 0$.

- Inverse Gamma-Verteilung 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 Gamma-Verteilung 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 $x, \vartheta \in \mathbb{R}$.

7.3 Blatt 05

7.4 Blatt 06: Zusammenhang zwischen Konfidenzbereichen und Tests

Für ein stat. Modell $(\mathcal{X}, \mathcal{F}, \mathbb{P}_\vartheta: \vartheta \in \theta)$ gilt

- Ist $C: \mathcal{X} \rightarrow \mathcal{P}(\theta)$ ein Konfidenzbereich für ϑ 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 $\theta_0 = \{\vartheta_0\}$ gegen die Alternative $\theta_1 = \{\vartheta: \vartheta \neq \vartheta_0\}$ zum Niveau α .
- Es sei für jedes $\vartheta \in \theta$ K_ϑ der kritische Bereich eines Tests von der Hypothese H_0 mit $\theta_0 = \{\vartheta\}$ gegen die Alternative H_0 mit $\{\tilde{\vartheta}: \tilde{\vartheta} \neq \vartheta_0\}$ zum Niveau α . Dann ist $C = \{(x, \vartheta) \in \mathcal{X} \times \vartheta: x \in K_\vartheta^c\}$ ein Konfidenzbereich zum Niveau $1 - \alpha$ für ϑ .