Wahrscheinlichkeitstheorie und Mathematische Statistik

Müller-Harknett, Ursula Ute, Dr. rer. nat.

Inhaltsverzeichnis

Kapitel 1

Einige Grundlagen

1.1 Diskrete Verteilungen

Gegeben $(\Omega, A, P), \Omega$ höchstens abzählbar
P bestimmt durch $p_w := P(\{w\})$, $w \in \Omega$, $p_w \geq 0$, $\sum\limits_{w \in \Omega} p_w = 1$

1.1.1 Bernoulli-Verteilung

$\mathrm{B}(n,p)$, $\mathrm{Bin}(n,p)$
Bsp. Münzwurf mit $p \in (0,1)$, '1' Kopf, '0' Zahl
$(\Omega, \mathcal{A}, P) = (\{0,1\}, \mathfrak{P}(\{0,1\}), P)$ mit $P(\{1\}) = p, P(\{0\}) = 1-p$, also $P(\{w\}) = p^w(1-p)^{1-w}$, $w = 0,1$

n-facher Münzwurf (unab. Wiederholungen)
$(\Omega, A, P) = (\{0,1\}^n, \mathfrak{P}(\Omega)P)$, $P(\{w\}) = \prod\limits_{j=1}^{n} p^{w_j}(1-p)^{1-w_j}$
Betrachte Zufallsvariable $X = \#\{\text{Münze zeigt Kopf}\} = \sum\limits_{j=1}^{n} w_j$, d.h. $(\Omega, A, P) \xrightarrow{X} (\Omega', A') = (\{0,1,...,n\}, \mathfrak{P}(\Omega'))$
Gesucht: Bildmaß von X
Betrachte erzeugende Funktion ('probability generating function') für diskrete Verteilungen (konv. $\forall z \in \mathbb{C}, |z| \leq 1$)
Allgemein

$$Ez^X = \sum_{k=0}^{\infty} z^k P(X=k) = \varphi_X(z), P(X=k) = \frac{\varphi^{(k)}(0)}{k!}$$

Hier

$$
\begin{aligned}
\varphi_X(z) &= E\left(z^{\sum_{j=1}^n w_j}\right) = E\prod_{j=1}^n z^{w_j} \overset{\text{unab.}}{=} \prod_{j=1}^n E z^{w_j} \\
&= \left(z^0(1-p)+z^1 p\right)^n = (zp+1-p)^n \underset{\text{Formel}}{\overset{\text{binom.}}{=}} \sum_{k=0}^n \binom{n}{k}(zp)^k(1-p)^{n-k} \\
&= \sum_{k=0}^n z^k \underline{\binom{n}{k}p^k(1-p)^{n-k}} = Ez^X = \sum_{k=0}^n z^k \underline{P(X=k)} \\
&\Rightarrow P(X=k) = \binom{n}{k}p^k(1-p)^{n-k} \text{ für } k=0,1,...,n \ \underline{\text{Binomialverteilung}}
\end{aligned}
$$

1.1.2 Negative Binomialverteilung

('Pascal Verteilung') $\mathrm{NB}(r,p)$
Betrachte Anzahl Münzwürfe bis genau r-mal Kopf beobachtet wurde
Sei $(Z_m)_m$ Folge von i.i.d. (u.i.v.), binärer (0/1) Zufallsvariable mit

$$
P(Z_m=1) = p \in (0,1), \quad Z := \min\left\{n \in \mathbb{N} : \sum_{j=1}^n Z_j = r\right\}
$$

Verteilung von Z?
Mit

$$
\{Z=n\} = \{Z_n = 1\} \cap \left\{\sum_{j=1}^{n-1} Z_j = r-1\right\}
$$

folgt

$$
\begin{aligned}
P(Z=n) &= \underbrace{P(Z_n=1)}_{p} \cdot \underbrace{P\left(\sum_{j=1}^{n-1} Z_j = r-1\right)}_{\sim \mathrm{Bin}(n-1,p)} \\
&= p\cdot\binom{n-1}{r-1}p^{r-1}(1-p)^{n-1-(r-1)} \\
&= \binom{n-1}{r-1}p^r(1-p)^{n-r} \quad \mathrm{NB}(r,p)\text{-Dichte}
\end{aligned}
$$

1.1.3 Hypergeometrische Verteilung

$\mathrm{Hyp}(N,M,n)$
Betrachte z.B. ja/nein Umfrage, #'ja', #'defekte Teile'
Urne mit N Kugeln, davon M rote, $N-M$ weiße Kugeln
Ziehe $n < N$ Kugeln (Stichprobe ohne Zurücklegen)
$X :=$ Anzahl rote Kugeln in Stichprobe

$$
\Rightarrow P(X=l) = \frac{\binom{M}{l}\cdot\binom{N-M}{n-l}}{\binom{N}{n}} \quad \text{hypergeom. Dichte}
$$

1.1.4 Multinomialverteilung

n Kugeln verteilen sich zufällig (unabhängig) auf m Fächer $F_1, ..., F_m$

$$P('F_j') = p_j \in (0,1), \quad \sum_{j=1}^{m} p_j = 1, \quad \Omega = \{1, ..., m\}^n$$
$$P(\{w\}) = P(\{w_1, ..., w_n\}) = P(\underbrace{X_1(w)}_{w_1}, ..., \underbrace{X_n(w)}_{w_n})$$

$X_1, X_2, ..., X_n$ iid, $P(X_k = j) = p_j$
$Y_j := \sum\limits_{k=1}^{n} 1_{\{j\}}(X_k)$ $\#\{\text{Kugeln in } F_j\}$, $j = 1, ...m$. Dann gilt

$$P(Y_1 = l_1, ..., Y_m = l_m) = \begin{cases} \frac{n!}{l_1! \cdot ... \cdot l_m!} \prod\limits_{j=1}^{m} p_j^{l_j} & \text{falls } \sum_{j=1}^{m} l_j = n \\ 0 & \text{sonst} \end{cases}$$

Spezialfall $m = 2$ Binomialverteilung

1.1.5 Poisson-Verteilung

Pois(λ), P(λ)
Betrachte Bin(n, p), n groß, p klein
z.B. n Unfälle/Jahr in Deutschland
p_n Unfallwahrscheinlichkeit (seltenes Ereignis)
Angenommen $p_n \approx \frac{\lambda}{n}$, λ 'Intensität', also $\lambda \approx np_n$ $(E(Bin(n,p)))$ für $n \to \infty, p_n \to 0$
Dies liefert die Poisson-Verteilung mit Parameter $\lambda > 0$.

$$P(\{k\}) = e^{-\lambda} \frac{\lambda^k}{k!}, \quad k \in \mathbb{N}_0$$

'Verteilung der seltenen Ereignisse'

Theorem: 1.1

Betrachte $(X_n)_n$ Folge von Zufallsvariablen, $X_n \sim$ Bin(n, p_n), $np_n \overset{n\to\infty}{\to} \lambda > 0$.
Dann gilt $X_n \overset{D}{\to} X \sim$ Pois(λ) $\quad (n \to \infty)$

1.2 Absolutstetige Verteilungen

W-Maße auf $(\mathbb{R}^d, \mathcal{B}(\mathbb{R})^d)$ mit Dichte bzgl. Lebesgue-Maß.

1.2.1 Normalverteilung

$N(\mu, \sigma^2)$, Erwartungswert $\mu \in \mathbb{R}$, Varianz $\sigma^2 > 0$

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right) \quad x \in \mathbb{R}$$

Bivariante Normalverteilung $(X_1, X_2) \sim N_2\left(\begin{pmatrix}\mu_1\\ \mu_2\end{pmatrix}, \begin{pmatrix}\sigma_1^2 & \sigma_{12}\\ \sigma_{12} & \sigma_2^2\end{pmatrix}\right)$, $\rho = Corr(X_1, X_2)$

$$f(x_1, x_2) = \frac{1}{2\pi\sqrt{\sigma_1^2\sigma_2^2 - \rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}\left(\left(\frac{X_1-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right) + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right)\right)$$

mit $\rho = Corr(X_1, X_2) = \frac{Cov(X_1,X_2)}{\sigma_1\sigma_2}$.
Beachte: $\rho = 0 \Rightarrow X_1, X_2$ unabhängig (Produkt eindimensionaler Dichten)

Lemma: k-variante Normalverteilung

Sei $Z \sim N_n(0, I_n)$ n-dim. normalverteilt, $A \in \mathbb{R}^{k\times n}$, $\mu \in \mathbb{R}^k$.
Dann ist

$$X := AZ + \mu$$

k-dim. normalverteilt, $X \sim N_k(\mu, \Sigma)$ mit

$$EX = \mu$$

und Kovarianzmatrix

$$CovX = E((X-\mu)(X-\mu)^T) = E((AZ)(AZ)^T) = AE(ZZ^T)A^T = AA^T =: \Sigma$$

Ist Σ invertierbar, dann hat X die Dichte

$$f(x) = \frac{1}{(2\pi)^{\frac{n}{2}}\sqrt{\det\Sigma}} \exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right)$$

1.2.2 Chi-Quadrat-Verteilung

χ_d^2

Betrachte $X_1, X_2, ..., X_d$ unabhängige Zufallsvariablen, $X_i \sim N(0,1)$, $Y := \sum_{j=1}^{d} X_j^2 \geq 0$

Verteilung von Y?
$\underline{d=1}$: $Y = X^2$ $(= X_j^2)$

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P(X^2 \leq y) = P(-\sqrt{y} \leq X \leq \sqrt{y}) \\ &= \frac{1}{\sqrt{2\pi}}\int_{-\sqrt{y}}^{\sqrt{y}} e^{-\frac{x^2}{2}}dx = 2\frac{1}{\sqrt{2\pi}}\int_0^{\sqrt{y}} e^{-\frac{x^2}{2}}dx \end{aligned}$$

Substitution mit $t = x^2, x = \sqrt{t}, dx/dt = \frac{1}{2}\cdot t^{-\frac{1}{2}}$ führt auf

$$\begin{aligned} &= \cancel{2}\frac{1}{\sqrt{2\pi}}\int_0^y e^{-\frac{t}{2}}\cdot\frac{\cancel{1}}{\cancel{2}}\cdot t^{-\frac{1}{2}}dt = \frac{1}{\sqrt{2\pi}}\int_0^y \frac{e^{-\frac{t}{2}}}{\sqrt{t}}dt \\ &\Rightarrow f_Y(t) = \frac{1}{\sqrt{2\pi}}\frac{e^{-\frac{t}{2}}}{\sqrt{t}}1_{(0,\infty)}(t) \quad \chi^2\text{-Dichte mit 1 Freiheitsgrad} \end{aligned}$$

Allgemein $d = \nu \in \mathbb{N}$: χ^2-Verteilung mit ν Freiheitsgeraden

$$f(x) = 2^{-\frac{\nu}{2}} \frac{1}{\Gamma(\frac{\nu}{2})} x^{\frac{\nu}{2}-1} e^{-\frac{x}{2}} 1_{(0,\infty)}$$

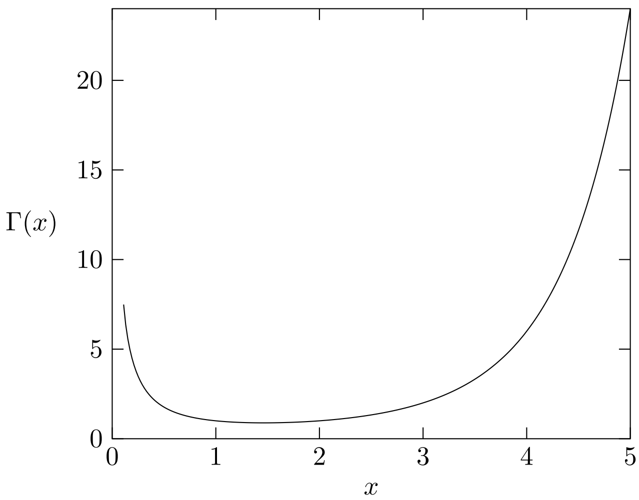
$\Gamma(.)$ ist die Gamma-Funktion

$$\Gamma(z) = \int_0^\infty x^{z-1} e^{-x} dx \text{ für } z > 0$$

Es gilt:

- $\Gamma(z+1) = z\Gamma(z)$
- $\Gamma(n+1) = n\Gamma(n) = n! \ \forall n \in \mathbb{N}_0$
- $\Gamma(\frac{1}{2}) = \sqrt{\pi}$

z.B. $\Gamma(4) = 3 \cdot 2 \cdot \Gamma(1) = 6 \int_0^\infty e^{-x} dx = 6 \cdot 1$ oder $\Gamma(1.5) = \frac{1}{2} \cdot \Gamma(\frac{1}{2}) = \frac{\sqrt{\pi}}{2}$



1.2.3 Gamma-Verteilung

$\Gamma(\lambda, r), \lambda, r \in \mathbb{R}_+$

$$f(x) = \frac{\lambda^r}{\Gamma(r)} x^{r-1} e^{-\lambda x} 1_{[0,\infty)}(x)$$

Spezialfälle:

- $r = \frac{\nu}{2}$, $\nu \in \mathbb{N}$, $\lambda = \frac{1}{2}$:

$$f(x) = \left(\frac{1}{2}\right)^{\frac{\nu}{2}} \frac{1}{\Gamma(\frac{\nu}{2})} x^{\frac{\nu}{2}-1} e^{-\frac{1}{2}x} 1_{[0,\infty)}(x) \quad \underline{\chi^2 - \text{Dichte}}$$

- $r = 1$: $f(x) = \lambda e^{-\lambda x}$, $x \geq 0$ Exponential-Dichte.

Herleitung der Charakteristische Funktion der $\Gamma(\lambda, r)$-Verteilung für $r \in \mathbb{N}$:
Betrachte $W \sim \Gamma(\lambda, r)$

$$\begin{aligned}
\varphi_W(t) = E(e^{itW}) &= \frac{\lambda^r}{\Gamma(r)} \int_0^\infty e^{itx} x^{r-1} e^{-\lambda x} dx \\
&= \frac{\lambda^r}{\Gamma(r)} \int_0^\infty x^{r-1} e^{x(it-\lambda)} dx \\
&= \frac{\lambda^r}{\Gamma(r)} i^{-(r-1)} \frac{d^{r-1}}{dt^{r-1}} \int_0^\infty \underbrace{e^{(it-\lambda)x}}_{e^{-(\lambda-it)x}} dx \cdot \frac{\lambda - it}{\lambda - it} \\
&= \frac{\lambda^r}{\Gamma(r)} i^{-(r-1)} \frac{d^{r-1}}{dt^{r-1}} \cdot 1 \cdot \frac{1}{\lambda - it} \\
&\text{mit } \frac{d}{dt}(\lambda - it)^{-1} = (-1)(\lambda - it)^{-2}(-i) = i(\lambda - it)^{-2} \\
&\text{und } \frac{d^{r-1}}{dt^{r-1}}(\lambda - it)^{-1} = i^{r-1}(\lambda - it)^{-r} \underbrace{(r-1)!}_{\Gamma(r)} \\
&= \frac{\lambda^r}{\Gamma(r)} i^{-(r-1)} i^{r-1} (\lambda - it)^{-r} \Gamma(r) \\
&= \underline{\lambda^r (\lambda - it)^{-r}}
\end{aligned}$$

Angenommen $W \sim \Gamma(\lambda, r)$, $V \sim \Gamma(\lambda, s)$ sind unab. dann gilt

$$\begin{aligned}
\varphi_{W+V}(t) = \varphi_W(t)\varphi_V(t) &= \lambda^r (\lambda - it)^{-r} \lambda^s (\lambda - it)^{-s} \\
&= \lambda^{r+s} (\lambda - it)^{-(r+s)} \quad \Gamma(\lambda, r+s) - \text{Verteilung}
\end{aligned}$$

$\Rightarrow$ Für unabhängig Zufallsvariablen $Z_1, ..., Z_d \sim N(0,1)$ gilt $\sum_{j=1}^d Z_j^2 \sim \chi^2(d)$ $(Z_i^2 \sim \Gamma(\frac{1}{2}, \frac{1}{2}))$

1.3 Transformationssatz und Deltamethode

Transformationssatz für Lebesgue-Dichten
Sei P W-Maß auf $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$
Vorbemerkung: $d = 1$
Betrachte X, $Y = h(X)$ bijektiv mit h^{-1} differenzierbar, Y abs. stetig
1. Fall: h streng monoton wachsend

$$\begin{aligned}
f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{d}{dy} P(h(X) \leq y) &\overset{\text{mon.}}{\underset{\text{wachs.}}{=}} \frac{d}{dy} P(X \leq h^{-1}(y)) \\
&= \frac{d}{dy} F_X(h^{-1}(y)) = f_X(h^{-1}(y)) \frac{d}{dy} h^{-1}(y) \\
&= \frac{1}{\underbrace{h'}_{>0}(h^{-1}(y))} f_X(h^{-1}(y))
\end{aligned}$$

2. Fall: h streng monoton fallend

$$
\begin{aligned}
f_Y(y) = \frac{d}{dy}P(h(x) \leq y) \overset{\text{mon.}}{\underset{\text{fall.}}{=}} \frac{d}{dy}P(X \geq h^{-1}(y)) &= \frac{d}{dy}(1 - F_X(h^{-1}(y))) \\
&= -\frac{1}{\underbrace{h'}_{<0}(h^{-1}(y))} f_X(h^{-1}(y))
\end{aligned}
$$

Insgesamt:

$$
f_Y(y) = \left| \frac{1}{h'(h^{-1}(y))} \right| f_X(h^{-1}(y))
$$

Merkregel:

$$
f_Y(y) = \left| \frac{dx}{dy} \right| f_X(y) \quad (x = h^{-1}(y))
$$

Beispiel
$X \sim \text{Expo}(\frac{1}{10})$, $Y = \ln(X) = h(X)$, $X = e^Y = h^{-1}(Y)$

$$
\begin{aligned}
f_Y(y) &= \left| \frac{dx}{dy} \right| \frac{1}{10} e^{-x \cdot \frac{1}{10}} 1_{[0,\infty)}(x) \\
&= \left| \frac{de^y}{dy} \right| \frac{1}{10} e^{-\frac{e^y}{10}} \underbrace{1_{[0,\infty)}(e^y)}_{=1} \\
&= \frac{e^y}{10} e^{-\frac{e^y}{10}} \quad \forall y \in \mathbb{R}
\end{aligned}
$$

Lemma: 1.1 Transformationssatz für mehr-dimensionale Zufallsvariablen

Betrachte $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, f Lebesgue-Dichte.
Angenommen $P(I) = 1$ für $I \subseteq \mathbb{R}^d$ offen, $T : I \to \mathbb{R}^d$ stetig differenzierbar injektiv, mit T^{-1} stetig differenzierbar auf $T(I)$.
Das Bildmaß von P unter T besitzt dann die Lebesgue-Dichte $f_T(y)$

$$
f_T(y) = \begin{cases} f(T^{-1}(y)) \cdot |\det(\partial T^{-1}(y))| & \forall y \in T(I) \\ 0 & \forall\, y \in \mathbb{R}^d \setminus T(I) \end{cases}
$$

mit Jacobi Matrix

$$
\partial T^{-1}(y) = \begin{pmatrix} \frac{\partial (T^{-1})_1(y)}{\partial y_1} & \dots & \frac{\partial (T^{-1})_1(y)}{\partial y_d} \\ \vdots & & \vdots \\ \frac{\partial (T^{-1})_d(y)}{\partial y_1} & \dots & \frac{\partial (T^{-1})_d(y)}{\partial y_d} \end{pmatrix}
$$

Einfache Version/ Merkregel für $d = 2$
(X, Y) mit Dichte $f_{X,Y}$, $U = g(X, Y)$, $V = h(X, Y)$, $(x, y) \mapsto (g(x, y), h(x, y))$

$$
f_{U,V}(u, v)\partial u \partial v = f_{X,Y}(x, y)\partial x \partial y \quad \frac{\partial x \partial y}{\partial u \partial v} = |\det(\text{Jacobi})|
$$

bzw.

$$f_{U,V}(u,v) = f_{X,Y}(x,y) \cdot \left| \det \begin{pmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{pmatrix} \right|$$

Beispiel

$$f_{X,Y}(x,y) = \frac{1}{2}(\lambda^2 e^{-\lambda(x+y)} + \mu^2 e^{-\mu(x+y)}) 1_{(0,\infty)}(x) \cdot 1_{(0,\infty)}(y)$$

$$T = X+Y, \quad W = \frac{X}{X+Y} \Rightarrow W = \frac{X}{T} \Leftrightarrow \underline{X = TW}$$
$$\underline{Y} = T - X = T - TW = \underline{T(1-W)}$$

$$\frac{\partial x \partial y}{\partial t \partial w} = \left| \det \begin{pmatrix} \frac{\partial x}{\partial t} & \frac{\partial x}{\partial w} \\ \frac{\partial y}{\partial t} & \frac{\partial y}{\partial w} \end{pmatrix} \right| = \left| \det \begin{pmatrix} w & t \\ 1-w & -t \end{pmatrix} \right|$$
$$= |-wt - t(1-w)| = |-w - t + w| = |-t| = t$$

$$f_{T,W}(t,w) = t \cdot f_{X,Y}(tw, t(1-w))$$
$$= t \cdot \frac{1}{2}(\lambda^2 e^{-\lambda(\cancel{tw}+t-\cancel{tw})} + \mu^2 e^{-\mu t}) 1_{(0,\infty)}(tw) 1_{(0,\infty)}(t(1-w))$$

mit $1_{(0,\infty)}(tw) = 1$ für $t > 0, w > 0$ und $1_{(0,\infty)}(t(1-w)) = 1$ für $t > 0$, $1 - w > 0 \Leftrightarrow w < 1$

$$\Rightarrow f_{T,W}(t,w) = \underbrace{t \cdot \frac{1}{2}(\lambda^2 e^{-\lambda t} + \mu^2 e^{-\mu t}) 1_{(0,\infty)}(t)}_{f_T(t)} \underbrace{1_{(0,1)}(w)}_{f_W(w) \sim U(0,1)}$$

Man sieht das T, W unabhängig sind

Theorem: 1.2 Delta-Methode

Sei $(Z_n)_n$ Folge von Zufallsvariablen mit $\sqrt{n}(Z_n - \theta) \xrightarrow{\mathrm{D}} Z$, Z Zufallsvariable mit Varianz $\sigma^2 \in (0,\infty)$, $g : \mathbb{R} \to \mathbb{R}$ eine in θ differenzierbare Funktion mit $g'(\theta) \neq 0$.
Dann gilt

$$\sqrt{n}(g(Z_n) - g(\theta)) \xrightarrow{\mathrm{D}} g'(\theta) Z$$

Ist $Z \sim N(0, \sigma^2)$, dann gilt $g'(\theta) \cdot Z \sim N(0, \sigma^2 g'(\theta)^2)$

Einschub/Notation:
Seien $\{a_n\}, \{b_n\}$ Folgen mit $a_n \in \mathbb{R}$, $b_n \in \mathbb{R}_+$
(a) $a_n = O(b_n)$ falls $\lim\limits_{n\to\infty} \frac{a_n}{b_n} < \infty$
(b) $a_n = o(b_n)$ falls $\lim\limits_{n\to\infty} \frac{a_n}{b_n} = 0$
Seien $\{X_n\}, \{Y_n\}$ Folgen von Zufallsvariablen
(a) $X_n = o_P(Y_n)$ falls $\frac{X_n}{Y_n} \xrightarrow{\mathrm{P}} 0$, d.h. $\frac{X_n}{Y_n} = o_P(1)$
(b) $X_n = O_P(Y_n)$ falls $X_n = o_P(Y_n)$ oder falls gilt:

$$\forall \varepsilon > 0 \exists n_0 \in \mathbb{N} : \exists L, R \in \mathbb{R} : P\left(L \leq \frac{X_n}{Y_n} \leq R\right) > 1 - \varepsilon \quad \forall n > n_0$$

(die Folge $\frac{X_n}{Y_n}$ ist 'straff')

Lemma: 1.3

Betrachte $R : \mathbb{R} \to \mathbb{R}$ mit $R(0) = 0$, Zufallsvariablen $X_n \xrightarrow{\text{P}} 0$
Dann gilt für $p > 0$

$$R(h) = o(|h|^p) \quad (h \to 0) \Rightarrow R(X_n) = o_P(|X_n|^p)$$

Definition: 1.4 Varianzstabilisierende Transformationen (VST)

Sei $(Z_n)_{n\in\mathbb{N}}$ eine Folge von Zufallsvariablen mit $\sqrt{n}(Z_n - \theta) \xrightarrow{\text{D}} N(0, \underbrace{\sigma(\theta)^2}_{>0})\ \forall\theta$.
Eine Funktion heißt Varianzstablisierende Transformation (VST) für $(Z_n)_n$, falls gilt

$$\sqrt{n}(g(Z_n) - g(\theta)) \xrightarrow{\text{D}} N(0,1)\ \forall\theta \quad (\text{oder } N(0,c) \text{ für } c \in \mathbb{R}_+ \text{ konstant})$$

Konstruktion: Sei g differenzierbar in θ, $g'(\theta) \neq 0\ \forall\theta$
Delta Methode: $\sqrt{n}(g(Z_n) - g(\theta)) \xrightarrow{\text{D}} N(0, g'(\theta)^2\sigma(\theta)^2)$
Wähle also g so dass $g'(\theta)\sigma(\theta) = 1 \Leftrightarrow g'(\theta) = \frac{1}{\sigma(\theta)}$, d.h.

$$g(\theta) = \int \frac{1}{\sigma(\theta)} d\theta$$

Beispiel (Poisson)
Sei $X_n \sim \text{Pois}(\theta)$, $EX_n = Var(X_n) = \theta = \sigma^2(\theta)$, dann gilt für

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad \text{Stichprobenmittel}$$

$$\sqrt{n}(\bar{X}_n - \theta) \xrightarrow{\text{D}} N(0,\theta)$$

Wähle

$$g(\theta) = \int \frac{1}{\sigma(\theta)} d\theta = \int \theta^{-\frac{1}{2}} d\theta = 2\theta^{\frac{1}{2}} + c$$

O.B.d.A. sei $c = 0$ (im Nachfolgenden kürzt sich c raus)
Dann gilt

$$\sqrt{n}\left(2\sqrt{\bar{X}} + \cancel{c} - 2\sqrt{\theta} - \cancel{c}\right) \xrightarrow{\text{D}} N(0,1)$$

$$\Leftrightarrow \sqrt{n}\left(\sqrt{\bar{x}} - \sqrt{\theta}\right) \xrightarrow{\text{D}} N\left(0, \frac{1}{4}\right)$$

Theorem: 1.5 Multivariate Delta-Methode

Gegeben: (X_n) Folge k-dim. Zufallsvariable, $\mu \in \mathbb{R}^k$, $\Sigma \in \mathbb{R}^{k \times k}$ positiv semi-definit, $c_n \in \mathbb{R}$, $n \in \mathbb{N}$,
$\lim\limits_{n \to \infty} c_n = 0$, $\frac{X_n - \mu}{c_n} \xrightarrow{\text{D}} N_k(0, \Sigma)$
Ist $g : \mathbb{R}^k \to \mathbb{R}^m$ $(m \leq k)$ total differenzierbar in μ mit Jacobi-Matrix $g'(\mu) = D^T \in \mathbb{R}^{m \times k}$ mit $\det(D^T D) \neq 0$, dann gilt

$$c_n^{-1}(g(X_n) - g(\mu)) \xrightarrow{\text{D}} N_m\left(0, D^T \Sigma D\right)$$

Kapitel 2

Einführung in die Statistik und Stichprobenmomente

Betrachte $X : \Omega \to \mathbb{R}^k$ Zufallsvariable/-vektor
W-Theorie: Eigenschaften P^X
Statistik: Rückschlüsse auf P^X (unbekannt) aufgrund von Beobchtung(en) von X

Beispiel (2.1)
(a)
Wahlprognose: Wird Kandidat A gewählt?
Befragung von n Personen (Stichprobe)
X :=Anzahl der Personen in Stichprobe die A wählen
Verteilung: $X \sim \text{Bin}(n, \vartheta)$, ϑ $(= p)$ unbekannt.
Gesucht: Schätzung $\hat{\vartheta}$ für ϑ (Schätzer)
(b)
Behauptung: Glühlampenhersteller: Lebensdauer ≥ 6 Jahre
Stimmt das? ($EX_j > 6$?)
Modell: $X_1, X_2, ..., X_n$ iid $\sim \text{Exp}(\frac{1}{\vartheta})$ (mit $EX_j = \vartheta$)
(c)
Gesucht: Anzahl Fische $\vartheta = N$ in Teich = ?
Methode: Capture/ recapture sampling
d.h. fangen K Fische, markieren sie, zurück in Teich
1 Woche später: fange n Fische, $X = \#\{\text{markierte Fische}\}$
$\frac{X}{n}$ = Anteil markierter Fische in Stichprobe
Dann ist $\frac{X}{n} \approx \frac{K}{\vartheta}$ und $\vartheta \approx \frac{K \cdot n}{X}$
oder Herleitung eines Schätzers $\hat{\vartheta}$ mittels [j Anzahl mark. Fische]

$$P(X = j) = \frac{\binom{k}{j}\binom{\vartheta - K}{n-j}}{\binom{\vartheta}{n}} \quad (K, n, j \text{ bekannt})$$

(d)
Volumen X von $0.75l$ Weinflaschen
Annahme: $X_1, X_2, ..., X_n$ iid $\sim N(\mu, \sigma^2)$
Gesucht: Intervall I mit $P(\mu \in I) = 0.95$

Modellannahmen (2.2)
(Ω, A, P) W-Raum mit unbekanntem $P \in \mathcal{P} = \{P_\vartheta : \vartheta \in \Theta \neq \emptyset\}$, Θ 'Parameterraum'
$(\mathcal{X}, B)$ messbarer Raum (Messraum), $\mathcal{X}$ Stichprobenraum
$X : \Omega \to \mathcal{X}$, $A - B$-messbar
Dann ist $\tilde{P} = \left\{P_\vartheta^X : \vartheta \in \Theta\right\}$ Familie von W-Maßen auf $\mathcal{X}$
$(\mathcal{X}, B, \tilde{P})$ 'stochastisches Modell'
X statistische Experiment, Realisation von X 'Stichprobe'
Das Modell ist parametrisch falls $\Theta \subseteq$ endlich-dimensionaler Vektorraum
Andernfalls ist das Modell nicht-parametrisch

Bemerkung (2.3)
(a) Oft wird nur die Zufallsvariable X mit Verteilung $\mathcal{P} = \{P_\vartheta : \vartheta \in \Theta\}$ angegeben, d.h. P_ϑ^X wird mit P_ϑ identifiziert
(b) Gilt $X \sim P_{\vartheta_0}$ für ein $\vartheta_0 \in \Theta$, dann heißt ϑ_0 '<u>wahrer</u> Parameter' (Annahme $\vartheta_1 \neq \vartheta_2 \Rightarrow P_{\vartheta_1} \neq P_{\vartheta_2}$)
(c) Oft gilt $X_1, X_2, ..., X_n$ iid, d.h. $X_1, X_2, ..., X_n \sim P_\vartheta^{(1)} (= P_\vartheta^{X_1})$

$$P_\vartheta = \bigotimes_{i=1}^n P_\vartheta^{(1)}, \quad \text{d.h. Angabe von } P_\vartheta^{(1)} \text{ genügt}$$

<u>Annahmen</u>: Sei $X : \Omega \to \mathcal{X}$ Zufallsvariable mit $X \sim P^X = P_\vartheta$ für ein $\vartheta \in \Theta$
X ist meist diskret <u>oder</u> stetig, $X(w) \in \mathbb{R}^k$ für ein $k \in \mathbb{N}$
Alternative Notation für $\mathcal{P}$ wenn X stetig:

$$\mathcal{P} = \{f_\vartheta : \vartheta \in \Theta\}, \ X \sim f_\vartheta$$

Beispiel (2.4 Medikamentenvergleich)
Medikamente M_1, M_2
M_1 : Heilung in 80% aller Fälle
M_2 : neues Medikament, Test an 50 Patienten
Realisationen $x_1, x_2, ..., x_{50} \in \{0, 1\}$ von Zufallsvariablen $X_1, X_2, ..., X_n$ iid $\text{Bin}(1, \vartheta)$ $(x_1 = X_1(w), ...)$
$\vartheta (= p)$ Heilungs- oder Erfolgswahrscheinlichkeit
Möglicher Schätzer $\hat{\vartheta} = \frac{1}{50} \sum_{i=1}^{50} x_i$, z.B. $\hat{\vartheta} = \frac{42}{50} = 0.84$
Wie zuverlässig ist $\hat{\vartheta} = 0.84$? Spricht dies deutlich für M_2?
Mögliche Fehler?

Ziel (2.5)
Aussagen über ϑ_0 (wahrer Parameter) aufgrund von Beobachtung(en) $x = X(w)$, $X \sim P^X = P_{\vartheta_0}$
(i) Punktschätzungen: $\hat{\vartheta} = \hat{\vartheta}(x)$ Schätzer von ϑ_0
Herleitung und Eigenschaften guter Schätzverfahren
(ii) Hypothesentests: Gilt $\vartheta \in \Theta_0$ oder $\vartheta \in \Theta_1 = \Theta \backslash \Theta_0$
(iii) Konfidenzbereiche: Zufällige Mengen $C(X) \in \Theta$ die ϑ_0 mit einer gegebenen Wahrscheinlichkeit enthalten (z.B. $0.9, 0.95, 0.99$)

2.1 Stichprobenmomente

<u>Im restlichen Kapitel</u> sei X Zufallsvariable mit Verteilungsfunktion F und $X_1, X_2, ..., X_n \overset{iid}{\sim} F$ (Stichprobe)
Par. Modell: $F = \mathcal{F} = \{F_\vartheta : \vartheta \in \Theta\}$, Θ endl. dim.
<u>Ziel</u>: Rückschlüsse auf ϑ_0 $(F = F_{\vartheta_0})$

Definition: 2.6 Statistik

Sei Y Zufallsvariable mit Werten in $\mathcal{Y}$. Die messbare Abbildung $h : \mathcal{Y} \to \mathbb{R}^k$ heißt **Statistik**, falls sie nicht von den unbekannten Parametern abhängt.
Ist $Y = (X_1, X_2, ..., X_n)$ mit $X_1, X_2, ..., X_n$ iid, dann wird $h(Y)$ auch **Stichprobenstatistik** genannt.

Definition: 2.7

Sei $X_1, X_2, ..., X_n$ eine Stichprobe. Die Zufallsvariable

$$a_k = \frac{1}{n} \sum_{i=1}^{n} X_i^k$$

heißt k-tes Stichprobenmoment. Insbesondere heißen

$$a_1 = \frac{1}{n} \sum_{i=1}^{n} X_i =: \bar{X}_n \text{ bzw. } \bar{X} \quad \underline{\text{Stichprobenmittel}}$$

$$b_k = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X}_n)^k \quad \underline{\text{zentrales k-tes Stichprobenmittel}}$$

$$\hat{S_n}^2 = \frac{n}{n-1} b_2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2 \quad \underline{\text{Stichprobenvarianz}}$$

Definition: 2.8

Sei $(X_1, Y_1), ..., (X_n, Y_n)$ Stichprobe einer bivarianten Verteilung. Dann heißt

$$S_{11} = S_{XY} = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y}) \text{ Stichprobenkovarianz}$$

und

$$\hat{\rho}_n = \frac{S_{11}}{\sqrt{S_1^2}\sqrt{S_2^2}} \text{ Stichprobenkorrelationskoeffizient/ Pearson-Koeffizient}$$

mit

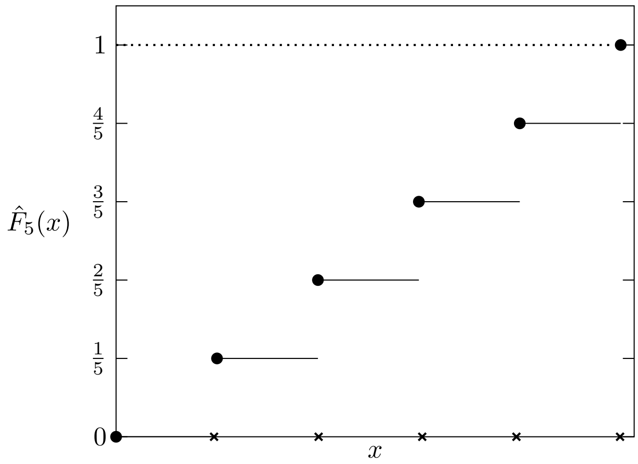
$$S_1^2 = S_{XX} = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2, S_2^2 = S_{YY} = \frac{1}{n-1} \sum_{i=1}^{n} (Y_i - \bar{Y})^2$$

Bemerkung (2.9)

Sei $(x_1, x_2, ..., x_n) = (X_1(w), X_2(w), ..., X_n(w))$ Realisationen von $(X_1, X_2, ..., X_n)$. Die Abbildung

$$\hat{F_n} : \mathbb{R} \to \mathbb{R} \text{ mit } \hat{F_n}(x) = \frac{1}{n} \sum_{i=1}^{n} 1(x_i \leq x)$$

heißt empirische Verteilungsfunktion.



$\hat{F_n}$ ist Verteilungsfunktion der Gleichverteilung (oder Laplace-Verteilung) auf $\{x_1, x_2, ..., x_n\}$ vorausgesetzt $x_i \neq x_j \; \forall i \neq j$.
Gilt $Y \sim \hat{F_n}$ dann gilt

$$\Rightarrow EY = \sum_{i=1}^{n} x_i P(X_i = x_i) = \frac{1}{n} \sum_{i=1}^{n} x_i = \bar{x}$$

$$EY^k = \sum_{i=1}^{n} x_i^k \cdot \frac{1}{n} = a_k \quad \text{(k-tes Stichprobenmoment)}$$

Theorem: 2.10

Mit der Notation wie in (2.7), (2.8) gilt

$$(1) \quad E\bar{X}_n = EX_1, \quad Var(X_n) = \frac{Var(X_1)}{n}$$

$$(2) \quad E\hat{S_n^2} = Var(X_1), \quad Var(\hat{S_n^2}) = \frac{1}{n} E[(X_1 - EX_1)^4] + \frac{3-n}{n(n-1)} (VarX_1)^2$$

$$(3) \quad ES_{11} = Cov(X_1, Y_1)$$

Lemma: 2.11

Sei $X_1, X_2, ..., X_n$ Stichproben, $E(X_1^{2k}) < \infty, k \in \mathbb{N}$.

$$m_k := EX_1^k, \quad m_{2k} = EX_1^{2k}.$$

Dann gilt mit $n \to \infty$ für $a_k = \frac{1}{n}\sum_{i=1}^n X_i^k$:

$$\sqrt{n}\frac{a_k - m_k}{\sqrt{m_{2k} - m_k^2}} \xrightarrow{\text{D}} N(0,1).$$

Bemerkung (2.12)
(1) Spezialfall: $\sqrt{n}(\bar{X}_n - EX_1) \xrightarrow{\text{D}} N(0, Var(X_1))$ ($\bar{X}_n$ Schätzer für EX_1, $Var(X_1)$ unbekannt)

$$\sqrt{n}\frac{\bar{X}_n - EX_1}{\sqrt{\hat{S}_n^2}} = \sqrt{n}\frac{\bar{X} - EX_1}{\sqrt{VarX_1}} \cdot \frac{\sqrt{VarX_1}}{\sqrt{\hat{S}_n^2}} \xrightarrow{\text{D}} N(0,1)$$

Dies gilt weil

$$\sqrt{n}\frac{\bar{X}_n - EX_1}{\sqrt{VarX_1}} \xrightarrow{\text{D}} N(0,1) \quad \text{(ZGW)}$$

$$\hat{S}_n^2 \xrightarrow{\text{P}} VarX_1$$

$$\overset{\text{cont.}}{\underset{\text{mapping}}{\Rightarrow}} \sqrt{\frac{VarX_1}{\hat{S}_n^2}} \xrightarrow{\text{P}} 1 \quad \text{und Slutskys Lemma } (X_n \xrightarrow{\text{D}} X, Y_n \xrightarrow{\text{P}} c, X_nY_n \xrightarrow{\text{D}} X \cdot c)$$

(2) Verteilung der Zentralen Sitchprobenmomente

$$\frac{1}{n}\sum (X_i - \bar{X})^k \text{ siehe (2.13)(1) für } k = 2$$

Lemma: 2.13

(1)
$X_1, X_2, ..., X_n$ iid mit $EX_1^4 < \infty$, $VarX_1 = \sigma^2 > 0$. Dann gilt

$$\sqrt{n}\left(\frac{1}{n}\sum (X_i - \bar{X})^2 - \sigma^2\right) = \sqrt{n}\left(\frac{n-1}{n}\hat{S}_n^2 - \sigma^2\right) \xrightarrow{\text{D}} N(0, \mu_4 - \mu_2^2)$$

mit

$$\mu_j := E((X_1 - EX_1)^j) \text{ (zentriertes } j\text{-tes Moment)}$$

(2)
$(X_1, Y_1), ..., (X_n, Y_n)$ iid $N_2\left(\begin{pmatrix}\mu_1\\ \mu_2\end{pmatrix}, \begin{pmatrix}\sigma^2 & \rho_{\sigma\tau}\\ \rho_{\tau\sigma} & \tau^2\end{pmatrix}\right)$ mit $VarX_1 = \sigma^2$, $VarY_1 = \tau^2$, $\rho = Corr(X_1, Y_1)$

$$\hat{\rho} = \frac{S_{11}}{\sqrt{S_1^2}\sqrt{S_2^2}} \quad \underline{\text{Stichprobenkorrelationskoeffizient}}$$

Dann gilt

$$\sqrt{n}(\hat{\rho} - \rho) \xrightarrow{\text{D}} N(0, (1-\rho^2)^2)$$

Theorem: 2.14

$X_1, X_2, ..., X_n$ iid $\sim N(\mu, \sigma^2)$, $\hat{S}^2 = \frac{1}{n-1}\sum (X_j - \bar{X})^2$. Es gilt

$$\begin{aligned}
&1)\quad \bar{X} \sim N(\mu, \frac{\sigma^2}{n})\\
&2)\quad \frac{(n-1)\hat{S}_n^2}{\sigma^2} = \frac{1}{\sigma^2}\sum_{i=1}^n (X_i - \bar{X})^2 \sim \chi^2_{n-1}\\
&3)\quad \bar{X} \perp\!\!\!\perp \hat{S}_n^2\\
&4)\quad \frac{\sqrt{n}(\bar{X} - \mu)}{\sqrt{\hat{S}_n^2}} \sim t_{n-1}
\end{aligned}$$

Theorem: 2.15

Seien $X_1, X_2, ..., X_m, Y_1, Y_2, ..., Y_n$ unabhängig. $X_1, X_2, ..., X_m$ iid $N(\mu_1, \sigma_1^2)$ und $Y_1, Y_2, ..., Y_n$ iid $N(\mu_2, \sigma_2^2)$

$$S_1^2 = \frac{1}{m-1}\sum_{i=1}^{m}(X_i - \bar{X})^2, S_2^2 = \frac{1}{n-1}\sum_{j=1}^{n}(Y_j - \bar{Y})^2$$

Dann gilt:
(a)

$$\frac{S_1^2}{\sigma_1^2} \div \frac{S_2^2}{\sigma_2^2} = \frac{S_1^2\sigma_2^2}{S_2^2\sigma_1^2} \sim F(m-1, n-1) \text{ } F\text{-Verteilung}$$

(b)

$$\frac{\bar{X}_m - \bar{Y}_n - (\mu_1 - \mu_2)}{\sqrt{(m-1)\frac{S_1^2}{\sigma_1^2} + (n-1)\frac{S_2^2}{\sigma_2^2}}} \cdot \sqrt{\frac{m+n-2}{\sigma_1^2/m + \sigma_2^2/n}} \sim t_{m+n-2}$$

Bemerkung
Für $\sigma_1^2 = \sigma_2^2$ gilt vereinfacht sich die Statistik in (b) zu

$$\frac{\bar{X}_m - \bar{Y}_n - (\mu_1 - \mu_2)}{\sqrt{(m-1)S_1^2 + (n-1)S_2^2}} \cdot \sqrt{\frac{mn(m+n-2)}{m+n}}$$

Kapitel 3

Schätzen

3.1 Eigenschaften von Punktschätzern

Bemerkung (3.1)
In diesem Kapitel: X Zufallsvariable mit Werten in $\mathcal{X}$, $X \sim P^X \in \{P_\vartheta : \vartheta \in \Theta\}, \Theta \subseteq \mathbb{R}^n$, Statistik $T : \mathcal{X} \to \mathbb{R}^l$.
Ziel: Schätzer für $\vartheta, \psi(\vartheta)$ basierend auf Stichprobe $X_1, X_2, ..., X_n$
Bezeichnung: E_ϑ, Var_ϑ etc. 'bezüglich' P_ϑ

Definition: 3.2

Betrachte $\psi : \Theta \subseteq \mathbb{R}^m \to \Gamma \subseteq \mathbb{R}^l : \vartheta \mapsto \psi(\vartheta)$
Jede Statistik $\delta : \mathcal{X} \to \Gamma$, $x \mapsto \delta(x)$ heißt (Punkt)schätzer (oder Schätzfunktion) für $\psi(\vartheta)$

$$\vartheta(x) = \vartheta(X(w))$$

heißt Schätzwert (estimator) oder Schätzung (estimate)

Beispiel (3.3)
$X_1, X_2, ..., X_n \sim \text{Bin}(1, \vartheta)$
(a)
$\psi(\vartheta) = \vartheta$, $\Theta = \Gamma = [0, 1]$
Mögliche Schätzer:

$$\begin{aligned} \delta_1(X_1, X_2, ..., X_n) &= \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \\ \delta_2(X_1, X_2, ..., X_n) &= \frac{1}{2} \\ \delta_3(X_1, X_2, ..., X_n) &= X_1 \end{aligned}$$

(b)
$\psi(\vartheta) = Var_\vartheta(X_1) = \vartheta(1 - \vartheta)$

Möglicher Schätzer:

$$\alpha_i = \delta_i(X_1, X_2, ..., X_n) \cdot (1 - \delta_i(X_1, X_2, ..., X_n)) \quad i = 1, 2, 3$$

Definition: 3.4 Erwartungstreue, Verzerrung

$\delta(X)$ sei Schätzer für $\psi(\vartheta)$
<u>Bias</u> oder Verzerrung von $\delta(X)$

$$b(\delta, \psi(\vartheta)) = E_\vartheta(\delta(X)) - \psi(\vartheta) \qquad (E\hat{\psi} - \psi)$$

δ heißt <u>erwartungstreu</u> oder <u>unverzerrt</u>/ <u>unbiased</u> falls

$$E_\vartheta(\delta(X)) = \psi(\vartheta) \qquad (bias = 0)$$

<u>Mittlere quadratische Abweichung</u> (<u>M</u>ean <u>S</u>quared <u>E</u>rror)

$$MSE_\vartheta(\delta) = E_\vartheta\left((\delta(X) - \psi(\vartheta))^2\right) = Var_\vartheta(\delta(X)) + b(\delta, \psi(\vartheta))^2$$

denn

$$\begin{aligned} E(\delta - \psi)^2 &= E(\delta - E\delta + E\delta - \psi)^2 \\ &= E(\delta - E\delta)^2 + E((\underbrace{E\delta - \psi}_{bias})^2) + 2E(\underbrace{(\delta - E\delta)}_{E=0}\underbrace{(E\delta - \psi)}_{\text{konst}}) \\ &= Var\delta + bias^2 \end{aligned}$$

Beispiel (3.5)
$X := X_1, X_2, ..., X_n$ iid $\sim U[0, \vartheta]$, $f_\vartheta = \frac{1}{\vartheta} 1_{[0,\vartheta]}(x)$, $EX_1 = \frac{\vartheta}{2}$
Mögliche Schätzer:
$\delta_1(X) = 2\bar{X} = 2\frac{1}{n}\sum_{i=1}^n X_i$ ($\rightsquigarrow$ Momentenmethode)
$\delta_2(X) = \max\limits_{i=1,\dots,n} X_i =: X_{(n)}$ ($\rightsquigarrow$ Max. Likelihood)

$$E\delta_1(X) = E(2\bar{X}) = 2\frac{1}{n}\sum_{i=1}^n \underbrace{EX_1}_{\frac{\vartheta}{2}} = \vartheta$$

erwartungstreu, d.h. $bias = 0$

$$MSE(\delta_1(X)) = Var(\delta_1(X)) = Var\left(2\frac{1}{n}\sum_{i=1}^n X_i\right) = \frac{4}{n^2}\sum_{i=1}^n \underbrace{VarX_i}_{\frac{1}{12}(\vartheta-0)^2} = \frac{4}{n^2} n \frac{1}{12}\vartheta^2 = \frac{\vartheta^2}{3n}$$

Dichte von $X_{(n)}$ (Übung $\frac{d}{dx}P(X_1 \leq x, X_2 \leq x, ..., X_n \leq x)$), $x \in (0, \vartheta)$

$$\begin{aligned} f_{X_{(n)}} &= n(F_\vartheta(x))^{n-1} f_\vartheta(x) \text{ mit } F_\vartheta(x) = \int_0^x \frac{1}{\vartheta} dt = \frac{x}{\vartheta} \\ &= n\left(\frac{x}{\vartheta}\right)^{n-1} \frac{1}{\vartheta} 1_{[0,\vartheta]}(x) \\ &= n\frac{x^{n-1}}{\vartheta^n} 1_{[0,\vartheta]}(x) \end{aligned}$$

$$E_\vartheta(\delta_2(X)) = \int_0^\vartheta x \cdot n\frac{x^{n-1}}{\vartheta^n}dx = \frac{n}{\vartheta^n}\int_0^\vartheta x^n dx = \frac{n}{\vartheta^n}\left[\frac{x^{n+1}}{n+1}\right]_0^\vartheta$$
$$= \frac{n}{\vartheta^n}\frac{\vartheta^{n+1}}{n+1} = \frac{n}{n+1}\vartheta \neq \vartheta \Rightarrow \text{ nicht erwartungstreu}$$

$$b(\delta_2(X),\vartheta) = E_\vartheta(\delta_2(X),\vartheta) - \vartheta = \vartheta\left(\frac{n}{n+1} - 1\right) = \vartheta\frac{n-(n+1)}{n+1}$$
$$= -\frac{1}{n+1}\vartheta \overset{n\to\infty}{\longrightarrow} 0 \text{ 'asymptotisch erwartungstreu'}$$

$$Var(\delta_2(X)) = \int_0^\delta x^2 n\frac{x^{n-1}}{\vartheta^n}dx - \left(\frac{n\vartheta}{n+1}\right)^2 \overset{\text{Ü}}{=} \frac{n\vartheta^2}{(n+1)^2(n+2)}$$

$$MSE(\delta_2) = Var(\delta_2) + (bias(\delta_2))^2 = \frac{n\vartheta^2 + \vartheta^2(n+2)}{(n+1)^2(n+2)} = \frac{\vartheta^2(2n+2)}{(n+1)^2(n+2)}$$
$$= \frac{2\vartheta^2}{(n+1)(n+2)} = O\left(\frac{1}{n^2}\right)$$

$MSE(\delta_1)$ war $O\left(\frac{1}{n}\right)$, d.h. δ_2 konvergiert schneller als δ_1.
$MSE(\delta_2) < MSE(\delta_1)$ für $n >?$

$$\Leftrightarrow \frac{2\vartheta^2}{(n+1)(n+2)} < \frac{\vartheta^2}{3n} \Leftrightarrow 6n < (n+1)(n+2) = n^2 + 3n + 2$$
$$\Leftrightarrow 3n < n^2 + 2 \text{ also für } n \geq 3$$

Bemerkung (3.6)
Sei δ Schätzer für $\psi(\vartheta)$ mit Werten in Γ.
Verlustfunktion $L : \Theta \times \Gamma \to \mathbb{R}_{\geq 0}$ (z.B. $L(\vartheta,\gamma) = |\psi(\vartheta) - \gamma|^r$), γ 'Schätzwert' mit $L(\vartheta,\psi(\vartheta) = 0 \;\forall\, \vartheta \in \Theta$
Risiko $R(\vartheta,\delta) = E_\vartheta(L(\vartheta,\delta(X)))$ (erwarteter Verlust) z.B. $E_\vartheta(\psi(\vartheta) - \delta(X))^2)$ MSE
Mögliche Ansätze für Schätzer: minimiere Risiko
z.B. Minimax-Ansatz: Wähle das δ welches das maximale Risiko $\sup_{\vartheta\in\Theta} R(\vartheta,\delta)$ minimiert
Bayes-Ansatz: minimiere $\int_\Theta R(\vartheta,\delta)w(\vartheta)d\vartheta$ bezüglich δ ($w(\vartheta)$ 'Vorinformation') 'Bayes-Risiko'

Definition: 3.7 asymptotische Eigenschaften

$X_1, X_2, ..., X_n$ Folge von Zufallsvariablen, $X^{(n)} := (X_1, X_2, ..., X_n)$ mit Werten in $\mathcal{X}_n$, $\delta_n : \mathcal{X} \to \Gamma$ sei Schätzer für $\psi(\vartheta), \psi : \Theta \to \Gamma$.
$\delta_n(X^{(n)})$ heißt
(a) asymptotisch erwartungstreu, falls für alle $\vartheta \in \Theta$ gilt

$$\lim_{n\to\infty} b(\delta_n, \psi(\vartheta)) = 0 \text{ d.h. } E_\vartheta(\delta_n(X^{(n)})) \overset{n\to\infty}{\longrightarrow} \psi(\vartheta)$$

(b) (schwach) konsistent, falls für alle $\vartheta \in \Theta$ gilt:

$$\delta_n(X^{(n)}) \overset{\text{P}}{\to} \psi(\vartheta) \text{ d.h. } \lim_{n\to\infty} P(||\delta_n(X^{(n)}) - \psi(\vartheta)|| \geq \varepsilon) = 0 \;\forall \varepsilon > 0$$

(stark konsistent, falls fast sichere Konvergenz)

Beispiel (3.8 Fortsetzung von 3.3a und 3.3b)
(a)
$\delta_1 = \bar{X}_n$ ist erwartungstreu und konsistent für ϑ (LLN Law-of-Large-Numbers)
$\delta_2 = \frac{1}{2}$ ist weder erwartungstreu noch konsistent
$\delta_3 = X_1$ ist erwartungstreu aber nicht konsistent, weil $P(|X_1 - \vartheta| > \varepsilon) \not\to 0$
(b)
$\alpha_1 = \bar{X}_n(1 - \bar{X}_n)$ ist konsistent da stetige Funktion von $\bar{X}$ (continuous mapping)
α_1 ist amsymptotisch erwartungstreu (Ü)
(c)
Sei $X_1, X_2, ..., X_n$ iid $E|X|^k < \infty$
$\frac{1}{n}\sum_{i=1}^n X_i^k \xrightarrow{\text{P}} EX_1^k$ (LLN) d.h. $\frac{1}{n}\sum_{i=1}^n X_i^k$ ist erwartungstreu und konsistent für $E(X_1^k)$

Lemma: 3.9

$\delta_n = \delta_n(X_1, X_2, ..., X_n)$ sei Folge von asymptotisch erwartungstreuer Schätzer für $\psi(\vartheta)$, dann gilt

$$\lim_{n\to\infty} Var_\vartheta(\delta_n) = 0 \text{ impliziert } \delta_n \text{ konsistent für } \psi(\vartheta)$$

Beispiel (3.10 Erwartungstreue? Konsistenz?)
Betrachte $X_1, X_2, ..., X_n$ iid, $VarX_1 < \infty$
(a) $\delta_2 = X_{(n)} = \max\limits_{i=1,...,n} X_i$, $X_i \sim U[0, \vartheta]$ (siehe Beispiel 3.5) ist nicht erwartungstreu, ist konsistent
(b) $\bar{X}_n = \frac{1}{n}\sum\limits_{i=1}^n X_i$ ist erwartungstreu, ist konsistent, da

$$E\bar{X} = EX_1, Var\bar{X}_n = \frac{VarX_1}{n}$$

(c) $\hat{S}_n^2 = \frac{1}{n-1}\sum (X_i - \bar{X})^2$ Stichprobenvarianz (Annahme $EX_1^4 < \infty$)

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

$$\begin{aligned}
E\hat{\sigma}^2 &= \frac{1}{n}\sum_{i=1}^n E(X_i - \bar{X})^2 \\
&= \frac{1}{n}\sum_{i=1}^n E(X_i - EX_1 + EX_1 - \bar{X})^2 \\
&= \frac{n-1}{n} VarX_1 \text{ (Ü bzw. siehe 2.10 (3) } ES_{11} = ...) \\
\Rightarrow E\hat{S}_n^2 &= \frac{n}{n-1}E\hat{\sigma}^2 \\
&= \frac{n}{\cancel{n-1}}\frac{\cancel{n-1}}{n} VarX_1 = VarX_1 \text{ erwartungstreu}
\end{aligned}$$

$$Var\hat{S}_n^2 \overset{\text{Ü}}{=} \frac{1}{n}E((X_1 - EX_1)^4) + \frac{3-n}{n(n-1)}(VarX_1)^2 = O\left(\frac{1}{n}\right) \overset{n\to\infty}{\to} 0$$

also auch konsistent.

Kapitel 4

Suffizienz und Vollständigkeit

Beispiel (4.1)
$X_1, X_2, ..., X_n$ Bin$(1, \vartheta)$ z.B. Daten: 00101110101100001
Frage: Enthält $\hat{\vartheta} = \frac{1}{n}\sum_{i=1}^n X_i$ alle Informationen über ϑ? Oder gibt es zusätzliche Informationen in der Stichprobe?

$$T(X) = \sum_{i=1}^n X_i, \quad u_i \in \{0,1\}$$

$$\begin{aligned}
P_\vartheta(X_1 = u_1, ..., X_n = u_n | T(X) = k) &= \frac{P_\vartheta(X_1 = u_1, ..., X_n = u_n, T(X) = k)}{P_\vartheta(T(X) = k)} (= 0 \text{ für } \sum u_i \neq k) \\
&= \frac{P(X_1 = u_1, .., X_n = u_n)}{P_\vartheta(T(X) = k)} \text{ für } \sum u_i = k \\
&= \frac{\prod_{i=1}^n P_\vartheta(X_i = u_1)}{P_\vartheta(T(X) = k)} \\
&= \frac{\prod_{i=1}^n \vartheta^{u_i}(1-\vartheta)^{1-u_i}}{\binom{n}{k}\vartheta^k(1-\vartheta)^{n-k}} \\
&= \frac{\vartheta^{\overbrace{\Sigma u_i}^{k}}(1-\vartheta)^{n-\overbrace{\Sigma u_i}^{k}}}{\binom{n}{k}\vartheta^k(1-\vartheta)^{n-k}} \\
&= \begin{cases} \frac{1}{\binom{n}{k}} & \text{für } \sum u_i = k \\ 0 & \text{sonst} \end{cases}
\end{aligned}$$

Die verbleibende Information ist also unabhängig von $\vartheta \in \Theta = (0,1)$.
(Zähl-)Dichte von $(X_1, X_2, ..., X_n)$:

$$f_\vartheta(x_1, ..., x_n) \overset{\text{oben}}{=} \vartheta^{\Sigma x_i}(1-\vartheta)^{n-\Sigma x_i} = \left(\frac{\vartheta}{1-\vartheta}\right)^{\Sigma x_i}(1-\vartheta)^k$$

Definition: 4.2

Eine Statistik $T = T(X_1, ..., X_n)$ heißt <u>suffizient</u> für ϑ (oder $\mathcal{P} = \{P_\vartheta : \vartheta \in \Theta\}$), falls die bedingte Verteilung von $(X_1, X_2, ..., X_n)$ gegeben $T(X) = t$ (fest) unabhängig von ϑ ist.

Bemerkung (4.3)
Betrachte $T(X)$ statt $X = (X_1, X_2, ..., X_n)$ (keine zusätzliche Info über ϑ)$\longrightarrow$ Datenreduktion
Wünschenswert: maximale Datenreduktion $\longrightarrow$ 'Minimal-Suffizienz' (später)
$T(X_1, X_2, ..., X_n) = (X_1, X_2, ..., X_n)$ ist suffizient, soll aber ausgeschlossen werden.
Beispiel

$$P(X = x|X = t) = \frac{P(X = x, X = t)}{P(X = t)} = \begin{cases} \frac{P(X=t)}{P(X=t)} = 1 & x = t \\ 0 & \text{sonst} \end{cases}$$

Maßtheorie: Suffizient = Eigenschaft von σ-Alg.

Beispiel (4.4)
$X, Y \sim \text{Pois}(\lambda)$, X, Y unabhängig, $T = X + Y$
$P^{X|T=t} = \text{Bin}(t, \frac{1}{2})$:

$$\begin{aligned} P(X = x|T = t) &= \frac{P(X = x, X + Y = t)}{P(X + Y = t)} \\ &= \frac{P(X = x, Y = t - x)}{P(X + Y = t)} \\ &= \frac{P(X = x)P(Y = t - x)}{P(\underbrace{X + Y}_{\sim \text{Pois}(2\lambda)} = t)} \\ &= \frac{e^{-\lambda}\frac{\lambda^x}{x!}e^{-\lambda}\frac{\lambda^{t-x}}{(t-x)!}}{e^{-2\lambda}\frac{(2\lambda)^t}{t!}} \\ &= \frac{e^{-2\lambda}t!\lambda^t}{e^{-2\lambda}x!(t-x)!(2\lambda)^t} \\ &= \binom{t}{x}\left(\frac{1}{2}\right)^t = \binom{t}{x}\left(\frac{1}{2}\right)^x\left(1 - \frac{1}{2}\right)^{t-x} \\ &\text{Bin}\left(t, \frac{1}{2}\right) \text{ Dichte für } x = 0, 1, ..., t \end{aligned}$$

Analog

$$\begin{aligned} P(X = x, Y = y|T = t) &= \frac{P(X = x, Y = y, X + Y = t)}{P(X + Y = t)} \\ &= \frac{P(X = x, Y = t - x, X + Y = t)}{P(X + Y = t)} \\ &\overset{\text{oben}}{=} \binom{t}{x}\left(\frac{1}{2}\right)^t \text{ für } X + Y = t, \quad 0 \text{ sonst} \end{aligned}$$

Die bedingte Verteilung von (X, Y) gegeben T ist $\text{Bin}\big(t, \frac{1}{2}\big)$, also unabhängig vom Parameter λ

$$\Rightarrow T \text{ ist suffizient für } \lambda$$

Theorem: 4.5 Faktorisierungssatz (Neyman-Kriterium)

Sei X stetige oder diskrete Zufallsvariable mit Dichte f_ϑ, $\vartheta \in \Theta$, $T : \mathcal{X} \to \mathbb{R}$ Statistik.
Dann gilt

$$\begin{aligned}
&T \text{ ist suffizient für } \vartheta \\
&\Leftrightarrow \exists \text{ Funktion } h : \mathcal{X} \to \mathbb{R}_{\geq 0} \text{ unabhängig von } \vartheta \\
&\text{ und zu jedem } \vartheta \in \Theta \text{ gibt es eine Funktion } g_\vartheta : \mathcal{T} \to \mathbb{R}_{\geq 0} \text{ mit} \\
&(*) \quad f_\vartheta(x) = g_\vartheta(T(x))h(x) \text{ für alle } x \in \mathcal{X}, \vartheta \in \Theta
\end{aligned}$$

Bemerkung (4.6)
Für den allgemeinen maßtheoretischen Beweis siehe Witting, Mathematische Statistik I, Seite 343.

Lemma: 4.7

Sei $T : \mathcal{X} \to \mathcal{T}$ suffizient für $\vartheta \in \Theta$, $k : \mathcal{T} \to \mathcal{T}^*$ bijektiv (messbar) dann gilt

$$T^* = k \circ T \text{ ist suffizient für } \vartheta$$

Beispiel (4.8)
(a)
In Beispiel 4.1: Dichte von $X = (X_1, X_2, ..., X_n)$, $X_1, X_2, ..., X_n$ iid $\sim$ Bin$(1, \vartheta)$,
$f_\vartheta(x) = \underbrace{\left(\frac{\vartheta}{1-\vartheta}\right)^{T(x)} (1-\vartheta)^n}_{g_\vartheta(T(X))} \cdot \underbrace{1}_{h(x)}$ mit $T(x) = \sum_{i=1}^n X_i \overset{4.5}{\Rightarrow} T$ suffizient
(b)
$X = (X_1, X_2, ..., X_n)$ mit $X_1, X_2, ..., X_n$ iid $\sim N(\mu, \sigma^2)$, $\vartheta = (\mu, \sigma^2)$

$$\begin{aligned}
f_\vartheta(x) &= \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp\left(-\frac{1}{2\sigma^2} \sum (X_i - \mu)^2\right) \\
&= \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp\left(-\frac{1}{2\sigma^2} \underbrace{\sum X_i^2}_{T_2} - \frac{1}{2\sigma^2} \sum \mu^2 + \frac{2}{2\sigma^2}\mu \underbrace{\sum X_i}_{T_1}\right) \\
&= g_\vartheta(T_1(x), T_2(x)) \cdot 1
\end{aligned}$$

$\Rightarrow T = (T_1, T_2)$ ist suffizient für $\vartheta = (\mu, \sigma^2)$

Definition: 4.9

Sei X Zufallsvariable (diskret/ stetig) mit Dichte $f \in \{f_\vartheta : \vartheta \in \Theta\}$, $T : \mathcal{X} \to \Gamma \subseteq \mathbb{R}^k$ Statistik.
T heißt <u>vollständig</u>, falls für jede (messbare) Funktion $g : \Gamma \to \mathbb{R}$ gilt:

$$E_\vartheta(g(T(X))) = 0 \text{ für alle } \vartheta \in \Theta \Rightarrow P_\vartheta(g(T(X)) = 0) = 1 \text{ für alle } \vartheta$$

d.h. g verschwindet fast sicher auf $T(\mathcal{X})$.

Beispiel (4.10)
(a)
$X_1, X_2, ..., X_n$ iid $\sim \text{Bin}(1, \vartheta)$, $\vartheta \in \Theta = (0,1)$
$T(X) = \sum_{i=1}^n X_i$ mit $T \sim \text{Bin}(n, \vartheta)$ ist suffizient.

$$E_\vartheta(g(T(X))) = \sum_{t=0}^n g(t) \binom{n}{t} \vartheta^t (1-\vartheta)^{n-t}$$

ist Polynom n-ten Grades in ϑ. Also ist das Polynom nur dann $\equiv 0$ wenn alle Koeffizienten gleich 0 sind, d.h.

$$g(t) = 0 \text{ für alle } t \in \{0, 1, ..., n\}$$

$\Rightarrow T$ ist vollständig.
(b)
$X_1, X_2, ..., X_n$ iid $\sim N(\vartheta, \vartheta^2)$
Bekannt: $T(X) = (\sum X_i, \sum X_i^2)$ ist suffizient (4.8), aber
T ist nicht vollständig:

$$\exists g(u,v) = 2u^2 - (n+1)v \neq 0$$

$$\begin{aligned} E_\vartheta(g(T(X))) &= E_\vartheta(g(\sum X_i, \sum X_i^2)) \\ &= E_\vartheta\left(2(\sum X_i)^2 - (n+1)\sum X_i^2\right) \\ &\overset{\text{Ü}}{=} 0 \end{aligned}$$

Theorem: 4.11

Sei $X : \Omega \to \mathbb{R}^n$ diskrete oder stetige Zufallsvariable mit Dichte $f_\vartheta(x) : \mathbb{R}^n \to \mathbb{R}$ und Träger $A = \{x \in \mathbb{R}^n : f_\vartheta(x) > 0\}$ der nicht von ϑ abhängt (also z.B. <u>nicht</u> $U(0, \vartheta)$).
Angenommen es gibt Abbildungen $Q_1, ..., Q_k, D : \Theta \to \mathbb{R}$ und $T_1, ..., T_k, S : \mathbb{R}^n \to \mathbb{R}$ $(k \leq n)$, sodass

$$f_\vartheta(x) = \exp\left(\sum_{i=1}^k Q_i(\vartheta) T_i(x) + D(\vartheta) + S(x)\right) \text{ für alle } \vartheta \in \Theta, x \in A$$

('k-parametrige Exponentialfamilie')
und die Menge

$$\mathcal{Q} = \{(Q_1(\vartheta), ..., Q_k(\vartheta)) : \vartheta \in \Theta\} \text{ ist offene Teilmenge von } \mathbb{R}^k$$

Dann ist

$$T = (T_1(X), ..., T_k(X))$$

suffizient und vollständig.

Beispiel (4.12)
Betrachte $X \sim N(\mu, \sigma^2)$, 2-parametrige Exponentialfamilie ($\vartheta = (\mu, \sigma^2)$):

$$\begin{aligned} f_\vartheta(x) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right) \\ &= \underbrace{\exp(\ln((2\pi\sigma^2)^{-\frac{1}{2}}))}_{\exp(-\frac{1}{2}\ln(2\pi\sigma^2))} \cdot \exp\left(-\frac{x^2}{2\sigma^2} + \frac{2x\mu}{2\sigma^2} - \frac{\mu^2}{2\sigma^2}\right) \\ &= \exp\left(\underbrace{-\frac{1}{2\sigma^2}}_{Q_1(x)} \underbrace{x^2}_{T_1(\vartheta)} + \underbrace{x}_{T_2} \underbrace{\frac{\mu}{\sigma^2}}_{Q_2} - \frac{1}{2}\left(\frac{\mu^2}{\sigma^2} + \ln(2\pi\sigma^2)\right)\right) \end{aligned}$$

Für $X_1, X_2, ..., X_n$ iid $N(\mu, \sigma^2)$ erhält man

$$T_1 = \sum X_i^2, \quad T_2 = \sum X_i \text{ mit } \mathcal{Q} = \mathbb{R}^- \times \mathbb{R} \text{ offen}$$

$\Rightarrow T = (T_1, T_2)$ suffizient und vollständig.

T vollständig $\Leftrightarrow \forall g$ gilt $E_\vartheta(g(T(X))) = 0 \Rightarrow P_\vartheta(g(T(X)) = 0) = 1 \ \forall\vartheta$

Bemerkung (4.13)
Suffiziente Statistiken sind <u>nicht</u> eindeutig.

Beispiel
$X_1, X_2, ..., X_n \overset{iid}{\sim} N(0, \sigma^2)$

$$\begin{aligned} f_{\sigma^2}(x) &= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2}\sum X_i^2\right) \cdot 1 \\ &= g_{\sigma^2}(T(x))h(x) \end{aligned}$$

Suffizient sind

$$\begin{aligned} T_1(X) &= (X_1^2, X_2^2, ..., X_n^2) \\ T_2(X) &= (X_1^2 + X_2^2, X_3^2, ..., X_n^2) \\ T_3(X) &= \sum_{i=1}^n X_i^2 \end{aligned}$$

T_1 ist nicht vollständig. Betrachte $g(x) = x_1 - x_2$.

$$Eg(T_1(X)) = E(X_1^2 - X_2^2) = 0 \text{ (obwohl } g \neq 0)$$

T_2 ist nicht vollständig. Betrachte $g(x) = x_1 - 2x_2$

$$Eg(T_2(X)) = E(X_1^2 + X_2^2 - 2X_3^2) = EX_1^2 + EX_2^2 - 2EX_3^2 = 0$$

T_3 ist vollständig, siehe 4.11/4.12

Definition: 4.14

Eine suffiziente Statistik T^* heißt minimal-suffizient für $\vartheta \in \Theta$, falls T^* über jede für ϑ suffiziente Statistik faktorisiert, d.h.

$$\forall\, T \text{ suffizient } \exists\, H : T^*(X) = H(T(X))$$

D.h. T^* ist suffizient und nimmt möglichst wenige Werte an.

Beispiel (4.15)
Sei $\Theta = (\vartheta_0, \vartheta_1, ..., \vartheta_k)$ und $\mathcal{P} = \{P_\vartheta : \vartheta \in \Theta\}$ eine endliche Familie von Verteilungen, mit Dichten f_{ϑ_j} für $j = 0, 1, ..., k$ die denselben Träger haben. Dann ist

$$T^*(X) = \left(\frac{f_{\vartheta_1}(X)}{f_{\vartheta_0}(X)}, ..., \frac{f_{\vartheta_k}(X)}{f_{\vartheta_0}(X)}\right)$$

minimal suffizient.

Begründung:
1.) Sei $h(x) = f_{\vartheta_0}(x)$ mit ϑ_0 fest, bekannt.
Definiere $g_{\vartheta_j}(x) = x_j$ für $j = 1, ..., k$, $x = (x_1, ..., x_k)$.
Dann gilt

$$f_{\vartheta_j} = \begin{cases} \frac{f_{\vartheta_j}(x)}{f_{\vartheta_0}(x)} f_{\vartheta_0}(x) = g_{\vartheta_j}(T^*(x))h(x) & j = 1, ..., k \\ h(x) \cdot 1 \text{ mit } g_{\vartheta_0}(T^*(x)) := 1 & j = 0 \end{cases}$$

Mit Theorem 4.5 folgt, dass T^* suffizient ist.
2.) Sei nun T beliebig suffizient. Mit Theorem 4.5 folgt

$$\exists\, \tilde{h}, \tilde{g}_\vartheta(T(x)) : f_\vartheta(x) = \tilde{g}_\vartheta(T(x))\tilde{h}(x) \quad \forall\, \vartheta$$

Daraus folgt

$$T^*(x) = \left(\frac{\tilde{g}_{\vartheta_1}(T(x))\cancel{\tilde{h}(x)}}{\tilde{g}_{\vartheta_0}(T(x))\cancel{\tilde{h}(x)}}, ..., \frac{\tilde{g}_{\vartheta_k}(T(x))\cancel{\tilde{h}(x)}}{\tilde{g}_{\vartheta_0}(T(x))\cancel{\tilde{h}(x)}}\right) = H(T(x))$$

mit

$$H(y) = \left(\frac{\tilde{g}_{\vartheta_1}(y)}{\tilde{g}_{\vartheta_0}(y)}, ..., \frac{\tilde{g}_{\vartheta_k}(y)}{\tilde{g}_{\vartheta_0}(y)}\right)$$

also ist T^* per Definition minimal-suffizient.

Theorem: 4.16

Suffiziente vollständige Statistiken sind minimal-suffizient.

Beispiel (4.17 nicht-parametrische suffiziente Statistik)
$X_1, X_2, ..., X_n$ stetige iid Zufallsvariablen mit Dichten $\vartheta \in \Theta = \{\vartheta : \mathbb{R} \to \mathbb{R} \text{ Dichtefunktionen}\}$ (∞-dimensional)
Die Ordnungsstatistik

$$T(X_1, X_2, ..., X_n) = (X_{(1)}, ..., X_{(n)}) \text{ mit } X_{(1)} \leq X_{(2)} \leq ... \leq X_{(n)}$$

ist suffizient für ϑ;
Die gemeinsame Dichte von $X = (X_1, X_2, ..., X_n)$ ist:

$$f_\vartheta(x_1, ..., x_n) = \prod_{i=1}^{n} \vartheta(x_i) = \underbrace{\prod_{i=1}^{n} \vartheta(x_{(i)})}_{g_\vartheta(T(x))} \cdot \underbrace{1}_{h(x)}$$

Kapitel 5

Erwartungstreue Schätzer

Definition: 5.1

(a)
Bekannt aus Kapitel 3: MSE oft nicht minimierbar.
Möglicher Ausweg: Beschränkung auf erwartungstreue Schätzer (Minimierung der Varianz).
Existiert ein erwartungstreuer Schätzer für $\psi(\vartheta)$, dann heißt ψ schätzbar.
(b)
$\psi : \Theta \to \mathbb{R}$ sei schätzbar. $\mathcal{U} := \left\{T : \mathcal{X} \to \mathbb{R} \text{ Schätzer}, E_\vartheta(T(x)) = \psi(\vartheta), E_\vartheta(T^2(x)) < \infty \ \forall \vartheta \in \Theta\right\}$ sei die Menge aller für $\psi(\vartheta)$ erwartungstreuen Schätzer mit existenter Varianz.
(Einschub: Das k-te Moment existiert, falls $E|X|^k < \infty$)
$T_0 \in \mathcal{U}$ heißt Uniformly Mimimum Variance Unbiased (UMVU) Schätzer (oder UMVUE), d.h. T_0 hat die gleichmäßig kleinste Varianz unter den erwartungstreuen Schätzern, falls gilt

$$E_\vartheta[(T_0(x) - \psi(\vartheta))^2] \leq E_\vartheta[(T(x) - \psi(\vartheta))^2] \quad \forall\, \vartheta \in \Theta, T \in \mathcal{U}$$

(c)
Beachte $E_\vartheta[(T(X) - \psi(\vartheta))^2] = Var_\vartheta(T(X)) = MSE_\vartheta(T(X))$ für alle $T \in \mathcal{U}$ (da erwartungstreu)

Bemerkung (5.2a)
Nicht jede Funktion ist schätzbar.

Beispiel
$X \sim Bin(1, p)$, $\psi(p) = p^2$.
Angenommen es gibt ein Schätzer $T(X)$ der erwartungstreu ist, dann gilt

$$\begin{aligned} p^2 &= E_p T(X) = T(1) \underbrace{P(X=1)}_{p} + T(0) \underbrace{P(X=0)}_{1-p} \\ \Leftrightarrow p^2 &= T(1)p + T(0) - T(0)p \\ \Leftrightarrow p^2 &- p(T(1) - T(0)) - T(0) = 0 \quad \underline{\forall p \in (0,1)} \end{aligned}$$

D.h. $\equiv 0$ unmöglich.

Bemerkung (5.2b)

Nicht jeder erwartungstreue Schätzer ist sinnvoll.

Beispiel
$X \sim \text{Pois}(\lambda)$, $\psi(\lambda) = e^{-3\lambda}$. $T(X) = (-2)^x$ ist erwartungstreu:

$$\begin{aligned} E_\lambda(T(X)) &= \sum_{k=0}^{\infty} (-2)^k \frac{\lambda^k}{k!} e^{-\lambda} \\ &= e^{-\lambda} \sum_{k=0}^{\infty} \frac{(-2\lambda)^k}{k!} \\ &= e^{-\lambda} e^{-2\lambda} = e^{-3\lambda} \end{aligned}$$

<u>Aber</u>: $\psi(\lambda) > 0$ für alle λ, aber $T(x) < 0$ für x ungerade.

Bemerkung (5.2c)
(i)
Sind T_1, T_2 fast sicher verschiedene, erwartungstreue Schätzer für $\psi(\vartheta)$, dann sind durch

$$R = \alpha T_1 + (1-\alpha) T_2 \quad \alpha \in (0,1)$$

unendlich viele erwartungstreue Schätzer gegeben.
Welches α ist sinnvoll?
(ii)
Sei T ein erwartungstreuer Schätzer für $\psi(\vartheta)$. Dann sind durch

$$T + V \quad V \neq 0, \quad E_\vartheta V = 0 \quad \forall \vartheta \in \Theta$$

<u>alle</u> erwartungstreuen Schätzer für $\psi(\vartheta)$ gegeben. (Angenommen $\tilde{T}$ erwartungstreu, $\tilde{T} = \underbrace{\tilde{T} - T}_{=:V} + T$)

Theorem: 5.3

Die Klasse $\mathcal{U}$ aller erwartungstreuen Schätzer sei nicht-leer.
(a)

$$\begin{aligned} &T_0 \in \mathcal{U} \text{ ist UMVUE} \\ &\Leftrightarrow E_\vartheta(T_0(x)V(x)) = 0 \;\forall\; \vartheta \in \Theta \text{ und} \\ &\quad \forall V \in \mathcal{U}_0 : \{V : \mathcal{X} \to \mathbb{R} : E_\vartheta V(X) = 0, Var_\vartheta V(x) \text{ existiert } \forall\; \vartheta\} \end{aligned}$$

(b)
Es gibt höchstens einen UMVU-Schätzer

Theorem: 5.4 Rao-Blackwell]

Sei $\mathcal{F} = \{F_\vartheta : \vartheta \in \Theta\}$ eine Familie von Verteilungen, ψ schätzbar (es existiert ein erwartungstreuer Schätzer) und $h \in \mathcal{U}$ ein erwartungstreuer Schätzer für $\psi(\vartheta)$.
$T = T(X)$ sei suffizient für $\mathcal{F}$. Dann ist

$$T_0 = E(h(X)|T)$$

unabhängig von ϑ und erwartungstreu für $\psi(\vartheta)$. Es gilt

$$E_\vartheta[(E(h(X)|T) - \psi(\vartheta))^2] \leq E_\vartheta[(h(X) - \psi(\vartheta))^2] \quad \forall \vartheta$$

mit Gleichheit genau dann wenn

$$P_\vartheta(h(X) = E(h(X)|T)) = 1 \quad \forall \vartheta$$

(d.h. T_0 ist gleichmäßig in ϑ gleich gut oder besser [im Bezug auf MSE/ Var] als $h(X)$)

Theorem: 5.5 Lehmann-Scheffé

Sei T suffizient und vollständig für ϑ, $h = h(X)$ ein erwartungstreuer Schätzer für $\psi(\vartheta)$. Dann gilt

$$T_0 = E[h(X)|T]$$

ist fast sicher eindeutiger UMVU Schätzer für $\psi(\vartheta)$.

Bemerkung (5.6a)
Sei T suffizient und vollständig, $h(X) = \tilde{h}(T)$ (d.h. h hängt nur von T ab) erwartungstreu für $\psi(\vartheta)$. Dann folgt mit (5.5), dass

$$E(h(X)|T) = E(\tilde{h}(T)|T) = \tilde{h}E(1|T) = \tilde{h}(T) = h(X)$$

ein UMVUE ist.
Also hängt ein Schätzer nur von T suffizient und vollständig ab, ist er UMVUE (einfach zu finden für exponentielle Familien siehe 4.11)

Beispiel
$X_1, X_2, ..., X_n \overset{iid}{\sim} N(\mu, \sigma^2)$, $(\sum X_i, \sum X_i^2)$ suffizient und vollständig (siehe 4.12)
Ü: $\left(\frac{1}{n}\sum X_i, \frac{1}{n-1}\sum (X_i - \bar{X})^2\right) = (\bar{X}, \hat{S}^2)$ suffizient und vollständig und erwartungstreu für (μ, σ^2). Also ist $(\bar{X}, \hat{S}^2)$ UMVUE

Bemerkung (5.6b)
(i)
$X_1, X_2, ..., X_n \overset{iid}{\sim} U(0, \vartheta)$, $X_{(n)} = \max\limits_{i=1,...,n} X_i$ ist suffizient und vollständig (Ü)
Gesucht: Erwartungstreuer Schätzer der nur von $X_{(n)}$ abhängt.
Bekannt: $E_\vartheta(X_{(n)}) = \frac{n\vartheta}{n+1} \Rightarrow T_0 = \frac{n+1}{n} X_{(n)}$ erwartungstreu für $\vartheta \overset{(a)}{\Rightarrow} T_0$ UMVUE.
Oder: $\delta_1(X) = \frac{2}{n}\sum X_i$ erwartungstreu für ϑ (siehe 3.5)

Ü: $\overset{5.5}{\Rightarrow} E(\delta_1(X)|X_{(n)}) = \frac{n+1}{n} X_{(n)}$ ist UMVUE.
(ii)
Bsp: Nicht-parametrischer UMVUE für Verteilungsfunktion F
$X_1, X_2, ..., X_n \overset{iid}{\sim} F$, $(X_{(1)}, ..., X_{(n)})$ suffizient (siehe 4.17) und vollständig für F (Ü)
Dann gilt: $h(X_1, X_2, ..., X_n) = 1_{\{X_1 \leq z\}}$ ist erwartungstreu:

$$E(1(X_1 \leq z)) = P(X_1 \leq z) = F(z)$$

$$\overset{5.5}{\Rightarrow} E(1(X_1 \leq z)|X_{(1)}, ..., X_{(n)}) = \frac{1}{n}\sum_{i=1}^{n} 1(X_{(i)} \leq z)$$

$$= \frac{1}{n}\sum_{i=1}^{n} 1(X_i \leq z) = \hat{F}_n(z) \text{ empirische Verteilungsfunktion ist UMVUE}$$

Theorem: 5.8 Cramer-Rao Ungleichung

X sei n-dim. Zufallsvariable mit Dichte f_ϑ (diskret oder stetig), $\vartheta \in \Theta$, $\Theta \subseteq \mathbb{R}^1$ offen
Annahme an f_ϑ:
(i) Träger hängt nicht von ϑ ab
(ii) $\frac{\partial}{\partial\vartheta} \ln f_\vartheta(x)$ existiert für alle x, ϑ
(iii) (a) $E\left(\frac{\partial}{\partial\vartheta} \ln f_\vartheta(x)\right) = 0$ und (b) $\frac{\partial}{\partial\vartheta} E_\vartheta(T(X)) = E_\vartheta\left(\frac{\partial}{\partial\vartheta} \ln f_\vartheta(X) \cdot T(X)\right)$,
dabei sei T eine Statistik mit $Var_\vartheta(T(X)) < \infty \forall\, \vartheta$.
Setze $\psi(\vartheta) := E_\vartheta(T(X))$ und

$$I_n(\vartheta) := E_\vartheta\left(\left(\frac{\partial}{\partial\vartheta} \ln f_\vartheta(X)\right)^2\right) \text{ 'Fischer-Information'}$$

Gilt $0 < I_n(\vartheta) < \infty$ für alle $\vartheta \in \Theta$, dann folgt

$$Var_\vartheta(T(X)) \geq \underbrace{\frac{(\psi'(\vartheta))^2}{I_n(\vartheta)}}_{\text{untere Schranke}} \qquad \forall\, \vartheta \in \Theta$$

Bemerkung (5.9)
(a)
Ist T erwartungstreu für $\psi(\vartheta) = \vartheta$, dann folgt $\psi'(\vartheta) = 1$ und für die Schranke gilt

$$Var_\vartheta(T(X)) \geq \frac{1}{I_n(\vartheta)}$$

(b)
Ist $X = (X_1, X_2, ..., X_n)$ mit $X_1, X_2, ..., X_n \overset{iid}{\sim} f_\vartheta^{(1)}$ eindimensionale Dichte, dann gilt (Ü)

$$I_n(\vartheta) = n \cdot I_1(\vartheta), \quad I_1(\vartheta) = \left(E\left(\frac{\partial}{\partial\vartheta} \ln f_\vartheta^{(1)}(X_1)\right)^2\right)$$

(c)

$$\frac{\psi'(\vartheta)^2}{I_n(\vartheta)} \left(\overset{b}{=} \frac{\psi'(\vartheta)^2}{nI_1(\vartheta)} \overset{n\to\infty}{\longrightarrow} 0\right) \text{ 'Cramér-Rao Schranke'}$$

(d)
Bedingung (iii)b setzt voraus, dass Differentiation und Integration vertauscht werden dürfen.
Stetiger Fall:

$$
\begin{aligned}
\frac{\partial}{\partial \vartheta} E_\vartheta(T(X)) &\overset{\text{Ann.}}{=} E_\vartheta\left(T(X)\frac{\partial}{\partial \vartheta}\ln f_\vartheta(x)\right)\\
&= \int T(X)\underbrace{\left(\frac{\partial}{\partial \vartheta}\ln f_\vartheta(x)\right)}_{\frac{1}{f_\vartheta(x)}\frac{\partial}{\partial \vartheta}f_\vartheta(x)} f_\vartheta(x)dx\\
&= \int T(x)\frac{\partial}{\partial \vartheta} f_\vartheta(x)dx\\
&\overset{\text{Ann.}}{=} \frac{\partial}{\partial \vartheta}\int T(x)f_\vartheta(x)dx\\
&= \frac{\partial}{\partial \vartheta}E_\vartheta(T(X))
\end{aligned}
$$

(e)
Existieren $\frac{\partial^2}{\partial \vartheta^2} f_\vartheta(x)$ und sind Vertauschungen erlaubt, dann gilt

$$
I_n(\vartheta) = E_\vartheta\left(\left(\frac{\partial}{\partial \vartheta}\ln f_\vartheta(x)\right)^2\right) = -E_\vartheta\left(\frac{\partial^2}{\partial \vartheta^2}\ln f_\vartheta(x)\right)
$$

(Nebenbemerkung: 2. Ableitung oft einfacher)

Definition: 5.10 effiziente Statistik

$T(X)$ heißt (unter den obigen Annahmen) effizient für $\psi(\vartheta)$, wenn

$$
Var_\vartheta(T(X)) = \frac{\psi'(\vartheta)^2}{I_n(\vartheta)}
$$

Beachte: UMVUE nicht immer effizient (unbiased starke Einschränkung),
Aus Erwartungstreue und Effizienz folgt aber UMVUE

Beispiel (5.11)
(a)

$X_1, X_2, ..., X_n \overset{iid}{\sim} \text{Bin}(1, \vartheta),\ f_\vartheta(x) = \vartheta^x(1-\vartheta)^{1-x}$ mit $x \in \{0,1\}$, $\vartheta \in (0,1)$

$$\begin{aligned}
I_1(\vartheta) &= E_\vartheta\left(\left(\frac{\partial}{\partial\vartheta}\ln f_\vartheta(x)\right)^2\right)\\
&= E_\vartheta\left(\left(\frac{\partial}{\partial\vartheta}\ln\left(\vartheta^{x_1}(1-\vartheta)^{1-x_1}\right)\right)^2\right)\\
&= E_\vartheta\left[\left(\frac{\partial}{\partial\vartheta}[x_1\ln\vartheta + (1-x_1)\ln(1-\vartheta)]\right)^2\right]\\
&= E_\vartheta\left[\left(\frac{x_1}{\vartheta} + (1-x_1)\frac{1}{1-\vartheta}(-1)\right)^2\right]\\
&= E_\vartheta\left[\left(\frac{x_1(1-\vartheta)-(1-x_1)\vartheta}{\vartheta(1-\vartheta)}\right)^2\right]\\
&= \frac{1}{\vartheta^2(1-\vartheta)^2}E_\vartheta\left[(x_1 - \cancel{\vartheta x_1} - \vartheta + \cancel{\vartheta x_1})^2\right]\\
&= \frac{E_\vartheta(x_1-\vartheta)^2}{\vartheta^2(1-\vartheta)^2}\\
&= \frac{Var_\vartheta(x_1)}{\vartheta^2(1-\vartheta)^2}\\
&= \frac{\vartheta(1-\vartheta)}{\vartheta^2(1-\vartheta)^2}\\
&= \frac{1}{\vartheta(1-\vartheta)}
\end{aligned}$$

Demnach ist

$$I_n(\vartheta) = nI_1(\vartheta) = \frac{n}{\vartheta(1-\vartheta)}$$

Für alle $T(X)$ mit $E_\vartheta T(X) = \vartheta$ gilt also:

$$Var_\vartheta(T(X)) \geq \frac{\vartheta(1-\vartheta)}{n}$$

Für $T^*(X) = \bar{X}$ ist T^* erwartungstreu ($E\bar{X} = \vartheta$) und $VarT^* = \frac{\vartheta(1-\vartheta)}{n} \leq Var_\vartheta T(X)$
Aus der Erwartungstreue und Effizienz folgt, dass $T^*(X) = \bar{X}$ UMVUE ist.
(b)
$X_1, X_2, ..., X_n \overset{iid}{\sim} N(\mu, \sigma^2)$ mit $\vartheta = \mu$ und σ^2 bekannt.

$$f_\vartheta^{(1)}(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{1}{2\sigma^2}(x-\vartheta)^2\right)$$

$$\begin{aligned}
I_1(\vartheta) &= E_\vartheta\left[\left(\frac{\partial}{\partial\vartheta}\left[\ln\frac{1}{\sqrt{2\pi\sigma^2}}-\frac{1}{2\sigma^2}(x_1-\vartheta)^2\right]\right)^2\right]\\
&= E_\vartheta\left[\left(0-\frac{1}{2\sigma^2}2(x_1-\vartheta)(-1)\right)^2\right]\\
&= E_\vartheta\left[\left(\frac{x_1-\vartheta}{\sigma^2}\right)^2\right]=\frac{1}{\sigma^4}\underbrace{VarX_1}_{\sigma^2}\\
&= \frac{1}{\sigma^2}
\end{aligned}$$

also ist

$$I_n(\vartheta) = n\cdot I_1(\vartheta)=\frac{n}{\sigma^2}$$

und demnach

$$Var_\vartheta(T(X))\geq\frac{\sigma^2}{n}$$

für alle $T(X)$ erwartungstreu mit $VarT<\infty$.
$\hat{\vartheta}=\bar{X}$ ist erwartungstreu und $Var\bar{X}=\frac{\sigma^2}{n}$ (effizient)
und daraus folgt, dass $\bar{X}$ UMVUE ist.

Bemerkung (5.12)
Sei X n-dimensionale Zufallsvariable mit Dichte f_ϑ, $\vartheta\in\Theta\subseteq\mathbb{R}^k$ offen.
Die Rao-Cramér Schranke kann analog 5.8 hergeleitet werden:
Annahmen:
(i) Der Träger ist unabhängig von ϑ
(ii) Für den 'Score-Vektor'

$$U_n(\vartheta)=\left(\frac{\partial}{\partial\vartheta_1}\ln f_\vartheta(x),...,\frac{\partial}{\partial\vartheta_k}\ln f_\vartheta(x)\right)^T$$

gilt

$$EU_n(\vartheta)=0\quad\forall\,\vartheta\in\Theta$$

(iii) $T=(T_1,...,T_l)^T$ sei $\mathbb{R}^l$-wertige Statistik mit

$$E_\vartheta(\underbrace{T(X)U_n(\vartheta)^T}_{(l\times k)\text{-Matrix}})=\left(\frac{\partial}{\partial\vartheta_1}E_\vartheta(T(X)),...,\frac{\partial}{\partial\vartheta_k}E_\vartheta(T(X))\right)=:G(\vartheta)$$

Angenommen die Fischer-Informationsmatrix

$$I_n(\vartheta)=E_\vartheta(U_nU_n^T)$$

existiert und ist positiv definit. Dann gilt

$$Cov(T(X))=E_\vartheta(T(X)T(X)^T)-E_\vartheta(T(X))E_\vartheta(T(X)^T)\geq G(\vartheta)I_n^{-1}(\vartheta)G(\vartheta)^T$$

Dabei gilt $A\geq B$ für Matrizen $A,B\Leftrightarrow A-B$ positiv semidefinit.

Beispiel (5.13)
$X_1, X_2, ..., X_n \overset{iid}{\sim} N(\mu, \sigma^2)$, $\vartheta = (\mu, \sigma^2)^T$

$$f_\vartheta^{(1)}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right)$$

$$I_n(\vartheta) \overset{iid}{=} nI_1(\vartheta) = nE_\vartheta(U_1(\vartheta)U_1(\vartheta)^T)$$

$$\ln f_\vartheta^{(1)}(x) = -\frac{1}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}(x-\mu)^2$$

$$\frac{\partial}{\partial\mu} \ln f_\vartheta^{(1)}(x) = -\frac{1}{2\sigma^2} 2(x-\mu)(-1) = \frac{x-\mu}{\sigma^2}$$

$$\begin{aligned} \frac{\partial}{\partial\sigma^2} \ln f_\vartheta^{(1)}(x) &= -\frac{1}{2\sigma^2} - \frac{(x-\mu)^2}{2}\frac{\partial}{\partial\sigma^2}(\sigma^2)^{-1} \\ &= -\frac{1}{2\sigma^2} - \frac{(x-\mu)^2}{2}(\sigma^2)^{-2}(-1) \\ &= -\frac{1}{2\sigma^2} + \frac{(x-\mu)^2}{2\sigma^4} \end{aligned}$$

$$\begin{aligned} U_1(\vartheta) &= \begin{pmatrix} \frac{\partial}{\partial\mu} \ln f_\vartheta^{(1)}(x) \\ \frac{\partial}{\partial\sigma^2} \ln f_\vartheta^{(1)}(x) \end{pmatrix} \\ &= \begin{pmatrix} \frac{x-\mu}{\sigma^2} \\ -\frac{1}{2\sigma^2} + \frac{(x-\mu)^2}{2\sigma^4} \end{pmatrix} \end{aligned}$$

$$I_n(\vartheta) = nE_\vartheta(U_1(\vartheta)U_1(\vartheta)^T) = n\begin{pmatrix} a & b \\ b & c \end{pmatrix}$$

mit

$$\begin{aligned} a &= E\left(\frac{X_1-\mu}{\sigma}\right)^2 = \frac{\sigma^2}{\sigma^4} = \frac{1}{\sigma^2} \\ b &= E\left[\left(\frac{X_1-\mu}{\sigma^2}\right)\left(-\frac{1}{2\sigma^2} + \frac{(X_1-\mu)^2}{2\sigma^4}\right)\right] \\ &= -\frac{1}{2\sigma^2}E\left(\frac{X_1-\mu}{\sigma^2}\right) + E\left(\frac{(X_1-\mu)^3}{2\sigma^6}\right) \\ &= 0 \end{aligned}$$

(ungerade Momente der $N(0, \sigma^2)$-Verteilung sind 0)

$$c = E\left[\left(-\frac{1}{2\sigma^2} + \frac{(X_1-\mu)^2}{2\sigma^4}\right)^2\right] \overset{\text{Ü}}{=} \frac{1}{2\sigma^4}$$

(Verwenden kann man dabei, dass $E(X_1 - \mu)^4 = 3\sigma^4$)

$$I_n(\vartheta) = \begin{pmatrix} \frac{n}{\sigma^2} & 0 \\ 0 & \frac{n}{2\sigma^4} \end{pmatrix}$$

Für $T = (T_1, T_2)^T$ erwartungstreu für $\vartheta = (\mu, \sigma^2)^T$ gilt

$$G(\vartheta) = \left(\frac{\partial}{\partial \mu} \underbrace{E_\vartheta(T(X))}_{=(\mu,\sigma^2)^T}, \frac{\partial}{\partial \sigma^2} E_\vartheta(T(X)) \right)$$

$$= \begin{pmatrix} \frac{\partial}{\partial \mu}\mu & \frac{\partial}{\partial \sigma^2}\mu \\ \frac{\partial}{\partial \mu}\sigma^2 & \frac{\partial}{\partial \sigma^2}\sigma^2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Also (5.12):

$$Cov(T) \geq I_n^{-1} = \begin{pmatrix} \frac{\sigma^2}{n} & 0 \\ 0 & \frac{2\sigma^4}{n} \end{pmatrix}$$

Der Schätzer $T^* = (\bar{X}, \hat{S}^2)$ ist erwartungstreu und UMVUE aber nicht effizient, da (Ü)

$$CovT^* = \begin{pmatrix} \frac{\sigma^2}{n} & 0 \\ 0 & \frac{2\sigma^4}{n\cdot(-1)} \end{pmatrix}$$

Kapitel 6

Momentmethode und Maximum-Likelihood Schätzung

Bemerkung und Definition (Momentenmethode)
Sei $X_1, X_2, ..., X_n \overset{iid}{\sim} P_\vartheta$, $\vartheta \in \Theta$ (Stichprobe)
Annahme: $E_\vartheta X_1, E_\vartheta X_1^2, ..., E_\vartheta X_1^k$ existieren, $m_j(\vartheta) := E_\vartheta X_1^j$
Wir wollen $\psi(\vartheta) = h(m_1(\vartheta), ..., m_k(\vartheta))$ schätzen.
Momentenmethode:

$$\hat{\psi}(\vartheta) = h(\hat{m}_1, ..., \hat{m}_k) \text{ (Substitution/ plug-in)}$$

mit

$$\hat{m}_j = \frac{1}{n} \sum_{i=1}^{n} X_i^j \text{ (Stichprobenmomente)}$$

(erwartungstreu, konsistent, asmyptotisch normal)
Für h stetig gilt (nach dem Continuous Mapping Theorem), dass $\hat{\psi}(\vartheta)$ konsistent ist, zusätzlich ist $\hat{\psi}$ asymptotisch normal, falls die Voraussetzungen für die multivariate Delta-Methode erfüllt sind.

Beispiel (6.2)
(a)
$X_1, X_2, ..., X_n \overset{iid}{\sim} P_\vartheta$ schätzen
$\psi(\vartheta) = Var_\vartheta(X_1) = E_\vartheta X_1^2 - (E_\vartheta X_1)^2 = m_2(\vartheta) - m_1(\vartheta)^2$
Method of Moments (MoM) Schätzer:

$$\hat{\sigma}^2 = \frac{1}{n} \sum X_i^2 - \left(\frac{1}{n} \sum X_i \right)^2 = \frac{1}{n} \sum (X_i - \bar{X})^2 = \frac{n-1}{n} \hat{S}_n^2$$

(b)
Beispiel (3.5), $X_1, X_2, ..., X_n \overset{iid}{\sim} U(0, \vartheta)$, $E_\vartheta X_1 = \frac{\vartheta}{2}$

$$\psi(\vartheta) = \vartheta = 2 \cdot \frac{\vartheta}{2} = 2E_\vartheta X_1 = 2m_1(\vartheta)$$

$$\Rightarrow \hat{\vartheta} = 2\bar{X}$$

(c)
$X_1, X_2, ..., X_n \overset{iid}{\sim} \text{Gam}(\alpha, \beta)$ mit $f_\vartheta = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$, $x > 0$
$E_\vartheta X_1 = \alpha\beta = m_1(\vartheta)$
$E_\vartheta X_1^2 = Var_\vartheta X_1 + (E_\vartheta X_1)^2 = \alpha\beta^2 + (\alpha\beta)^2 = m_2(\vartheta)$

$$\beta = \frac{\alpha\beta^2}{\alpha\beta} = \frac{Var_\vartheta(X_1)}{E_\vartheta X_1} = \frac{m_2(\vartheta) - (m_1(\vartheta))^2}{m_1(\vartheta)}$$

$$\alpha = \frac{(\alpha\beta)^2}{\alpha\beta^2} = \frac{(E_\vartheta X_1)^2}{Var_\vartheta X_1} = \frac{m_1(\vartheta)^2}{m_2(\vartheta) - (m_1(\vartheta))^2}$$

$$\Rightarrow \hat{\beta} = \frac{\hat{\sigma}^2}{\bar{X}}, \quad \hat{\alpha} = \frac{\bar{X}^2}{\hat{\sigma}^2} \quad (\hat{\sigma}^2 \overset{(a)}{=} \frac{1}{n} \sum (X_i - \bar{X})^2)$$

Beispiel (6.3 Zur Maximum-Likelihood (ML) Schätzung)
Zu Schätzen: Anzahl $N(= \vartheta)$ der Fische in einem Teich mittels capture/ recapture sampling:
Fange und markiere K Fische, Zeit später fange n Fische, von denen sind x markiert.
$(\Rightarrow \frac{x}{n} \approx \frac{K}{N} \Rightarrow N \approx \frac{K \cdot n}{x})$
Sei X Zufallsvariable Anzahl der markierten Fische (von n)

$$P_N(X = j) = \frac{\binom{K}{j}\binom{N-K}{n-j}}{\binom{N}{n}} \quad N \geq \max(K, n) =: N^*$$

Angenommen $X(w) = x$
<u>Idee</u>: Wähle $N(\geq N^*)$ für das $P_N(X = x)$ maximal ist.
Es gilt:

$$\frac{P_N(X = x)}{P_{N-1}(X = x)} = \frac{\cancel{\binom{K}{x}}\binom{N-K}{n-x}}{\binom{N}{n}} \cdot \frac{\binom{N-1}{n}}{\cancel{\binom{K}{x}}\binom{N-1-K}{n-x}}$$

Nebenrechnung:

$$\frac{\binom{N-1}{n}}{\binom{N}{n}} = \frac{(N-1)!}{\cancel{n!}(N-1-n)!} \cdot \frac{\cancel{n!}(N-n)!}{N!} = \frac{N-n}{N}$$

Nebenrechnung vorbei.
Nach der Nebenrechnung gilt

$$\frac{P_N(X = x)}{P_{N-1}(X = x)} = \frac{(N-n)(N-K)}{N(N-K-(n-x))}$$

Für das größte N mit $P_N(X = x) \geq P_{N-1}(X = x)$ gilt

$$(N-n)(N-K) \geq N(N-K-n-x) \Leftrightarrow nK \geq Nx \Leftrightarrow N \leq \frac{n \cdot K}{x}$$

daraus folgt

$$\Rightarrow \hat{N}(x) = \max\left\{i \in \mathbb{N} : i \leq \frac{nK}{x}\right\} = \left\lfloor \frac{nK}{x} \right\rfloor$$

$\hat{N}(x)$ heißt ML-Schätzer für N.

Definition: 6.4

Sei X Zufallsvariable mit (diskreter oder stetiger) Dichte f_ϑ, $\vartheta \in \Theta$. Man nennt

$$L : \Theta \to \mathbb{R} : \vartheta \to L(\vartheta, x) = f_\vartheta(x) \quad (x \text{ fest}) \ \underline{\text{Likelihood-Funktion}}$$

$$l = \ln L \ \underline{\text{log-Likelihood-Funktion}}$$

Ein Schätzer $\hat{\vartheta} : \mathcal{X} \to \Theta$ heißt Maximimum-Likelihood Schätzer, falls gilt

$$\forall\, x \in \mathcal{X} : \sup_{\vartheta \in \Theta} L(\vartheta, x) = L(\hat{\vartheta}, x)$$

d.h. $\hat{\vartheta} = \arg\sup_{\vartheta\in\Theta} f_\vartheta(x)$

Bemerkung (6.5)
(a)
Für X diskret: wähle das 'wahrscheinlichste' ϑ, also $\hat{\vartheta}$ mit $P_{\hat{\vartheta}}(X = x) \geq P_\vartheta(X = x)$
(b)
Oft einfacher: maximiere $l = \ln L$ statt L (geht, da ln streng monoton)
(c)
Für $X_1, X_2, ..., X_n \overset{iid}{\sim} f_\vartheta^{(1)}$ gilt:

$$L(\vartheta, x_1, ..., x_n) = \prod_{i=1}^{n} f_\vartheta^{(1)}(x_i)$$
$$l(\vartheta, x_1, ..., x_n) = \sum_{i=1}^{n} \ln f_\vartheta^{(1)}(x_i)$$

(d)
Ist $f_\vartheta(x), \vartheta \mapsto f_\vartheta(x)$ differenzierbar, dann erfüllt der ML Schätzer die (log-)Likelihood Gleichungen:

$$\frac{\partial}{\partial \vartheta_i} l(\vartheta, x_1, ..., x_n)\Big|_{\vartheta=\hat{\vartheta}} = 0 \quad \forall\, j = 1, ..., k \ \ \forall\, x \in \mathcal{X}$$

$(\vartheta = (\vartheta_1, ..., \vartheta_k) \in \Theta \subseteq \mathbb{R}^k)$

Beispiel (6.6)
(a)
$X_1, X_2, ..., X_n \overset{iid}{\sim} N(\mu, \sigma^2)$, $\vartheta := (\mu, \sigma^2)^T \in \Theta := \mathbb{R} \times (0, \infty)$, $\nu := \sigma^2$.

$$L_x(\vartheta) := L(\vartheta, x) = \prod_{j=1}^{n} f_j(X_j|\vartheta)$$
$$l_x(\vartheta) := l(\vartheta, x) = \sum_{j=1}^{n} -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\nu) - \frac{1}{2}\nu^{-1}\sum_{j=1}^{n}(x_j - \mu)^2$$

$$\frac{\partial l_x(\vartheta)}{\partial \mu} = \nu^{-1}\sum_{j=1}^{n}(X_j-\mu) \stackrel{!}{=} 0 \Leftrightarrow \hat{\mu} = \bar{X}_n$$
$$\frac{\partial l_x(\vartheta)}{\partial \nu} = -\frac{n}{2}\nu^{-1} + \frac{1}{2}\nu^{-2}\sum_{j=1}^{n}(X_j-\mu)^2 \stackrel{!}{=} 0 \Rightarrow \hat{\nu} = \frac{1}{n}\sum_{j=1}^{n}(X_j-\mu)^2$$

Zudem:

$$\lim_{|\mu|\to\infty} L_x(\vartheta) = 0 = \lim_{\nu\to\infty} L_x(\vartheta),\ \text{also } \hat{\vartheta}_{ML}(X) = \left(\bar{X}_n, \frac{1}{n}\sum_{j=1}^{n}(X_j-\bar{X}_n)^2\right)^T = (\hat{\mu},\hat{\nu})^T$$

d.h. der MLE (Maximum Likelihood Estimator) stimmt mit dem MomE (Method of moments estimator) überein,
$\hat{\vartheta} = (\hat{\mu},\hat{\sigma})$ ist konsistent, $\hat{\mu}$ erwartungstreu, $\hat{\nu}$ asymptotisch unerwartungstreu
Beachte: μ bekannt $\Rightarrow \tilde{\nu} = \frac{1}{n}\sum_{j=1}^{n}(X_j-\mu)^2$, ν bekannt $\Rightarrow \tilde{\mu} = \hat{\mu} \sim N(\mu, \frac{\nu}{n})$ (beide erwartungstreu)

$$\frac{n\tilde{\nu}}{\nu} = \sum_{i=1}^{n}\Big(\underbrace{\frac{X_i-\mu}{\sigma}}_{\sim N(0,1)}\Big)^2 \sim \chi^2_n \ \text{ aber } \ \frac{n\hat{\nu}}{\nu} = \frac{(n-1)\hat{S}_n^2}{\sigma^2} \sim \chi^2_{n-1}$$

(b)
$X_1, X_2, ..., X_n \stackrel{iid}{\sim} U(0,\vartheta)$, $\Theta = (0,\infty)$, $y \in (0,\infty)^n$,

$$L_y(\vartheta) = \vartheta^{-n}1_{(0,\vartheta)}(\max\{y_1,...,y_n\}) = \vartheta^{-n}1(0 < Y_{(n)} < \vartheta)$$

ϑ^{-n} fallend in ϑ, $Y_{(n)} < \vartheta$ wähle also das kleinste $\vartheta > Y_{(n)}$

$$\Rightarrow \hat{\vartheta}_{ML}(X) = \max\{X_1, X_2, ..., X_n\} = X_{(n)} \neq MomE$$

Nebenbemerkung:

$$L(\vartheta, x) := f_\vartheta(x) = \begin{cases} P_\vartheta[\{x\}] & \text{diskret} \\ \frac{dP_\vartheta}{d\gamma}(x) & \text{stetig} \end{cases}$$

Nebenbemerkung vorbei.
Hier und im Allgemeinen gilt ML-Schätzer und MoM-Schätzer stimmen nicht überein.

Theorem: 6.7

Sei T suffiziente Statistik für ϑ und $\hat{\vartheta}$ MLE für $\vartheta \Rightarrow \hat{\vartheta}$ ist Funktion von T; $\xi \circ T$

Theorem: 6.8

Die Voraussetzung von (5.8) (Cramér-Rao bound) seien erfüllt.
Sei $T(X)$ erwartungstreu und effizient für ϑ ($\psi = \mathrm{id}_\vartheta$) d.h. $Var(T(X)) = I_n(\vartheta)$
Dann hat die Likelihood-Gleichung die eindeutige Lösung

$$\hat{\vartheta}(x) = T(X),\ \text{d.h. } \vartheta_{ML} = T(X)$$

Bemerkung (6.9)
$\hat{\vartheta}$ erwartungstreu, effizient und Anmerkung (5.8) sei erfüllt. Dann folgt

$$\hat{\vartheta} \text{ ist MLE}$$

Aber: nicht jeder MLE ist erwartungstreu oder effizient.

Bemerkung (6.10)
Ist $\hat{\vartheta}$ MLE für ϑ, dann ist $\hat{\gamma} = \gamma(\hat{\vartheta})$ (plug-in) MLE für $\gamma = \gamma(\vartheta)$.
Rechtfertigung:

Theorem: 6.11

Es seien $\Theta \subseteq \mathbb{R}^k$, $\psi : \Theta \to \Gamma = \psi(\Theta) \subseteq \mathbb{R}^p$, $1 \leq p \leq k$, $\{f_\vartheta : \vartheta \in \Theta\}$ Familie von diskreten oder stetigen Dichten und $\hat{\vartheta}$ MLE für ϑ.
Dann maximiert $\hat{\gamma} = \psi(\hat{\vartheta})$ die von γ induzierte Likelihood $M(\gamma, x) = \sup\limits_{\vartheta \in \Theta_\gamma} L(\vartheta, x)$, wobei

$\Theta_\gamma := \{\vartheta \in \Theta : \psi(\vartheta) = \gamma\} \overset{(**)}{\subseteq} \Theta$.

Beispiel (6.12)
$X_1, X_2, ..., X_n \overset{iid}{\sim} \text{Bin}(1, p)$, $\Theta := (0,1)$, $\psi := V_p X_1 = p(1-p)$

$$\hat{p}_{ML} = \bar{X}_n \text{ ist MLE (Ü)} \Rightarrow \psi(\hat{p}) = \bar{X}_n(1 - \bar{X}_n) \text{ ist MLE für } \psi(p)$$

Bemerkung (6.13)
MLEs sind (unter Regularitätsvoraussetzungen) asymptotisch normal; $\sqrt{n}(\hat{\vartheta} - \vartheta) \xrightarrow{\text{D}} N_k(0, I_1(\vartheta)^{-1})$ wobei $I_1(\vartheta)$ die Fischer-Matrix mit Einträgen $E_\vartheta(\frac{\partial}{\partial \vartheta_j} \ln f_\vartheta^{(1)}(X_1) \frac{\partial}{\partial \vartheta_k} \ln_\vartheta^{(1)}(X_1))$ ist.

Kapitel 7

Lineare Regression

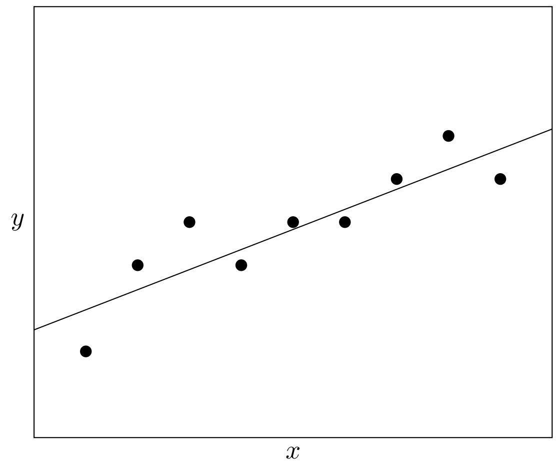
Wozu? Modellierung der linearen Abhängigkeit zwischen einer Zufallsvariablen X und einer Zufallsvariablen Y.

Beispiel
- Länge einer Feder Y hängt ab ab von Belastung X
- Blutdruck Y hängt ab von Alter X

Beobachtungen: (x_j, y_j), $j = 1, ..., n$,
x_j 'Designpunkte' (deterministisch),
y_j 'Beobachtungen' (zufällig)

Lineares Regressionsmodell

$y_j = \underline{a}x_j + \underline{b} + \varepsilon_j \rightarrow$ Messfehler (zufällig)
Annahmen: $\varepsilon_1, ..., \varepsilon_n$ unabhängig, $E\varepsilon_j = 0$, $Var\varepsilon_j = \sigma^2$, $j = 1, ..., n$
Ziel: Schätze die Regressions-Koeffizienten a, b und erkläre damit den Zusammenhang zwischen X und Y.



<u>Ansätze für Schätzer</u>: Beschränkung auf lineare Schätzer, d.h.

$$\hat{a} = \sum_{i=1}^{n} a_i y_i, \quad \hat{b} = \sum_{i=1}^{n} b_i y_i$$

Suchen Koeffizienten a_i, b_i, $i = 1, ..., n$ die von den Designpunkten abhängen.
<u>Erwartungstreue</u>

$$\begin{aligned} E\hat{a} = a \Leftrightarrow E\left(\sum_{i=1}^{n} a_i y_i\right) &= \sum_{i=1}^{n} a_i E(y_i) \\ &= \sum_{i=1}^{n} a_i(ax_i + b) = a \underbrace{\sum_{i=1}^{n} a_i x_i}_{\stackrel{!}{=}1} + b \underbrace{\sum_{i=1}^{n} a_i}_{\stackrel{!}{=}0} = a \end{aligned}$$

Analog

$$E\hat{b} = a \underbrace{\sum_{i=1}^{n} b_i x_i}_{\stackrel{!}{=}0} + b \underbrace{\sum_{i=1}^{n} b_i}_{\stackrel{!}{=}1} = b$$

<u>Fordern</u>:

$$\begin{aligned} &\sum_{i=1}^{n} a_i x_i = 1, \quad \sum_{i=1}^{n} a_i = 0 \\ &\sum_{i=1}^{n} b_i x_i = 0 \quad \sum_{i=1}^{n} b_i = 1 \end{aligned}$$

<u>In Matrixschreibweise</u>:

$$C = \begin{pmatrix} a_1 & ... & a_n \\ b_1 & ... & b_n \end{pmatrix} \quad X = \begin{pmatrix} x_1 & ... & x_n \\ 1 & ... & 1 \end{pmatrix}$$

$$CX^T = \begin{pmatrix} a_1 & ... & a_n \\ b_1 & ... & b_n \end{pmatrix} \begin{pmatrix} x_1 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^{n} a_j x_j & \sum_{j=1}^{n} a_j \\ \sum_{j=1}^{n} b_j x_j & \sum_{j=1}^{n} b_j \end{pmatrix} \stackrel{!}{=} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \Leftrightarrow \underbrace{CX^T = I}_{:\Leftrightarrow (F1)}$$

<u>Minimale Varianz</u>

$$\begin{aligned} Var(\hat{a}) = Var\left(\sum_{i=1}^{n} a_i y_i\right) &= \sum_{i=1}^{n} a_i^2 Var(y_i) = \sigma^2 \sum_{i=1}^{n} a_i^2 \\ &= \sigma^2 \cdot \left\| \begin{pmatrix} 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} a_1 & ... & a_n \\ b_1 & ... & b_n \end{pmatrix} \right\|_2^2 = \sigma^2 ||e_1^T C||_2^2 \end{aligned}$$

Analog

$$Var(\hat{b}) = \sigma^2 ||e_2^T C||_2^2$$

<u>Ansatz für C</u>: $C = DX,\ D \in \mathbb{R}^{2\times 2}$ (F2)
Angenommen wir haben Matrix C, die $(F1)$ erfüllt und $\tilde{C}$ sei eine beliebige weitere solche Matrix. Wann gilt

$$\sigma^2 ||e_j^T C||_2^2 \leq \sigma^2 ||e_j^T \tilde{C}||_2^2 \text{ ?} \quad \text{(min. Varianz)}$$

Dann folgt

$$\tilde{C}X^T - CX^T = (\underbrace{\tilde{C} - C}_{=:\Delta})X^T = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

und damit

$$\Delta C^T = \Delta X^T D^T = 0 D^T = 0$$

$$\Rightarrow ||e_j^T \tilde{C}||_2^2 = ||e_j^T C + e_j^T \Delta||_2^2 = ||e_j^T C||_2^2 + \underbrace{||e_j^T \Delta||_2^2}_{\geq 0} + \underbrace{e_j^T \Delta C^T e_j}_{=0} + \underbrace{e_j^T C \Delta^T e_j}_{=0} \geq ||e_j^T C||_2^2$$

also minimiert C die Varianz, d.h. wähle die Einträge von C für die Koeffizienten der Schätzer.
<u>Zu zeigen</u>: Es existiert auch wirklich eine Matrix C mit $CX^T = I$ und $C = DX$.

$$\Rightarrow DXX^T = I \Rightarrow D = (XX^T)^{-1} \text{ falls } XX^T \text{ invertierbar ist.}$$

$$\begin{aligned} XX^T &= \begin{pmatrix} x_1 & \dots & x_n \\ 1 & \dots & 1 \end{pmatrix} \begin{pmatrix} x_1 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{pmatrix} = \begin{pmatrix} \sum x_i^2 & \sum x_i \\ \sum x_i & n \end{pmatrix} \\ &= n \begin{pmatrix} \bar{x}^2 & \bar{x} \\ \bar{x} & 1 \end{pmatrix} \end{aligned}$$

$$\det(XX^T) = n^2 \left(\bar{x}^2 - (\bar{x})^2\right)$$

XX^T invertierbar $\Leftrightarrow \bar{x}^2 \neq (\bar{x})^2$. Wann ist $\bar{x}^2 = (\bar{x})^2$?
Betrachte Zufallsvariable $Z \sim U(\{x_1, ..., x_n\})$.

$$EZ = \sum \frac{1}{n} x_i = \bar{x}$$

$$VarZ = \bar{x}^2 - (\bar{x})^2$$

Wann gilt $VarZ = 0$? Falls $Z = EZ$ P-fast sicher $\Leftrightarrow x_1 = ... = x_n$
$\Rightarrow$ Sofern $x_i \neq x_j$ für ein Paar $(i,j) \in \{1, ..., n\}^2$ (F3) ist XX^T invertierbar.
Damit ergibt sich

$$D = (XX^T)^{-1}, \quad C = DX = (XX^T)^{-1}X \text{ mit } \underline{y} = (y_1, ..., y_n)^T \text{ und}$$

$$\begin{pmatrix} \hat{a} \\ \hat{b} \end{pmatrix} = \begin{pmatrix} \underline{a}^T & \underline{y} \\ \underline{b}^T & \underline{y} \end{pmatrix} = Cy$$

Theorem

Wir betrachten das lineare Regressionsmodell mit der Zusatzvoraussetzung, dass nicht alle x_i $(i = 1, ..., n)$ übereinstimmen. Dann hat der Regressionsschätzer

$$\begin{pmatrix} \hat{a} \\ \hat{b} \end{pmatrix} = (XX^T)^{-1}X\underline{y}$$

für Koeffizienten a und b unter allen linearen erwartungstreuen Schätzern die kleinste Varianz. ('BLUE' Best Linear Unbiased Estimator)

Bemerkung
Die Gerade $\hat{a}x + \hat{b}$ ist die Ausgleichsgerade aus der Numerik, die man mit der Methode der kleinsten Quadrate erhält.
Minimierung mit Euklidischen Abstand $\triangleq$ Minimierung mit der Varianz.

Kapitel 8

Hypothesentests

8.1 Grundlagen und Beispiele

Beispiel (8.1)
$X : \Omega \to \mathcal{X}$ Zufallsvariable, $X \sim \{P_\vartheta : \vartheta \in \Theta\}$, $\Theta = \Theta_0 \cup \Theta_1$ disjunkt, nicht-leer ($\Theta_0 \dot\cup \Theta_1, \Theta_0 + \Theta_1$)
Gegeben: $X(w) = x$
Entscheide ob $H_0 : \vartheta \in \Theta_0$ oder $H_1 : \vartheta \in \Theta_1$?
Bsp: 2 Medikamente M_1 (etabliert), M_2 (neu)
Heilungswahrscheinlichkeit von $M_1 : \theta = 0.8$ bekannt, von $M_2 : \vartheta$ unbekannt
$H_0 : \vartheta \leq 0.8$, also $\vartheta \in [0, 0.8]$ (M_2 nicht besser als M_1)
$H_1 : \vartheta \in \Theta_1 = (0.8, 1]$ (M_2 besser als M_1)
Teste M_2 an 50 Personen: $X_1, ..., X_{50} \overset{\text{iid}}{\sim} Bin(1, \vartheta)$, '$X_i = 1$'$\triangleq$ geheilt
Angenommen $\sum_{i=1}^n X_i = 42$, $\hat\vartheta = \frac{\sum X_i}{n} = \frac{42}{50} = 0.84$
Entscheidung?

Definition: 8.2

Sei X Zufallsvariable, $\vartheta \in \Theta = \Theta_0 \dot\cup \Theta_1$.
Ein Hypothesentest ist eine (messbare) Abbildung

$$\varphi : \mathcal{X} \to [0, 1] : x \mapsto \varphi(x) = P(\text{'Entscheidung für } H_1\text{'}) \text{ falls } x = X(w) \text{ beobachtet wurde}$$

φ heißt nicht-randomisiert, falls $\varphi(\mathcal{X}) \subseteq \{0, 1\}$ (im Allgemeinen $\varphi(\mathcal{X}) = \{0, 1\}$) ($\triangleq$ üblicher Fall)

Bemerkung (8.3 Übliche Sprechweisen)
Entscheidung für H_1: 'H_0 wird verworfen'
Entscheidung für H_0: 'H_0 wird beibehalten/ nicht verworfen' (nicht genug Anhaltspunkte, dass H_1 gilt)

Mögliche Fehler

Entscheidung für	$\vartheta \in \Theta_0$	$\vartheta \in \Theta_1$
H_0	✓	Fehler 2. Art
H_1	Fehler 1. Art	✓

Kontrolliert wird die Wahrscheinlichkeit für den Fehler 1. Art ($\leq \alpha$ im Allgemeinen $\alpha \in \{0.01, 0.05, 0.1\}$)

Definition: 8.4

Sei $\varphi : \mathcal{X} \to [0,1]$ ein Test mit $H_0 : \vartheta \in \Theta_0$ vs. $H_1 : \vartheta \in \Theta_1$. Die Funktion

$$\beta_\varphi : \Theta \to [0,1] : \vartheta \mapsto \beta_\varphi(\vartheta) = E_\vartheta(\varphi(x))$$

heißt Güte-Funktion von φ.
Sei $\alpha \in [0,1]$, dann heißt φ Test zum Niveau α (Level α, Signifikanzniveau α), falls

$$\beta_\varphi(\vartheta) \leq \alpha \quad \forall\, \vartheta \in \Theta_0$$

Für $\vartheta \in \Theta_1$ heißt $\beta_\varphi(\vartheta)$ Power/ Güte.
Φ_α sei die Menge aller Tests zum Niveau α.

$$\sup\left\{\beta_\varphi(\vartheta) : \vartheta \in \Theta_0\right\}$$

heißt effektives Niveau von φ ('size')

Bemerkung (8.5)
$\beta_\varphi(\vartheta) =$ Wahrscheinlichkeit Ablehnung von H_0 falls ϑ wahr.
φ nicht randomisiert: $\beta_\varphi(\vartheta) = 1 \cdot P_\vartheta(\varphi(x) = 1) + 0$
φ randomisiert: Für X diskret (analog stetig) gilt

$$\begin{aligned}
P_\vartheta(\varphi(x) = 1) &= \sum_x P_\vartheta(\varphi(x) = 1, X = x) \\
&= \sum_x \underbrace{P_\vartheta(\varphi(x) = 1 | X = x)}_{\varphi(x)} P_\vartheta(X = x) \\
&= E_\vartheta(\varphi(x))
\end{aligned}$$

$\vartheta \in \Theta_0$: $\beta_\varphi(\vartheta)$ Wahrscheinlichkeit Fehler 1. Art
$\vartheta \in \Theta_1$: $\beta_\varphi(\vartheta)$, $1 - \beta_\varphi(\vartheta)$ Wahrscheinlichkeit Fehler 2. Art

Beispiel (8.6 Einseitiger Binomialtest, vgl. Beispiel 8.1)
$X_1, X_2, ..., X_n \overset{iid}{\sim} Bin(1, \vartheta)$, sei $\vartheta_0 \in (0,1) = \Theta$ konstant, $H_0 : \vartheta \leq \vartheta_0$, $H_1 : \vartheta > \vartheta_0$
Ansatz: Lehne H_0 ab, falls $\hat{\vartheta} = \bar{X}$ bzw. $\sum_{i=1}^n X_i$ 'groß', d.h.

$$\varphi(x) = \begin{cases} 1 & \sum X_i > k \\ 0 & \text{sonst} \end{cases} \qquad \text{(nicht randomisiert)}$$

Bestimmung von k, sodass $\varphi \in \Phi_\alpha$:

$$\begin{aligned}\beta_\varphi(\vartheta) &= P_\vartheta(\varphi(x)=1)\\ &= \underbrace{P_\vartheta\Bigg(\overbrace{\sum_{i=1}^n X_i}^{\sim Bin(n,\vartheta)} > k\Bigg)}_{\text{mon. wachs.}}\\ &= \sum_{j=k+1}^n \underbrace{\binom{n}{j}\vartheta^j(1-\vartheta)^{n-j}}_{=:b(j,n,\vartheta)}\end{aligned}$$

$\beta_\varphi(\vartheta)$ monoton wachsend in $\vartheta \Rightarrow \quad \underbrace{\beta_\varphi(\vartheta)}_{\text{Fehlerw. Fehler 1. Art}} \leq \beta_\varphi(\vartheta_0)\ \forall\ \vartheta \leq \vartheta_0$

Wähle k, sodass

$$\beta_\varphi(\vartheta_0) = \sum_{j=k+1}^n b(j,n,\vartheta_0) \leq \alpha \quad \text{Niveau } \alpha$$

und $\forall\ \vartheta \geq \vartheta_0$:

$$1-\beta_\varphi(\vartheta_0) = \sum_{j=0}^k b(j,n,\vartheta) \text{ mit } k \text{ möglichst klein}$$

d.h. $k := \min\left\{l \in \{-1,0,...,n\} : \sum_{j=l+1}^n b(l,n,\vartheta_0) \leq \alpha\right\}$
Im Beispiel 8.1:
$\vartheta_0 = 0.8$, $n=50$, $\sum_{i=1}^{50} X_i = 42$, $\alpha = 0.05$

$$\sum_{j=l+1}^n b(j,50,0.8) = \begin{cases} 0.1034 > \alpha & l=43\\ 0.048 < \alpha & l=44\end{cases} \Rightarrow k=44$$

Da $\sum X_i = 42 \not> 44$ wird H_0 beibehalten.
Hier

$$\sup_{\vartheta\in\Theta} \beta_\varphi(\vartheta) = \beta_\varphi(\vartheta_0) = 0.048 < \alpha \text{ effektives Niveau}$$

d.h. Niveau wird nicht ausgeschöpft.
Ausweg: Randomisieren
Betrachte $H_0' : \vartheta = \vartheta_0$, $H_1 : \vartheta > \vartheta_0$

$$\tilde\varphi(x) = \begin{cases} 1 & \sum X_i > k\\ \gamma & \sum X_i = k\\ 0 & \sum X_i < k\end{cases}$$

mit

$$\gamma = \frac{\alpha - P_{\vartheta_0}(\sum_{i=1}^n X_i > k)}{P_{\vartheta_0}(\sum X_i = k)}$$

$\tilde{\varphi}(x)$ ist randomisierter Test mit

$$\beta_{\tilde{\varphi}}(\vartheta_0) = E_{\vartheta_0}(\tilde{\varphi}(x)) = 1 \cdot P_{\vartheta_0}(\sum X_i > k) + \underbrace{\gamma \cdot P_{\vartheta_0}(\sum X_i = k)}_{\alpha - P_{\vartheta_0}(\sum X_i > k)} + 0 = \alpha$$

In der Programmiersprache R:
$b(44, 50, 0.8)$ =dbinom$(44, 50, 0.8)$ (dbinom Dichte)
$\sum_{j=45}^{50} b(j, 50, 0.8)$ = 1-pbinom$(44, 50, 0.8)$ (pbinom Verteilungsfunktion)

Beispiel (8.7 Einseitiger Gaußtest)
$X_1, X_2, ..., X_n \overset{iid}{\sim} N(\mu, \sigma^2)$, $\vartheta = \mu$, σ^2 bekannt, μ_0 fest mit $\mu \leq \mu_0$
a) H_0: $\mu \leq \mu_0$ vs. $H_1 : \mu > \mu_0$
$\Theta_0 = (-\infty, \mu_0]$, $\Theta_1 = (\mu_0, \infty)$
Ansatz: verwende $\bar{X}$ (effizient und erwartungstreu)
Lehne H_0 ab für $\bar{X}$ groß, $\varphi(x) = 1_{(\bar{X} > c)}$
Bestimmung von c, sodass $\varphi \in \Phi_\alpha$:

$$\begin{aligned}\beta_\varphi(\mu) = \underbrace{P_\mu(\bar{X} > c)}_{\text{mon. wachs. in } \mu} &= P_\mu\left(\underbrace{\frac{\sqrt{n}(\bar{X} - \mu)}{\sigma}}_{\sim N(0,1)} > \frac{\sqrt{n}(c - \mu)}{\sigma}\right) \\ &= 1 - \Phi\left(\frac{\sqrt{n}(c - \mu)}{\sigma}\right) \overset{\mu \leq \mu_0}{\leq} 1 - \Phi\left(\frac{\sqrt{n}(c - \mu_0)}{\sigma}\right) \overset{!}{=} \alpha\end{aligned}$$

Auflösen nach c:

$$\begin{aligned}1 - \alpha = \Phi\left(\frac{\sqrt{n}(c - \mu_0)}{\sigma}\right) &\Leftrightarrow \Phi^{-1}(1 - \alpha) = \frac{\sqrt{n}(c - \mu_0)}{\sigma} \\ &\Leftrightarrow \Phi^{-1}(1 - \alpha)\frac{\sigma}{\sqrt{n}} = c - \mu_0\end{aligned}$$

$$\Rightarrow \Phi(x) = 1(\bar{X} > c) = 1(\bar{X} > \mu_0 + \underbrace{\Phi^{-1}(1 - \alpha)\frac{\sigma}{\sqrt{n}}}_{(i)})$$

(i) ist $U_{1-\alpha}$ das $(1 - \alpha)$-Quantil der $N(0, 1)$-Verteilung, $P(N(0, 1) \leq U_{1-\alpha}) = \Phi(U_{1-\alpha}) = 1 - \alpha$

(b)
$H_0 : \mu \geq \mu_0$, $H_1 : \mu < \mu_0$ analog zu (a)

$$\begin{aligned}\varphi(x) &= 1\left(\bar{X} < \mu_0 - \frac{\sigma}{\sqrt{n}} U_{1-\alpha}\right) \\ &= 1\left(\frac{\sqrt{n}(\bar{X} - \mu_0)}{\sigma} < \underbrace{-U_{1-\alpha}}_{=U_\alpha}\right)\end{aligned}$$

(c)
$H_0 : \mu = \mu_0$, $H_1 : \mu \neq \mu_0$ (zweiseitiger Gaußtest zum Niveau α)

$$\varphi(x) = 1\left(\frac{\sqrt{n}|\bar{X} - \mu_0|}{\sigma} > U_{1-\alpha/2}\right)$$

Beispiel (8.8 t-Test)
$X_1, X_2, ..., X_n \overset{iid}{\sim} N(\mu, \sigma^2)$, $\vartheta = (\mu, \sigma^2)$, μ_0 fest
(a)
$H_0 : \mu \leq \mu_0$, $H_1 : \mu > \mu_0$
Analog zu Beispiel 8.7:

$$\begin{aligned}\varphi(x) &= 1\left\{\bar{X} > \mu_0 + \frac{\sqrt{\hat{S}^2}}{\sqrt{n}} t_{n-1,1-\alpha}\right\} \\ &= 1\Bigg\{\underbrace{\sqrt{n}\frac{\bar{X}-\mu_0}{\sqrt{\hat{S}^2}}}_{\overset{\mu=\mu_0}{\sim} t_{n-1}} > \underbrace{t_{n-1,1-\alpha}}_{(1-\alpha)-\text{Quantil}}\Bigg\}\end{aligned}$$

$\varphi \in \Phi_\alpha$:

$$\begin{aligned}\beta_\varphi(\vartheta) &= P_\vartheta\left(\frac{\sqrt{n}(\bar{X} - \mu_0 + \mu - \mu)}{\sqrt{\hat{S}_n^2}} > t_{n-1,1-\alpha}\right) \\ &= P_\vartheta\left(\frac{\sqrt{n}(\bar{X}-\mu)}{\sqrt{\hat{S}^2}} > t_{n-1,1-\alpha} + \frac{\sqrt{n}(\mu_0 - \mu)}{\sqrt{\hat{S}_n^2}}\right) \\ &\leq P_\vartheta\Bigg(\underbrace{\frac{\sqrt{n}(\bar{X}-\mu)}{\sqrt{\hat{S}_n^2}}}_{\sim t_{n-1}} > t_{n-1,1-\alpha}\Bigg) \text{ für } \mu \leq \mu_0 \\ &= 1 - F_{t_{n-1}}(t_{n-1,1-\alpha}) = 1 - (1-\alpha) = \alpha\end{aligned}$$

(b)
$H_0 : \mu \geq \mu_0$, $H_1 : \mu < \mu_0$

$$\varphi(x) = 1\left\{\frac{\sqrt{n}(\bar{X}-\mu_0)}{\sqrt{\hat{S}_n^2}} < -t_{n-1,1-\alpha}\right\}$$

(c)
$H_0 : \mu = \mu_0$, $H_1 : \mu \neq \mu_0$

$$\varphi(x) = 1\left\{|\bar{X} - \mu_0| > \frac{\sqrt{\hat{S}_n^2}}{\sqrt{n}} t_{n-1,1-\alpha/2}\right\}$$

Beispiel (8.9 χ^2-Test/ Chi-Quadrat-Test)
$X_1, X_2, ..., X_n \overset{iid}{\sim} N(\mu, \sigma^2)$, $\vartheta = (\mu, \sigma^2)$, σ_0^2 fest
(a)
$H_0 : \sigma^2 = \sigma_0^2$, $H_1 : \sigma^2 \neq \sigma_0^2$
<u>Ansatz</u>: Lehne H_0 ab, falls $\frac{\hat{S}_n^2}{\sigma_0^2} \ll 1$ oder $\frac{\hat{S}_n^2}{\sigma_0^2} \gg 1$, also $\frac{\hat{S}_n^2}{\sigma_0^2} \notin [c_1, c_2]$.

Bestimme c_1, c_2 sodass $\varphi \in \Phi_\alpha$.
Für $\vartheta \in \Theta_0 = \mathbb{R} \times \{\sigma_0^2\}$ gilt:

$$\begin{aligned}\beta_\varphi(\vartheta) &= 1 - P_\vartheta\left(c_1 \leq \frac{\hat{S}^2}{\sigma_0^2} \leq c_2\right)\\ &\overset{\vartheta\in\Theta_0}{=} 1 - P_{(\mu,\sigma_0^2)}\left((n-1)c_1 \leq \underbrace{\frac{(n-1)\hat{S}^2}{\sigma_0^2}}_{\sim\chi^2_{n-1}} \leq (n-1)c_2\right)\\ &= 1 - \left[F_{\chi^2_{n-1}}((n-1)c_2) - F_{\chi^2_{n-1}}((n-1)c_1)\right]\end{aligned}$$

Wähle zum Beispiel

$$c_1 = \frac{\chi^2_{n-1,\alpha/2}}{n-1} \quad c_2 = \frac{\chi^2_{n-1,1-\alpha/2}}{n-1}$$

$$\Rightarrow \beta_\varphi(\vartheta) \overset{\vartheta\in\Theta_0}{=} 1 - (1 - \alpha/2 - \alpha/2) = \alpha$$

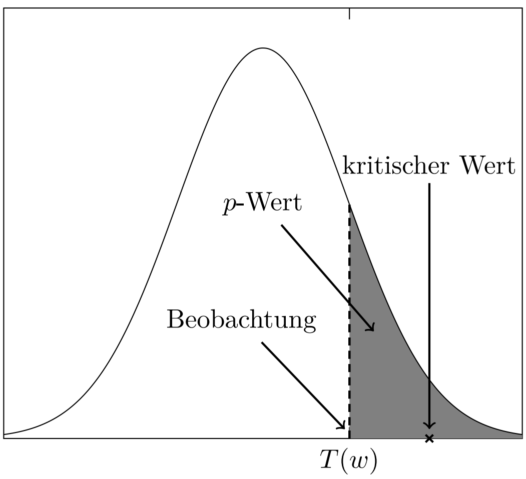
$\Rightarrow$ Zweiseitiger χ^2-Test:
Lehne H_0 ab, falls

$$\frac{(n-1)\hat{S}^2}{\sigma_0^2} \notin \left[\chi^2_{n-1,\alpha/2}, \chi^2_{n-1,1-\alpha/2}\right]$$

(b)
Einseitiger χ^2-Test
$H_0 : \sigma^2 \leq \sigma_0^2$ und $H_1 : \sigma^2 > \sigma_0^2$ (oder $\geq$ und $<$)
Ablehnung von H_0 falls

$$\frac{(n-1)\hat{S}^2}{\sigma_0^2} > \chi^2_{n-1,1-\alpha} \quad (< \chi^2_{n-1,\alpha})$$

Bemerkung (p-Wert/ p-Value)



Definition: 8.10

Sei $\alpha \in [0, 1]$. Ein Test $\varphi_0 \in \Phi_\alpha$ heißt <u>gleichmäßig bester</u> oder '<u>U</u>niformly <u>M</u>ost <u>P</u>owerful' Test zum Niveau α, falls

$$\beta_{\varphi_0}(\vartheta) = \sup_{\varphi \in \Phi_\alpha} \beta_\varphi(\vartheta) \quad \forall\, \vartheta \in \underline{\Theta_1}$$

Kapitel 9

Neyman-Pearson-Lemma

Vorbemerkung (9.1)
Einfache Hypothesen, $\vartheta = \vartheta_0$ vs. $\vartheta = \vartheta_1$, $\Theta = \{\vartheta_0, \vartheta_1\}$
UMP (Uniformly Most Powerful) Test maximiert $\beta_\varphi(\vartheta_1) = E_{\vartheta_1}\varphi(x)$ unter den Tests $\varphi \in \Phi_\alpha$, d.h. $E_{\vartheta_0}(\varphi(x)) \leq \alpha$
Modell: $\{f_{\vartheta_0}, f_{\vartheta_1}\}$ bzw. $\{f_0, f_1\}$

Theorem: 9.2 Neyman-Pearson Lemma

(a)
Jeder (Neyman-Pearson) Test der Form

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

mit $k \geq 0$, $\gamma \in [0,1]$ ist UMP Test zum Niveau $\alpha := \beta_\varphi(\vartheta_0)$ für $H_0 : \vartheta = \vartheta_0$ vs. $H_1 : \vartheta = \vartheta_1$
(b)
Für jedes $\alpha \in (0,1)$ existiert ein NP (Neyman-Pearson) Test φ mit

$$\gamma(x) \equiv \gamma \text{ und } \beta_\varphi(\vartheta_0) = \alpha$$

Mit (a) folgt φ ist UMP.

Bemerkung
Konstruktion nur für diskrete Zufallsvariablen notwendig, sonst gilt $\gamma = 0$.

kurze Wiederholung $H_0 : \vartheta = \vartheta_0$, $H_1 : \vartheta = \vartheta_1$
NP-Lemma

$$\begin{cases} 1 & > \\ \gamma(x) & f_1(x) = k \cdot f_0(x) \\ 0 & < \end{cases}$$

$(k \geq 0, \gamma \in [0,1])$
UMP-Test $\alpha = \beta_\varphi(\vartheta_0)$
kurze Wiederholung vorbei

Erweiterung Für $\alpha = 0$ und $\alpha = 1$
$\underline{\alpha = 1}$: abgedeckt durch (a) $k = 0$, $\gamma = 1$, d.h.

$$\varphi(x) = \begin{cases} 1 & f_1(x) \geq 0 \\ 0 & f_1(x) < 0 \end{cases}$$

$\underline{\alpha = 0}$: Setze $\varphi(x) = 1\{f_0(x) = 0\}$ d.h. $\varphi \equiv 0$ fast sicher.
Das entspricht '$k = \infty$'

$$\varphi(x) = \begin{cases} 1 & f_1(x) \geq \infty f_0(x) \\ 0 & f_1(x) < \infty f_0(x) \end{cases}$$

$$\beta_\varphi(\vartheta_0) = \int 1\{f_0(x) = 0\} f_0(x) dx = 0 \quad (= \alpha)$$

Bleibt zu zeigen, dass das der beste Test ist. Sei dazu $\psi \in \Phi_\alpha$, d.h.

$$\underbrace{E_{\vartheta_0}\psi}_{\int \psi(x) f_0(x) dx} \leq \alpha = 0$$

also $\psi = 0$ im Träger von f_0

$$\Rightarrow \beta_\varphi(\vartheta_1) - \beta_\psi(\vartheta_1) = \int 1\{f_0 > 0\}(\underbrace{\varphi(x) - \psi(x)}_{0-0}) f_1(x) dx + \int 1\{f_0 = 0\} \underbrace{(\varphi(x) - \psi(x))}_{\geq 0} f_1(x) dx \geq 0$$

$$\Rightarrow \beta_\varphi(\vartheta_1) \geq \beta_\psi(\vartheta_1)$$

Bemerkung (9.3)
(a)
Man kann zeigen: Ist φ UMP Test für $H_0 : \vartheta = \vartheta_0$, $H_1 : \vartheta = \vartheta_1$, dann ist φ NP-Test außer auf Nullmengen A mit $P_{\vartheta_0}(A) = P_{\vartheta_1}(A) = 0$
(b)
Vgl. Beispiel 5.15 ($\Theta = \vartheta_0, \vartheta_1, ..., \vartheta_k$): $\frac{f_1(x)}{f_0(x)}$ ist suffizient für $\vartheta \in \{\vartheta_0, \vartheta_1\}$
(c)
Ist T suffizient, dann ist $\frac{f_{\vartheta_1}(x)}{f_{\vartheta_0}(x)} = \frac{g_{\vartheta_1}(T(x))}{g_{\vartheta_0}(T(x))}$ NP-Test der nur von $T(x)$ abhängt.
Allgemein (Θ beliebig): Ist T suffizient, dann reicht es Tests zu betrachten, die nur von $T(x)$ abhängen, denn:
Ist φ Test $\Rightarrow E(\varphi(x)|T(x))$ ist auch Test (unabhängig von ϑ) mit Gütefunktion

$$E_\vartheta(E(\varphi(x)|T(x))) = E_\vartheta(\varphi(x)) = \beta_\varphi(\vartheta)$$

d.h. die Gütefunktionen stimmen überein.

Beispiel (9.4)
(a)

$X \sim f \in \{f_0, f_1\} \triangleq \{\vartheta_0, \vartheta_1\}$ mit

$$f_0(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad f_1(x) = \frac{1}{2} e^{-|x|}$$

$$\frac{f_1(x)}{f_0(x)} = \frac{\sqrt{2\pi}}{2} e^{-|x|+\frac{x^2}{2}} = \sqrt{\frac{\pi}{2}} e^{\frac{1}{2}(x^2-2|x|+1-1)} = \sqrt{\frac{\pi}{2}} e^{-\frac{1}{2}} e^{\frac{1}{2}(|x|-1)^2} \begin{cases} > \\ = \\ < \end{cases} k \Leftrightarrow ||x|-1| \begin{cases} > \\ = \\ < \end{cases} \tilde{k}$$

für $\tilde{k}$ mit

$$\alpha = E_{\vartheta_0}(\varphi(x)) = \int \underbrace{1\{||x|-1| > \tilde{k}\}}_{\varphi(x)} f_0(x) dx$$

φ ist UMP Test für $H_0 : f = f_0$, $H_1 : f = f_1$

(b)
$X_1, X_2, ..., X_n \overset{iid}{\sim} \text{Bin}(1, \vartheta)$, $\vartheta = \{\vartheta_0, \vartheta_1\}$, $\vartheta_0 < \vartheta_1$

$$f_i(x) = \vartheta_i^{T(x)} (1-\vartheta_i)^{n-T(x)} \quad (i = 0, 1), \quad T(x) = \sum_{i=1}^{n} X_i \text{ (suffizient)}$$

$$\varphi(x) = 1\left\{\frac{f_1(x)}{f_0(x)} > k\right\} + \gamma 1\left\{\frac{f_1(x)}{f_0(x)} = k\right\} \text{ NP Test}$$

φ ist UMP $\in \Phi_\varphi$ falls $\alpha = E_{\vartheta_0}(\varphi(x))$

$$\begin{aligned} \frac{f_1(x)}{f_0(x)} &= \left(\frac{\vartheta_1}{\vartheta_0}\right)^{T(x)} \left(\frac{1-\vartheta_1}{1-\vartheta_0}\right)^{n-T(x)} \\ &= \Bigg(\underbrace{\frac{\vartheta_1}{\vartheta_0}}_{>1}\Bigg)^{T(x)} \left(\frac{1-\vartheta_1}{1-\vartheta_0}\right)^{n} \Bigg(\underbrace{\frac{1-\vartheta_0}{1-\vartheta_1}}_{>1}\Bigg)^{T(x)} \end{aligned}$$

streng monoton wachsend in $T(x)$.
Finde also $\tilde{k}, \tilde{\gamma}$ mit

$$\tilde{\varphi}(x) = \begin{cases} 1 & T(x) > \tilde{k} \\ \tilde{\gamma} & T(x) = \tilde{k} \\ 0 & T(x) < \tilde{k} \end{cases} \text{ und } E_{\vartheta_0}(\tilde{\varphi}(x)) = \alpha$$

Diese Bedingungen erfüllt der einseitige Binomialtest aus (8.6) ($H_0 : \vartheta = \vartheta_0, H_1 : \vartheta > \vartheta_0$)
Der Test ist auch UMP unter $H_0 : \vartheta \leq \vartheta_0$, $H_1 : \vartheta > \vartheta_0$

Kapitel 10

Monotone Dichtequotienten

Bemerkung (10.1)
Parameterraum beim NP Lemma: $\Theta = \{\vartheta_0, \vartheta_1\}$ (einfache Hypothesen)
Dennoch ist NP Lemma wichtig für kompliziertere Hypothesen, z.B.

$$H_0 : \vartheta \leq \vartheta_0 \quad H_1 : \vartheta > \vartheta_0 \text{ einseitige Tests}$$

falls die Verteilungsfamilie einen isotonen (monoton wachsenden) Dichtequotienten hat.

Definition: 10.2

Die Familie von Dichten $\{f_\vartheta : \vartheta \in \Theta \subseteq \mathbb{R}\}$ hat einen (strikt) isotonen Dichtequotienten in $T : \mathcal{X} \to \mathbb{R}$ falls $\forall\, \vartheta_0, \vartheta_1 \in \Theta$ mit $\vartheta_0 < \vartheta_1$ gilt:
Es existiert eine streng monoton wachsende Funktion

$$H_{\vartheta_0,\vartheta_1}(T(.)) : \mathbb{R} \to [0, \infty)$$

mit

$$\frac{f_{\vartheta_1}(x)}{f_{\vartheta_0}(x)} = H_{\vartheta_0,\vartheta_1}(T(x)) \quad \forall\, x \text{ mit } f_{\vartheta_1}(x) + f_{\vartheta_0}(x) > 0$$

Beispiel (10.3)
(a)
$f_\vartheta(x) = \exp(Q(\vartheta)T(x) + S(x) + D(\vartheta))$ ein-parametrige Exponentialfamilie mit Q (streng) monoton wachsend hat einen (streng) monotonen Dichtequotienten (DQ) in T:

$$\begin{aligned}\frac{f_{\vartheta_1}(x)}{f_{\vartheta_0}(x)} &= \exp\big(\underbrace{[Q(\vartheta_1) - Q(\vartheta_0)]}_{\geq 0 \text{ für } \vartheta_0 < \vartheta_1} T(x)\big) \underbrace{\exp\left(D(\vartheta_1) - D(\vartheta_2)\right)}_{>0} \\ &= H_{\vartheta_0,\vartheta_1}(T(x))\end{aligned}$$

(b)

$X_1, X_2, ..., X_n \overset{iid}{\sim} U[0,\vartheta]$, $\vartheta_0 < \vartheta_1$, $\vec{x}$ mit $f_{\vartheta_0}(\vec{x}) + f_{\vartheta_1}(\vec{x}) > 0$. Dann gilt

$$\begin{aligned}\frac{f_{\vartheta_1}(\vec{x})}{f_{\vartheta_0}(\vec{x})} &= \frac{\left(\frac{1}{\vartheta_1}\right)^n 1(X_{(n)} \le \vartheta_1)}{\left(\frac{1}{\vartheta_0}\right)^n 1(X_{(n)} \le \vartheta_0)} \\ &= \begin{cases} \left(\frac{\vartheta_0}{\vartheta_1}\right)^n & X_{(n)} \in [0, \vartheta_0] \\ \infty & X_{(n)} \in (\vartheta_0, \vartheta_1] \end{cases} \\ &= H_{\vartheta_0,\vartheta_1}(X_{(n)})\end{aligned}$$

streng monoton wachsend
$\Rightarrow T(x)$ hat monoton wachsenden DQ (Dichte-Quotienten) in $T(x) = X_{(n)}$
(c)
Cauchy-Verteilung

$$f_\vartheta(x) = \frac{1}{\pi(1 + (x - \vartheta)^2)} \quad \vartheta \in \mathbb{R}, x \in \mathbb{R}$$

$$\frac{f_{\vartheta_1}(x)}{f_{\vartheta_0}(x)} = \frac{1 + (x - \vartheta_0)^2}{1 + (x - \vartheta_1)^2} \text{ kein isotoner DQ}$$

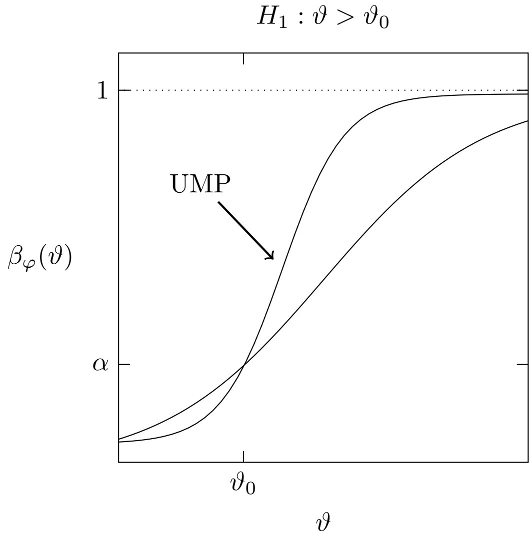
Theorem: 10.4

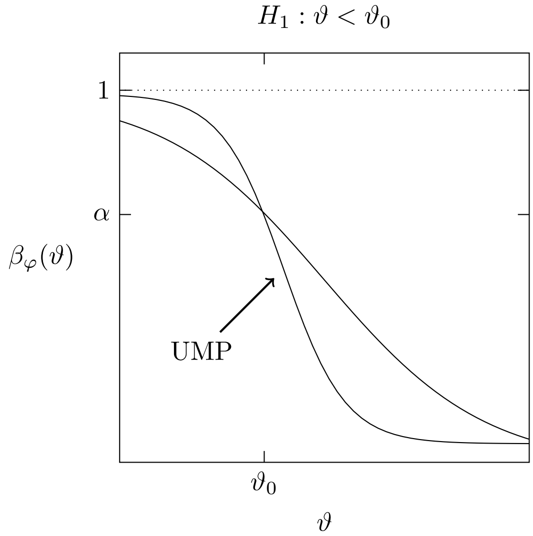
Betrachte $H_0 : \vartheta \le \vartheta_0$, $H_1 : \vartheta > \vartheta_0$, Zufallsvariable $X : \Omega \to \mathcal{X}$ mit Dichte f_ϑ, $\vartheta \in \Theta \subseteq \mathbb{R}$.
Angenommen $\{f_\vartheta : \vartheta \in \Theta\}$ hat einen monotonen Dichtequotienten in $T : \mathcal{X} \to \mathbb{R}$, dann gilt
(a)
Jeder Test der Form (mit noch zu bestimmenden t_0)

$$\varphi(x) = \begin{cases} 1 & T(x) > t_0 \\ \gamma & T(x) = t_0 \\ 0 & T(x) < t_0 \end{cases}$$

hat eine monoton wachsende Gütefunktion $\beta_\varphi(\vartheta)$.
φ ist UMP Test zum Niveau $\alpha = E_{\vartheta_0}(\varphi(x))$ (falls $\alpha > 0$) für die Hypothesen $H_0 : \vartheta \le \vartheta_0$, $H_1 : \vartheta > \vartheta_0$.
(b)
Für jedes $\alpha \in (0, 1]$ und jedes $\vartheta_0 \in \Theta$ existiert ein $t_0 \in [-\infty, \infty]$ und $\gamma \in [0, 1]$, sodass der Test aus Teilaussage (a) ein UMP Test zum Niveau α für obige Hypothesen ist.

Bemerkung (10.5)
Analog: UMP Test $H_0 : \vartheta \ge \vartheta_0$, $H_1 : \vartheta < \vartheta_0$ (Ablehnen von H_0 falls $T(X) < t_0$)
Diese Tests minimieren die Fehlerwahrscheinlichkeit 1. und 2. Art.
Vgl. Binomialtest: Konstruktionsvorschrift für γ in 10.4 (b) $\rightsquigarrow$ UMP





Beispiel (10.6 Einseitiger Gauß-Test)
$H_0 : \mu \leq \mu_0$, $H_1 : \mu > \mu_0$, $X_1, X_2, ..., X_n$ iid $N(\mu, \sigma^2)$ mit σ^2 bekannt (sonst t-Test)
Dann gilt

$$\begin{aligned}
\frac{f_{\mu_1}(x)}{f_{\mu_0}(x)} &= \frac{\exp\left(-\sum_{i=1}^n \frac{(X_i-\mu_1)^2}{2\sigma^2}\right)}{\exp\left(-\sum_{i=1}^n \frac{(X_i-\mu_0)^2}{2\sigma^2}\right)} \\
&= \exp\left(\frac{1}{2\sigma^2}\sum_{i=1}^n -\left[\cancel{X_i^2} + \mu_1^2 - 2\mu_1 X_i - \cancel{X_i^2} - \mu_0^2 + 2\mu_0 X_i\right]\right) \\
&= \exp\left(\frac{1}{\sigma^2}\underbrace{\sum_{i=1}^n X_i}_{=T(X)}(\mu_1-\mu_0)\right)\exp\left(\frac{n}{2\sigma^2}(\mu_0^2-\mu_1^2)\right) \\
&= H_{\mu_0,\mu_1}(T(x))
\end{aligned}$$

$\Rightarrow T(X)$ isotoner Dichte-Quotient (monoton wachsend für $\mu_1 - \mu_0 > 0$)
$T(X) \sim N\left(n\mu, n\sigma^2\right)$

$$P_\mu(T(x) = t_0) = 0 \Rightarrow \gamma = 0$$

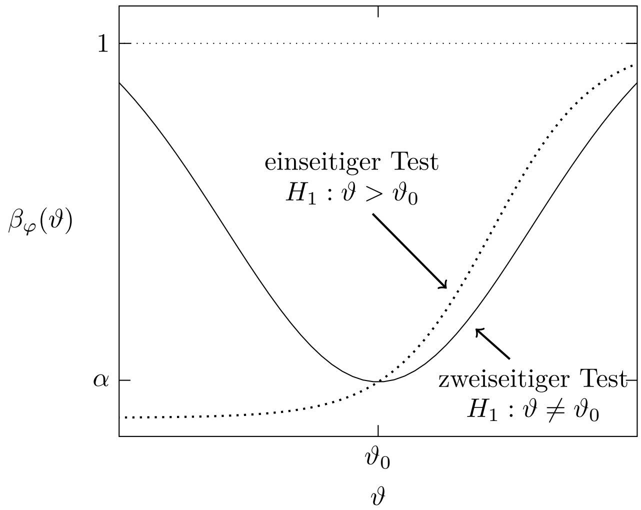
$$\varphi(x) = 1(T(x) > t_0) \text{ mit } t_0 \text{ sodass } E_{\mu_0}(\varphi(x)) = \alpha :$$

$$\begin{aligned}
P_{\mu_0}(T(x) > t_0) &= P_{\mu_0}\left(\underbrace{\frac{T(x) - n\mu_0}{\sqrt{n\sigma^2}}}_{\sim N(0,1)} > \frac{t_0 - n\mu_0}{\sqrt{n\sigma^2}}\right) \\
&= 1 - \Phi\underbrace{\left(\frac{t_0 - n\mu_0}{\sqrt{n\sigma^2}}\right)}_{=U_{1-\alpha}=\Phi^{-1}(1-\alpha)} \stackrel{!}{=} \alpha
\end{aligned}$$

Auflösen nach t_0 (Quantil in Tabelle nachgucken z.B. $U_{0.95} = 1.645$)

Bemerkung (10.7)
(a)
$H_0 : \vartheta \notin (\vartheta_1, \vartheta_2)$, $H_1 : \vartheta \in (\vartheta_1, \vartheta_2)$
UMP Test existiert in ein-parametrigen Exponentialfamilien mit streng isotonen Dichtequotienten (siehe Lehmann)

(b)
$H_0 : \vartheta \in [\vartheta_1, \vartheta_2]$, $H_1 : \vartheta \notin [\vartheta_1, \vartheta_2]$ bzw. $\vartheta = \vartheta_0$, $\vartheta \neq \vartheta_0$
UMP Test existiert im Allgemeinen nicht, weitere Einschränkung nötig (siehe 11.2)



Kapitel 11

Unverfälschte Tests

Für zweiseitige Tests wie in Bemerkung 10.7(b) ist eine weitere Einschränkung nötig:

Definition: 11.1

Ein Test $\varphi \in \Phi_\alpha$ für $H_0 : \vartheta \in \Theta_0$, $H_1 : \vartheta \in \Theta_1$ heißt unverfälscht zum Niveau α (unbiased), falls

$$\beta_\varphi(\vartheta) \geq \alpha \quad \forall\, \vartheta \in \Theta_1$$

$\Phi_{\alpha\alpha}$ bezeichne die Menge dieser Tests.
Der Test $\varphi_0 \in \Phi_{\alpha\alpha}$ heißt gleichmäßig bester unverfälschter Test (UMPU - UMP Unbiased), falls

$$\beta_{\varphi_0}(\vartheta) \geq \beta_\varphi(\vartheta) \quad \forall\, \vartheta \in \Theta_1 \forall\, \varphi \in \Phi_{\alpha\alpha}$$

Theorem: 11.2

Betrachte ein-parametrige Exponentialfamilie $\{f_\vartheta : \vartheta \in \Theta\}$ mit streng isotonen Dichte-Quotienten T. Sei $\alpha \in (0,1)$, $\vartheta_0, \vartheta_1, \vartheta_2 \in \Theta$, $\vartheta_1 < \vartheta_2$.
(i)
Es existiert ein UMPU Test $\varphi \in \Phi_{\alpha\alpha}$ für $H_0 : \vartheta \in [\vartheta_1, \vartheta_2]$, $H_1 : \vartheta \notin [\vartheta_1, \vartheta_2]$ mit (t_1, t_2 noch zu bestimmen)

$$\varphi(x) = \begin{cases} 1 & T(x) \notin [t_1, t_2] \\ \gamma_i & T(x) = t_1 \text{ oder } T(x) = t_2 \\ 0 & \text{sonst} \end{cases}$$

mit γ_i, t_i, sodass $\beta_\varphi(\vartheta_1) = \beta_\varphi(\vartheta_2) = \alpha$

(ii)
Für $H_0 : \vartheta = \vartheta_0$, $H_1 : \vartheta \neq \vartheta_0$ existiert ein Test φ wie in Teilaussage (i) der die folgenden zwei Bedingungen erfüllt:
$\beta_\varphi(\vartheta_0) = \alpha$ und $E_{\vartheta_0}(\varphi(x) \cdot T(x)) = \underbrace{E_{\vartheta_0}(\varphi(x))}_{=\alpha} E_{\vartheta_0}(T(x))$

Bemerkung (11.3)
Die Bedingungen aus 11.2(ii) bedeutet, dass die Gütefunktion die Steigung Null hat.

Beispiel (11.4)
$X_1, X_2, ..., X_n \overset{iid}{\sim} N(\mu, \sigma^2)$, σ^2 bekannt
(a)
<u>Gaußtest auf relevanten Unterschied</u>
$H_0 : \mu \in [-\varepsilon, \varepsilon]$, $H_1 : \mu \notin [-\varepsilon, \varepsilon]$
In Theorem 10.6 gezeigt: strikt isotoner DQ in $T(x) = \sum_{i=1}^n X_i$
Theorem 11.2: $\varphi(x) = 1\{T(x) \notin [t_1, t_2]\} \in \Phi_{\alpha\alpha}$ UMPU (T stetige Zufallsvariable $\Rightarrow \gamma = 0$)
Es gilt

$$\begin{aligned}
\beta_\varphi(\mu) &= 1 - P_\mu\left(t_1 < \sum X_i < t_2\right) \\
&= 1 - P_\mu\left(\frac{t_1 - n\mu}{\sqrt{n\sigma^2}} < \underbrace{\frac{\sum X_i - n\mu}{\sqrt{n\sigma^2}}}_{\sim N(0,1)} < \frac{t_2 - n\mu}{\sqrt{n\sigma^2}}\right) \\
&= 1 - \Phi\left(\frac{t_2 - n\mu}{\sqrt{n\sigma^2}}\right) + \Phi\left(\frac{t_1 - n\mu}{\sqrt{n\sigma^2}}\right) \overset{!}{=} \alpha \text{ für } \mu = \varepsilon, \mu = -\varepsilon
\end{aligned}$$

Wähle $0 < t_1 = -t_2 < 0$ (Symmetrie)
Für $\mu = \varepsilon$ und $\mu = -\varepsilon$ erhält man

$$\beta_\varphi(\mu)\Big|_{\mu=\pm\varepsilon} = 1 - \underbrace{\Phi\left(\frac{t_2 - n\varepsilon}{\sqrt{n\sigma^2}}\right)}_{\Phi^{-1}(1-\alpha/2)} + \underbrace{\Phi\left(\frac{-t_2 - n\varepsilon}{\sqrt{n\sigma^2}}\right)}_{\Phi^{-1}(\alpha/2)} \overset{!}{=} \alpha$$

$[\Phi(x) = 1 - \Phi(-x)]$
Kann eindeutig nach t_2 aufgelöst werden.

(b)
<u>Zweiseitiger Gaußtest</u>
$H_0 : \mu = \mu_0 = 0$, $H_1 : \mu \neq 0$
$\varphi(x) = 1\{\sum X_i \notin [t_1, t_2]\}$ mit t_1, t_2 sodass

$$\alpha = \beta_\varphi(\mu_0) = \beta_\varphi(0) = 1 - \Phi\left(\frac{t_2 - n \cdot 0}{\sqrt{n\sigma^2}}\right) + \Phi\left(\frac{t_1 - n \cdot 0}{\sqrt{n\sigma^2}}\right)$$

$$\begin{aligned}
0 = \beta'_\varphi(0) = \frac{d}{d\mu}\beta_\varphi(\mu)\Big|_{\mu=0} &= -\Phi'\left(\frac{t_2}{\sqrt{n\sigma^2}}\right)\left(-\frac{n}{\sqrt{n\sigma^2}}\right) + \Phi'\left(\frac{t_1}{\sqrt{n\sigma^2}}\right)\left(-\frac{n}{\sqrt{n\sigma^2}}\right) \\
&= \sqrt{\frac{n}{\sigma^2}}\left(\Phi'\left(\frac{t_2}{\sqrt{n\sigma^2}}\right) - \Phi'\left(\frac{t_1}{\sqrt{n\sigma^2}}\right)\right) = 0 \text{ für } t_2 = -t_1
\end{aligned}$$

($\Phi' = \varphi$ Dichte von $N(0,1)$-Verteilung)
Mit $t_2 = -t_1 = \sqrt{n\sigma^2}\ \underbrace{U_{1-\alpha/2}}_{\Phi^{-1}(1-\alpha/2)}$ gilt auch

$$\beta_\varphi(0) = 1 - \Phi\left(\frac{t_2}{\sqrt{n\sigma^2}}\right) + \Phi\left(\frac{t_1}{\sqrt{n\sigma^2}}\right) = 1 - \Phi(U_{1-\alpha/2}) + \Phi(U_{\alpha/2}) = 1 - (1 - \alpha/2) + \alpha/2 = \alpha$$

$$\varphi(x) = 1\left\{\left|\sum X_i\right| > \sqrt{n\sigma^2}U_{1-\alpha/2}\right\} = 1\left\{|\bar{X}| > \frac{\sigma}{\sqrt{n}}U_{1-\alpha/2}\right\}$$

(bzw. äquivalent $\left|\frac{\sum X_i - 0}{\sqrt{n\sigma^2}}\right| > U_{1-\alpha/2}$)

<u>Allgemeiner</u>: $H_0 : \mu = \mu_0$, $H_1 : \mu \neq \mu_0$

$$\varphi(x) = 1\left\{|\bar{X} - \mu_0| > \frac{\sigma}{\sqrt{n}}U_{1-\alpha/2}\right\}$$

Ist σ^2 <u>un</u>bekannt ergibt sich mit $\hat{S}^2 \xrightarrow{P} \sigma^2$ der <u>zwei-seitige t-Test</u>:

$$\varphi(x) = 1\left\{|\bar{X} - \mu_0| > \sqrt{\frac{\hat{S}^2}{n}}t_{n-1,1-\alpha/2}\right\}$$

(c)
$X_1, X_2, ..., X_n$ iid $N(0, \sigma^2)$, $H_0 : \sigma = \sigma_0$, $H_1 : \sigma \neq \sigma_0$.
Bekannt: $\sum X_i^2$ suffizient. Die Familie hat strikt isotonen Dichtequotienten in $\sum X_i^2$ denn (Definition 10.2) $\forall\, \sigma_0, \sigma_1 \in \Theta$ mit $\sigma_0 < \sigma_1$ gilt

$$\frac{f_{\sigma_1}(x)}{f_{\sigma_0}(x)} = \frac{\sigma_0}{\sigma_1}\frac{\exp\left(-\frac{\sum X_i^2}{2\sigma_1^2}\right)}{\exp\left(-\frac{\sum X_i^2}{2\sigma_0^2}\right)} = \underbrace{\frac{\sigma_0}{\sigma_1}}_{>0} \exp\left(\frac{1}{2}\sum X_i^2\underbrace{\left(\frac{1}{\sigma_0^2} - \frac{1}{\sigma_1^2}\right)}_{>0}\right)$$

Nach Theorem 11.2 gilt

$$\varphi(x) = 1\{\sum X_i^2 \notin (\tilde{t}_1, \tilde{t}_2)\} = 1\left\{\underbrace{\frac{\sum X_i^2}{\sigma_0^2}}_{T(x)\overset{H_0}{\sim}\chi_n^2} \notin \left(\underbrace{\frac{\tilde{t}_1}{\sigma_0^2}, \frac{\tilde{t}_2}{\sigma_0^2}}_{t_1, t_2}\right)\right\} \in \Phi_{\alpha\alpha}$$

1. Bedingung

$$\beta_\varphi(\sigma_0^2) = 1 - P_{\sigma_0}\left(t_1 < \frac{\sum X_i^2}{\sigma_0^2} < t_2\right) \overset{!}{=} \alpha$$

$$\Rightarrow \int_{t_1}^{t_2} f_n(x) \overset{!}{=} 1 - \alpha$$

mit f_n Dichte der χ_n^2-Verteilung
2. Bedingung

$$\begin{aligned}
E_{\sigma_0}(\varphi(x)T(x)) &= \alpha E_{\sigma_0}(T(x)) = \alpha n \\
&= \int 1_{(t_1,t_2)^c}(x)xf_n(x)dx \\
&\overset{\text{Ü}}{=} n\int 1_{(t_1,t_2)^c}f_{n+2}(x)dx \\
&\Leftrightarrow 1 - \int_{t_1}^{t_2} f_{n+2}(x)dx = \alpha \Leftrightarrow 1 - \alpha = \int_{t_1}^{t_2} f_{n+2}(x)dx
\end{aligned}$$

Bestimme t_1, t_2 numerisch sodass beide Bedingungen gelten.

Kapitel 12

Likelihood-Quotienten Tests

Allgemeines Prinzip zur Herleitung von Tests der Form $H_0 : \vartheta \in \Theta_0$ vs. $\vartheta \in \Theta_1$, $\Theta := \Theta_1 \dot{\cup} \Theta_0$, $\Theta_1 \neq \emptyset \neq \Theta_0$

Definition: 12.1

Gegeben sei das statistiche Experiment/ Modell mit Dichten $\{f_\vartheta : \vartheta \in \Theta\}$.
Wir nennen

$$\lambda(x) := \frac{\sup\limits_{\vartheta \in \Theta_0} f_\vartheta(X)}{\sup\limits_{\vartheta \in \Theta} f_\vartheta(X)} = \frac{\sup\limits_{\vartheta \in \Theta_0} L(\vartheta, x)}{\sup\limits_{\vartheta \in \Theta} L(\vartheta, x)} \text{ ('klein' spricht gegen } H_0)$$

Likelihood-Quotient (LQ) (LR - likelihood ratio).
Ein Test der Form

$$\varphi(x) := 1\{\lambda(X) < c\} + \gamma 1\{\lambda(X) = c\}, \gamma, c \in [0, 1]$$

Likelihood-Quotienten Test (LQT).
Erläuterung:
Für $\vartheta \in \Theta_1$ gilt '$\lambda(x)$ klein'
Für $\vartheta \in \Theta_0$ gilt '$\lambda(x) \approx 1$'
Ist f_ϑ stetig, dann ist $\gamma = 0$
- Spezialfall: $\Theta_0 = \{\vartheta_0\}, \Theta_1 = \{\vartheta_1\}$ Neyman-Pearson Test
- in ein-parametrige Exponentialfamilien mit isotonen Dichtequotienten sind LQT UMPT für $H_0 : \vartheta \leq \vartheta_0$ vs. $H_1 : \vartheta > \vartheta_0$ zum Niveau $\alpha = \sup_{\vartheta \in \Theta_0} \beta_\varphi(\vartheta)$

Beispiel (12.2)
(a)
$X_1, X_2, ..., X_n$ iid $N(\mu, \sigma^2)$, $\nu := \sigma^2$, $\theta := (\mu, \sigma^2)$

$H_0 : \mu = \mu_0$ vs. $H_1 : \mu \neq \mu_0$, $\Theta_0 = \{\mu_0\} \times (0, \infty)$, $\Theta_1 = \mathbb{R}\backslash\{\mu_0\} \times (0, \infty)$

$$L(X, \theta) = (2\pi\nu)^{-n/2} \exp\left(-\frac{1}{2}\nu^{-1} \sum_{j=1}^{n} (X_j - \mu)^2\right)$$

Dann gilt

$$\sup_{\vartheta \in \Theta} L(\vartheta, x) = L(\hat{\vartheta}, x) \qquad \sup_{\vartheta \in \Theta_0} L(\vartheta, x) = L(\tilde{\vartheta}_0, x)$$

mit $\hat{\vartheta}$ ML Schätzer für $\vartheta = (\mu, \sigma^2)$, $\hat{\vartheta} = (\bar{X}, \hat{\sigma}^2)$
und $\tilde{\vartheta}_0 = (\mu_0, \tilde{\sigma}_0^2)$ ML Schätzer für σ^2 bei bekanntem $\mu = \mu_0$, $\tilde{\sigma}_0^2 = \frac{1}{n} \sum (X_i - \mu_0)^2$

$$\begin{aligned}
\Rightarrow \lambda(X) = \frac{L(\tilde{\vartheta}_0, x)}{L(\hat{\vartheta}_0, x)} &= \left(\frac{\hat{\sigma}^2}{\tilde{\sigma}^2}\right)^{\frac{n}{2}} \exp \underbrace{\left(-\frac{1}{2\tilde{\sigma}^2} \sum_{j=1}^{n} (X_j - \mu_0)^2 + \frac{1}{2\hat{\sigma}^2} \sum_{j=1}^{n} (X_j - \bar{X}_n)^2\right)}_{-\frac{n}{2} + \frac{n}{2} = 0} \\
&= \left(\frac{\sum_{j=1}^{n} (X_j - \bar{X}_n)^2}{\sum_{j=1}^{n} (X_j - \mu_0)^2}\right)^{n/2} \cdot 1
\end{aligned}$$

Nebenrechnung

$$\begin{aligned}
\sum_{j=1}^{n} (X_i - \mu_0)^2 &= \sum_{j=1}^{n} ((X_j - \bar{X}_n) + (\bar{X}_n - \mu_0))^2 \\
&= \sum_{j=1}^{n} (X_j - \bar{X})^2 + n(\bar{X} - \mu_0)^2 + 2 \sum_{j=1}^{n} (X_j - \bar{X})(\bar{X} - \mu_0) \\
&= \sum_{j=1}^{n} (X_j - \bar{X})^2 + n(\bar{X} - \mu_0)^2 + 2n(\bar{X} - \mu_0) \underbrace{\sum_{j=1}^{n} X_j - \bar{X}}_{n\bar{X} - n\bar{X} = 0}
\end{aligned}$$

Nebenrechnung vorbei.

$$\begin{aligned}
\lambda(x) &= \left(\frac{\sum_{j=1}^{n} (X_j - \bar{X})^2}{\sum_{j=1}^{n} (X_j - \bar{X})^2 + n(\bar{X} - \mu_0)^2}\right) \\
&= \left(\frac{1}{1 + \underbrace{\frac{n(\bar{X} - \mu_0)^2}{\sum_{j=1}^{n} (X_j - \bar{X})^2}}_{=: T^2(X_{(n)})}}\right)^{\frac{n}{2}}
\end{aligned}$$

(Sei jetzt $X_{(n)} = (X_1, ..., X_n)$) streng monoton fallend in

$$T(X_{(n)}) = \frac{|\bar{X} - \mu_0|}{\sqrt{\sum_{j=1}^{n} (X_j - \bar{X})^2}} \geq 0$$

Die Tests

$$\varphi(x) := 1_{(-\infty,c)}(\lambda(X)) \text{ und } \tilde{\varphi}(X) := 1_{(\tilde{c},\infty)}(T(X))$$

sind äquivalent.
$T(X) > \tilde{c} \Leftrightarrow \sqrt{\frac{n}{n-1}}T(X) > \sqrt{\frac{n}{n-1}}\tilde{c} = \tilde{\tilde{c}}$, betr. also

$$\tilde{\tilde{\varphi}}(X) = 1\Bigg\{\underbrace{\left|\frac{\sqrt{n}(\bar{X}-\mu_0)^2}{\sqrt{\hat{S}_n^2}}\right|}_{\overset{H_0}{\sim} t_{n-1}} > \tilde{\tilde{c}}\Bigg\}$$

$$E(\tilde{\tilde{\varphi}}(X)) = P(|\tilde{T}| > \tilde{\tilde{c}}) = 2 - 2F(\tilde{\tilde{c}}) \overset{!}{\leq} \alpha$$

folgt mit Symmetrie und Stetigkeit

$$\rightsquigarrow \tilde{c} = t_{n-1,1-\frac{\alpha}{2}}$$

Nachrechnung

$$|\tilde{T}| \leq \tilde{\tilde{c}} \Leftrightarrow -\tilde{\tilde{c}} \leq \tilde{T} \leq \tilde{\tilde{c}}$$

$$F(\tilde{\tilde{c}}) - F(-\tilde{\tilde{c}}) = F(\tilde{\tilde{c}}) - (1 - F(\tilde{\tilde{c}})) = 2F(\tilde{\tilde{c}}) - 1$$

Nachrechnung vorbei
(b)
Mit dem LQ Ansatz erhält man auch andere t-Tests aus (8.8):

$$H_0 : \mu \begin{cases} \leq \mu_0 \\ \geq \mu_0 \end{cases} \quad \text{vs. } H_1 : \mu \begin{cases} > \mu_0 \\ < \mu_0 \end{cases} \quad ,$$

lehne H_0 ab, falls

$$\frac{\sqrt{n}(\bar{X}-\mu)}{\hat{S}_n} \begin{cases} > t_{n-1,1-\alpha} \\ < t_{n-1,\alpha} \end{cases}$$

(c)
Der zweiseitige χ^2-Test aus (8.9) ist auch LRT, ebenso die einseitigen Versionen

$$H_0 : \sigma^2 \begin{cases} \leq \\ \geq \\ = \end{cases} \sigma_0^2 \text{ vs. } H_1 : \sigma^2 \begin{cases} > \\ < \\ \neq \end{cases} \sigma_0^2$$

H_0 ablehnen, wenn

$$\frac{(n-1)\hat{S}_n^2}{\sigma_0^2} \begin{cases} > \chi^2_{n-1,1-\alpha} \\ < \chi^2_{n-1,\alpha} \\ > \chi^2_{n-1,1-\frac{\alpha}{2}} \text{ oder } < \chi^2_{n-1,\frac{\alpha}{2}} \end{cases}$$

Beispiel (12.3 (Zu 12.2a asymptotisch äquivalenter Test))

$$\lambda(X) = \left(\frac{\sum_{j=1}^{n}(X_j - \mu_0)^2}{\sum_{j=1}^{n}(X_j - \bar{X})^2}\right)^{-n/2} = \left(1 + \frac{n(\bar{X} - \mu_0)^2}{\sum_{j=1}^{n}(X_j - \bar{X})^2}\right)^{-n/2} = \left(1 + \frac{n(\bar{X} - \mu_0)}{(n-1)\hat{S}^2}\right)^{-n/2}$$

$$\Rightarrow -2\ln(\lambda(X)) = n\ln\left(1 + \frac{n(\bar{X} - \mu_0)^2}{\sum_{j=1}^{n}(X_j - \bar{X})^2}\right) = n\ln\left(1 + \frac{\left(\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma}\right)^2}{(n-1)\frac{\hat{S}_n^2}{\sigma^2}}\right)$$

(Siehe Slutsky Lemma)
Es gilt $Z := \left(\frac{\sqrt{n}(\bar{X}-\mu_0)}{\sigma}\right) \overset{H_0}{\sim} N(0,1)$
$\frac{\hat{S}_n^2}{\sigma^2} \overset{P}{\to} 1$ (da Stichprobenvarianz konsistent), also haben $-2\ln(\lambda(X))$ und $T_n(X) = n\ln\left(1 + \frac{Z^2}{n-1}\right)$ dieselbe Grenzverteilung

$$T_n(X) \xrightarrow{D} Z^2 \sim \chi_1^2 \text{ weil } \exp(T_n(y)) = \left(1 + \frac{y^2}{n-1}\right)^n \longrightarrow \exp(y^2) \forall\, y \in \mathbb{R}$$

Kapitel 13

Konfidenzbereiche

Bereiche/ Intervalle in denen ein Parameter ϑ mit großer Wahrscheinlichkeit liegt (mehr Info als bei einem Punktschätzer)

Definition: 13.1

Sei $X \sim P_\vartheta, \vartheta \in \Theta$.
Eine Abbildung $S : \mathcal{X} \to \mathfrak{P}(\Theta)$ heißt Konfidenzbereich für ϑ zum Niveau $1-\alpha$ ($\alpha \in (0,1)$), falls

$$P_\vartheta(\vartheta \in S(x)) = P_\vartheta(x \in \mathcal{X} : \vartheta \in S(x)) \geq 1-\alpha \quad \forall\, \vartheta \in \Theta$$

Für eine Realisation $x := X(w)$ heißt $S(x)$ konkreter Konfidenzbereich und weiter heißt

$$\inf_{\vartheta \in \Theta} P(\vartheta \in S(x)) \text{ effektives Konfidenzniveau}$$

Ist $\Theta \subseteq \mathbb{R}$, $S(x)$ Intervall $\forall\, x \in \mathcal{X}$ dann nennt man $S(x)$ Konfidenzintervall.

Beispiel (13.2)
$X_1, X_2, ..., X_n \overset{iid}{\sim} N(\mu, \sigma^2)$
(a)
Konfidenzintervall für $\vartheta = \mu$ (Varianz ebenfalls unbekannt). Ansatz:

$$S(X) := [\underbrace{\bar{X}}_{\hat{\mu}} - \varepsilon_1, \bar{X} + \varepsilon_2]$$

$$\begin{aligned}
P_\mu(\mu \in S) &= P(\bar{X} - \varepsilon_1 \leq \mu \leq \bar{X} + \varepsilon_2) \\
&= P_\mu(-\varepsilon_2 \leq \bar{X} \leq \varepsilon_1) \\
&= P_\mu\left(\underbrace{-\sqrt{\frac{n}{\hat{S}_n^2}}\varepsilon_2}_{a} \leq \underbrace{\frac{\sqrt{n}(\bar{X}-\mu)}{\sqrt{\hat{S}_n^2}}}_{\sim t_{n-1}} \leq \underbrace{\sqrt{\frac{n}{\hat{S}_n^2}}\varepsilon_1}_{b} \right) \\
&= F_{t_{n-1}}(b) - F_{t_{n-1}}(a) \overset{!}{\geq} 1-\alpha
\end{aligned}$$

Für $b = -a = t_{n-1,1-\frac{\alpha}{2}}$ (Symmetrie) (kürzestes KI)

$$\Rightarrow \varepsilon_1 = b \cdot \frac{\hat{S}_n}{\sqrt{n}} = t_{n-1,1-\frac{\alpha}{2}} \frac{\hat{S}_n}{\sqrt{n}}, \quad \varepsilon_2 = -a \frac{\hat{S}_n}{\sqrt{n}} = t_{n-1,1-\frac{\alpha}{2}} \frac{\hat{S}_n}{\sqrt{n}}$$

$$\Rightarrow S(X) = \left[\bar{X} \mp \frac{\hat{S}_n}{\sqrt{n}} t_{n-1,1-\frac{\alpha}{2}} \right]$$

(b)
Konfidenzintervall für $\vartheta = \sigma^2$. Ansatz:

$$[c_1 \hat{S}_n^2, c_2 \hat{S}_n^2]$$

$$\begin{aligned} P_\vartheta(\sigma^2 \in S(X)) &= P_\vartheta(c_1 \hat{S}_n^2 \leq \sigma^2 \leq c_2 \hat{S}_n^2) \\ &= P_\vartheta\left(\frac{1}{c_1 \hat{S}_n^2} \geq \frac{1}{\sigma^2} \geq \frac{1}{c_2 \hat{S}_n^2} \right) \\ &= P_\vartheta\left(\frac{n-1}{c_2} \leq \underbrace{\frac{(n-1)\hat{S}_n^2}{\sigma^2}}_{\sim \chi^2_{n-1}} \leq \frac{n-1}{c_1} \right) \overset{!}{\geq} 1 - \alpha \end{aligned}$$

Wähle z.B.

$$\frac{n-1}{c_1} = \chi^2_{n-1,1-\alpha/2}, \quad \frac{n-1}{c_2} = \chi^2_{n-1,\alpha/2}$$

$$\Rightarrow S(X) = \left[\frac{n - 1\hat{S}_n^2}{\chi^2_{n-1,1-\alpha/2}}, \frac{(n-1)\hat{S}_n^2}{\chi^2_{n-1,\alpha/2}} \right]$$

<u>oder</u>

$$S(X) = \left[\frac{(n-1)\hat{S}_n^2}{a}, \frac{(n-1)\hat{S}_n^2}{b} \right]$$

mit a, b so gewählt, dass die intervallänge minimal ist und das Konfidenzniveau $1 - \alpha$.
(c)
$\vartheta = (\mu, \sigma^2)$
Man wählt im Allgemeinen Konfidenzellipsoide statt Quader $S_1(X) \times S_2(X)$ mit kleinerer Fläche.

Bemerkung (13.3)
Betrachte $H_0 : \vartheta = \vartheta_0$, $H_1 : \vartheta \neq \vartheta_0$, $S(X)$ Konfidenzbereich für ϑ zum Niveau $1 - \alpha$.
Dann gilt

$$\varphi(X) = \begin{cases} 1 & \vartheta_0 \notin S(X) \\ 0 & \vartheta_0 \in S(X) \end{cases}$$

ist Niveau α Test, denn
Fehler 1. Art:

$$P_{\vartheta_0}(\varphi(x) = 1) = 1 - \underbrace{P_{\vartheta_0}(\underbrace{\varphi(x) = 0}_{\vartheta_0 \in S(X)})}_{\geq 1-\alpha} \leq 1 - (1 - \alpha) = \alpha$$

Bsp. 13.2.a:

$$\mu_0 \notin S(X) = \left(\bar{X} - \frac{\sqrt{\hat{S}_n^2}}{\sqrt{n}} t_{n-1,1-\alpha/2}, \bar{X} + \frac{\sqrt{\hat{S}_n^2}}{\sqrt{n}} t_{n-1,1-\alpha/2}\right)$$
$$\Leftrightarrow |\bar{X} - \mu_0| \geq \frac{\sqrt{\hat{S}_n^2}}{\sqrt{n}} t_{n-1,1-\alpha/2} \text{ zweiseitiger t-Test}$$

Theorem: 13.4

Sei $H_0 : \vartheta = \vartheta_0,\ H_1 : \vartheta \in \Theta_1(\vartheta_0)$ (z.B. $\Theta \backslash \{\vartheta_0\}, (\vartheta_0, \infty)$)
Für jedes $\vartheta_0 \in \Theta$ (Θ hier ganzer Bereich, nicht nur $\Theta_1(\vartheta_0)$) sei (mit $\mathcal{A}(\vartheta_0)$ Annahmebereich von H_0)

$$\varphi_0(X) = 1 - 1\{X \in \mathcal{A}(\vartheta_0)\} = 1\{X \notin \mathcal{A}_{\vartheta_0}\}$$

UMP-Test zum Niveau $\alpha \in (0,1)$. Dann gilt

$$S(X) = \{\vartheta \in \Theta : X \in \mathcal{A}(\vartheta)\}$$

ist Konfidenzbereich zum Niveau $1-\alpha$ mit minimaler Fehlerwahrscheinlichkeit 2. Art

$$P_\vartheta(\vartheta_0 \in S(X)) \quad \forall\, \vartheta \in \Theta_1(\vartheta_0)$$

unter allen Konfidenzbereichen zum Niveau $1-\alpha$.

Beispiel (13.5 Exaktes Konfidenzintervall für Binomialverteilungen)
Sei $X \sim \mathrm{Bin}(n, \vartheta)$ z.B. $X = \sum_{i=1}^n X_i,\ X_i \overset{iid}{\sim} \mathrm{Bin}(1, \vartheta)$
Gesucht: Konfidenzintervall $S(X) = (l(X), L(X))$ mit

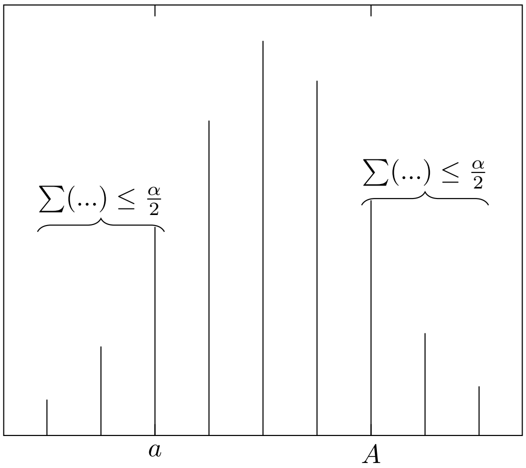
$$P_\vartheta(\vartheta \in S(X)) \geq 1 - \alpha \quad \forall\, \vartheta \in \Theta = [0,1]$$

Bestimme zunächst

$$\mathcal{A}(\vartheta) = \{x \in \{0,1,...,n\} : a(\vartheta) \leq x \leq A(\vartheta)\}$$

Wähle z.B.

$$a(\vartheta) = \max\left\{k \in \{0,1,...,n\} : \sum_{j=0}^{k-1} \binom{n}{j} \vartheta^j (1-\vartheta)^{n-j} \leq \alpha/2\right\}$$
$$A(\vartheta) = \min\left\{k \in \{0,1,...,n\} : \sum_{j=k+1}^{n} \binom{n}{j} \vartheta^j (1-\vartheta)^{n-j} \leq \alpha/2\right\}$$



$a(\vartheta), A(\vartheta)$ monoton wachsend in ϑ, $a(\vartheta)$ rechts-stetig, $A(\vartheta)$ links-stetig, also
$a(\vartheta) \leq x \leq A(\vartheta) \Leftrightarrow l(x) < \vartheta < L(x)$ mit
$L(x) = \sup\{\vartheta : a(\vartheta) = x\}$, $L(x)$ löst

$$\sum_{j=0}^{x} \binom{n}{j} \vartheta^j (1-\vartheta)^{n-j} = \frac{\alpha}{2} \text{ bzgl. } \vartheta$$

$l(x) = \inf\{\vartheta : A(\vartheta) = x\}$, $l(x)$ löst

$$\sum_{j=x}^{n} \binom{n}{j} \vartheta^j (1-\vartheta)^{n-j} = \frac{\alpha}{2} \text{ bzgl. } \vartheta$$

Numerisch lösbar oder mit 'Pearson-Clopper' Schranken (Zusammenhang Bin- und F-Verteilung)

Kapitel 14

Stochastische Prozesse und Versicherungsmathematik

Definition: 14.1

Betrachte $(\Omega, \mathcal{A}, P)$ und $T \neq \emptyset$ Indexmenge.
Eine Familie von Zufallsvariablen $\{X_t\}_{t\in T}$ auf $(\Omega, \mathcal{A}, P)$ heißt stochastischer Prozess.
Gilt $T \subseteq \mathbb{N}_0$, dann nennt man $\{X_t\}_{t\in T}$ Kette.

Motivation (14.2 Versicherungsmathematisches Modell)
Annahme: Versicherung bekommt von Kunden pro Zeiteinheit (z.B. Jahr) eine Prämie γ.
Sei $N(t) = \#\{\text{Schadensfälle in } [0,t]\}$
$N(t)$ wird im Allgemeinen durch Poisson-Prozess modelliert.
Auch von Interesse:
$\#\{$Schadensfälle bis variablen Zeipunkt $t\}$ (z.B. Zahlungsunfähigkeit)
$(Y_n)_{n\in\mathbb{N}}$ Höhe der Schadensfälle ($Y_1, Y_2, ..., Y_n$ iid)
$(X_t)_{t\in T}$ Kapital der Versicherung zum Zeitpunkt t (stochastischer Prozess)

$$X(t) = a + \gamma t - \sum_{j=1}^{N(t)} Y_j \quad (t \geq 0)$$

(a Startkapital)
Annahme: Zahlungen an bzw. von Kunden erfolgen zu äquidistanten Zeitpunkten $t_k = \frac{k}{N}$, $k \in \mathbb{N}_0$, der Einfachheit halber sei $t_k = k$ (d.h. $N = 1$)
Von Interesse/Ziel:

Bestimmung von $P_a = P(\text{'rote Zahlen bei Startkapital } a\text{'}) = P\left(\inf\limits_{k\in\mathbb{N}_0} X(k) < 0\right)$
Sei $X_k = X(k)$ $(= X(t_k))$, τ_a Zufallsvariable (zufälliger Zeitpunkt) mit $\tau_a = \inf\{k : X_k < 0\}$
Sei $\tau_a = \infty$ falls $X_k < 0$ nie eintritt.
Dann gilt

$$P_a = P(\inf X(k) < 0) = P(\tau_a < \infty)$$

Definition: 14.3

Ein stochastischer Prozess $N(.,t)$, $t \geq 0$ heißt Zählprozess, falls

$$(i)\ N(w,t) \in \mathbb{N}_0\ \forall\, w \in \Omega \text{ sowie } N(.,0) = 0 \text{ fast sicher}$$
$$(ii) \text{ Der Pfad } t \mapsto N(w,t) \text{ wächst monoton für jedes } w \in \Omega$$

Ein Zählprozess heißt (homogener) Poisson-Prozess mit Intensität $\lambda > 0$ wenn seine Zuwächse unabhängig und Poisson-verteilt sind, d.h. für beliebige $0 \leq t_1 < t_2 < ... < t_n$ sind die Zufallsvariablen

$$N(t_{j+1}) - N(t_j), \quad j = 1, ..., n-1$$

unabhängig und jeweils Poisson-verteilt mit Parameter $\lambda(t_{j+1} - t_j)$

Theorem: 14.4

Betrachte $X_1, X_2, ..., X_n \overset{iid}{\sim} \text{Exp}(\lambda)$ mit Dichte $\lambda e^{-\lambda x}, \lambda > 0, x > 0$
$S_n := \sum_{i=1}^{n} X_i$ und den Zählprozess

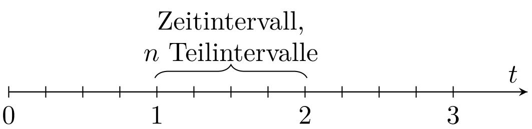
$$N(t) = N(t,w) = \sup\{n \in \mathbb{N}_0 : S_n(w) \leq t\}$$

(mit rechts-seitig stetigen Pfaden).
Sei $0 \leq t_1 < t_2 < ... \leq t_n$ dann gilt:

$$N(t_{j+1}) - N(t_j), \quad j = 1, ..., n-1$$

sind unabhängig und Pois($\lambda(t_{j+1} - t_j)$)-verteilt.

Heuristik zum Poisson-Prozess



Betrachte unabhängige Bernoulli-Experimente, Ausgang 0/1 in jedem Teilintervall.
Angenommen wir erwarten λ Erfolge pro Zeitintervall (z.B. $\lambda = 2.2$ Tore pro Fußballspiel)
Dann gilt

$$p = P(\text{'1'}) = \frac{\lambda}{n}$$

sind die Teilintervalle sehr klein, dann gilt

$$P(\text{'2 oder mehr Erfolge'}) \approx 0$$

Zeit bis zum 1. Erfolg $\approx \frac{1}{n} \cdot \underbrace{\#\{\text{Experimente bis zum 1. Erfolg}\}}_{=:G_n}$ mit $G_n \sim \text{Geo}(p)$, $p = \frac{\lambda}{n}$ (diskrete

Analogon zur Exponentialverteilung) und

$$\frac{G_n}{n} \xrightarrow{\text{D}} \text{Exponentialverteilung}$$

d.h. Wartezeit ist Exponentialverteilt.
Betrachte $[0,t]$, t fest, d.h. $\approx n \cdot t$ Experimente.
Dann gilt

$$X_t = \#\{\text{Erfolge in } [0,t]\} \approx \text{ Bin}(nt,p) = \text{ Bin}\left(nt, \frac{\lambda}{n}\right) \overset{n \text{ groß}}{\approx} \text{Pois}(ntp) = \text{Pois}(t\lambda)$$

X_t ist der Poisson-Prozess.
Die Erfolge in disjunkten Intervallen $[0,t)$, $[t, t+s]$ sind unabhängig, da die Experimente unabhängig sind.

Definition: 14.5

Betrachte Kette $\{X_t\}_{t\in\mathbb{N}_0}$ auf $(\Omega, \mathcal{A}, P)$
(a)
Eine Folge von $\{A_t\}_{t\in\mathbb{N}_0}$ von σ-Algebren über Ω heißt Filtration, wenn

$$\mathcal{A}_t \subseteq \mathcal{A}_s \subseteq \mathcal{A} \quad \forall\, t < s$$

(b)
$\{X_t\}_{t\in\mathbb{N}_0}$ heißt $\{\mathcal{A}_t\}_{t\in\mathbb{N}_0}$ adaptiert, wenn

$$X_t \text{ bzgl. } \mathcal{A}_t \text{ messbar ist für alle } t \in \mathbb{N}_0$$

(c)
$\{X_t\}_{t\in\mathbb{N}_0}$ heißt Martingal bezüglich der Filtration $\{\mathcal{A}_t\}_{t\in\mathbb{N}_0}$ wenn (b) gilt und

$$\begin{aligned} (i) \quad & E|X_t| < \infty \quad \forall\, t \in \mathbb{N}_0 \\ (ii) \quad & E(X_s|\mathcal{A}_t) = X_t \quad \forall\, s > t \end{aligned}$$

Gilt $\leq$ (bzw. $\geq$) in (ii) spricht man von einem Supermartingal (bzw. Submarginal)

Definition: 14.6

Betrachte $(\Omega, \mathcal{A}, P)$ mit einer Filtration $\{\mathcal{A}_t\}_{t\in\mathbb{N}_0}$.
Eine Zufallsvariable τ auf $(\Omega, \mathcal{A}, P)$ mit Werten in $\mathbb{N}_0 \cup \{\infty\}$ mit der Eigenschaft

$$\{w \in \Omega : \tau(w) = t\} \in \mathcal{A}_t \quad \forall\, t \in \mathbb{N}_0$$

heißt Stoppzeit bzgl. $\{\mathcal{A}_t\}_{t\in\mathbb{N}_0}$.
Ist $P(\tau = \infty) = 0$ heißt τ endliche Stoppzeit.

Theorem: 14.7 Version des Satzes vom optionalen Stoppen

Sei $(M_t)_{t\in\mathbb{N}_0}$ ein Martingal bzgl. der vom ihm erzeugten Filtration (d.h. $\mathcal{A}_t$ ist die von $M_0, M_1, ..., M_t$ erzeugte σ-Algebra) und τ eine beschränkte Stoppzeit bzgl. $\{\mathcal{A}_t\}_{t\in\mathbb{N}_0}$ (d.h. $P(\tau > t_s) = 0$ für ein $t_s \in \mathbb{N}_0$)
Dann gilt

$$EM_\tau = EM_0$$

Herleitung der Lundberg'schen Ungleichung (14.8)
Betrachte Modell (14.2) mit $X_k = a + \gamma k - \sum_{j=1}^{N(k)} Y_j$, setze

$$M_k := \exp(-sX_k) \quad (k \in \mathbb{N}_0) \text{ für ein } s > 0$$

Sei $\mathcal{A}_k$ die von $X_0, X_1, ..., X_k$ erzeugte σ-Algebra (Filtration), dann ist M_k $\{\mathcal{A}_k\}$-adaptierter Prozess.
Es gilt mit $g(s) := E(\exp(sY_1))$

$$\begin{aligned}
E|M_k| = EM_k &= E\Big\{\exp\Big(-s\Big[a+\gamma k-\sum_{j=1}^{N(k)} Y_j\Big]\Big)\Big\}\\
&= E\Big\{\exp\Big(-s[a+\gamma k]-\sum_{j=1}^{N(k)} sY_j\Big)\Big\}\\
&= \exp(-s[a+\gamma k])E\Big\{E\Big(\prod_{j=1}^{N(k)}\exp(sY_j)\Big|N(k)\Big)\Big\}\\
&= \exp(-s[a+\gamma k])\sum_{k=0}^{\infty}\underbrace{E\Big(\prod_{j=1}^{N(k)}\exp(sY_j)\Big)}_{\prod E\exp(sY_j)=g(s)^n}P(N(k)=n)\\
&= \exp(-s[a+\gamma k])E(g(s)^{N(k)}) < \infty \text{ falls } g(s)=E(\exp(sY_1))<\infty\\
&\overset{\sim\text{Pois}(\lambda k)}{=} \exp(-s[a+\gamma k])\sum_{j=0}^{\infty} g^j(s)\frac{(\lambda k)^j}{j!}\exp(-\lambda k)\\
&= \exp(-s[a+\gamma k]-\lambda k)\exp(g(s)\lambda k)\\
&= \exp(-s[a+\gamma k]+\lambda k[g(s)-1]) < \infty \text{ für } g(s)<\infty
\end{aligned}$$

Außerdem gilt

$$\begin{aligned}
E(M_{k+1}|\mathcal{A}_k) &= E(\exp(-sX_k)\exp(-s[X_{k+1}-X_k]|\mathcal{A}_k)\\
&= M_k \cdot E(\exp(-s[X_{k+1}-X_k])|\mathcal{A}_k)
\end{aligned}$$

mit

$$\begin{aligned}
X_{k+1}-X_k &= a+\gamma(k+1)-\sum_{j=1}^{N(k+1)} Y_j - a - \gamma k + \sum_{j=1}^{N(k)} Y_j\\
&= \gamma - \sum_{j=N(k)+1}^{N(k+1)} Y_j = \gamma - \sum_{j=1}^{N(k+1)-N(k)} Y_{N(k)+j}
\end{aligned}$$

Sei $\mathcal{A}'_k = \mathcal{A}_k \cup \sigma(N(k))$. Dann gilt fast sicher

$$\begin{aligned}
E(\exp(-s[X_{k+1}-X_k])|\mathcal{A}_k) &= \exp(-s\gamma)E\Bigg(E\Bigg\{1_{\mathbb{N}_0}(N(k))\exp\left[s\sum_{j=1}^{N(k+1)-N(k)} Y_{N(k)+j}\right]\Bigg|\mathcal{A}'_k\Bigg\}\Bigg|\mathcal{A}_k\Bigg)\\
&= \exp(-s\gamma)E\Bigg(E\Bigg\{\sum_{l=0}^{n} 1_{\{l\}}(N(k))\exp\left[s\sum_{j=1}^{N(k+1)-N(k)} Y_{l+j}\right]\Bigg|\mathcal{A}'_k\Bigg\}\Bigg|\mathcal{A}_k\Bigg)\\
&= \exp(-s\gamma)\sum_{l=0}^{\infty} E\Bigg(\underbrace{1_{\{l\}}(N(k))E\Bigg[\exp\Bigg(s\sum_{j=1}^{N(k+1)-N(k)} Y_{l+j}\Bigg)\Bigg|\mathcal{A}'_k\Bigg]}_{(*)}\Bigg|\mathcal{A}_k\Bigg)
\end{aligned}$$

Gilt $N(k) = l \Rightarrow \mathcal{A}'_k$ und die σ-Algebra die von den $N(k+1)-N(k)$ und Y_{l+j}, $j \geq 1$ erzegt werden, sind unabhängig. (Zukünftige Zuwächse und die Schadenshöhe von der Gegenwart k aus gesehen)
Also folgt

$$\begin{aligned}
(*) &= 1_{\{l\}}(N(k))E\Bigg[\exp\Bigg(s\sum_{j=1}^{N(k+1)-N(k)} Y_{l+j}\Bigg)\Bigg]\\
&= 1_{\{l\}}(N(k))E\Bigg[\prod \exp(sY_{l+j})\Bigg] \quad (Y_1, Y_2, \ldots \text{ iid})\\
&= 1_{\{l\}}(N(k))E(g(s)^{N(k+1)-N(k)})\\
&= 1_{\{l\}}(N(k))\sum_{j=0}^{\infty} g(s)^j \frac{\lambda^j}{j!}e^{-\lambda} = 1_{\{l\}}(N(k))e^{-\lambda}e^{\lambda g(s)}
\end{aligned}$$

Zusammengefasst gilt

$$\begin{aligned}
E(M_{k+1}|\mathcal{A}_k) &= M_k \cdot E(\exp(-s[X_{k+1}-X_k])|\mathcal{A}_k)\\
&= M_k \cdot \exp(-s\gamma)\sum_{l=0}^{\infty} E(1_{\{l\}}(N(k))e^{\lambda(g(s)-1)}|\mathcal{A}_k)\\
&= M_k \cdot \exp(\underbrace{\lambda g(s) - \lambda - s\gamma}_{=:G(s)}) \cdot \underbrace{\sum_{l=0}^{\infty} E(1_{\{l\}}(N(k))|\mathcal{A}_k)}_{=1}\\
&= M_k \cdot \exp(G(s))
\end{aligned}$$

Setze $W_k := \exp(-kG(s))M_k$.
$\{W_k\}_k$ ist eine $\{\mathcal{A}_k\}_k$-adaptierte Kette mit $E|W_k| < \infty$ (da $E|M_k| < \infty$) $\forall\, k \in \mathbb{N}_0$. Dann

$$E(W_{k+1}|\mathcal{A}_k) = \exp(-(k+1)G(s))\underbrace{E(M_{k+1}|\mathcal{A}_k)}_{M_k\exp(G(s))} = \exp(-kG(s))M_k = W_k$$

Also gilt

$$\begin{aligned} E(W_{k+m}|\mathcal{A}_k) &\overset{\text{eben}}{=} E(\underbrace{E(W_{k+m}|\mathcal{A}_{k+m-1})}_{W_{k+m-1}}|\mathcal{A}_k) \\ &= E(W_{k+m-1}|\mathcal{A}_k) \text{ fast sicher} \\ &= E(W_{k+m-2}|\mathcal{A}_k) = ... \\ &= E(W_k|\mathcal{A}_k) = W_k \end{aligned}$$

$\Rightarrow W_k$ Martingal bzgl. $\{\mathcal{A}_k\}_k$
Sei t_s fest, $\tau_a = \inf\{k : X_k < 0\}$, setze $\tau_a' = \min\{\tau_a, t_s\}$ (festgelegter Zeitpunkt $X_k < 0$ oder t_s). Dies ist eine beschränkte Stoppzeit, denn

$$\{w : \tau_a' > k\} = \begin{cases} \emptyset \in \mathcal{A}_k & k \geq t_s \\ \bigcap_{j \leq k} \{X_j \geq 0\} & k < t_s \end{cases} \in \mathcal{A}_k$$

Damit gilt:

$$\{w \in \Omega : \tau_a'(w) = k\} = \underbrace{\{w : \tau_a' > k-1\}}_{\in \mathcal{A}_{k-1} \subseteq \mathcal{A}_k} \setminus \underbrace{\{w : \tau_a' > k\}}_{\in \mathcal{A}_k}$$

Mit dem Satz von optionalen Stopping folgt

$$\begin{aligned} \exp(-sa) = EW_0 (&\overset{\text{Def.}}{=} E[\exp(-0 \cdot G(s))M_0] = EM_0) \\ &\overset{\text{Satz}}{=} EW_{\tau_a'} \geq E(W_{\tau_a'} \cdot 1_{[0,t_s]}(\tau_a)) = E(W_{\tau_a} \cdot 1_{[0,t_s]}(\tau_a)) \geq P(\tau_a \leq t_s) \end{aligned}$$

<u>falls</u>: $W_{\tau_a} \geq 1$: mit

$$W_{\tau_a} = \underbrace{e^{-\tau_a G(s)}}_{\geq 1 \text{ für } G(s)<0} M_{\tau_a} \geq M_{\tau_a} = e^{-sX_{\tau_a}} \geq 1$$

($X_{\tau_a} \leq 0$ ab Zeitpunkt τ_a) $\Rightarrow$ 1. Forderung: $G(s) < 0$ (F1)
Wegen $\{\tau_a < \infty\} = \bigcup\limits_{t_s \in \mathbb{N}_0} \{\tau_a \leq t_s\}$ gilt außerdem

$$P(\tau_a < \infty) = \lim_{t_s \to \infty} P(\tau_a \leq t_s) \overset{\text{oben}}{\leq} \exp(-sa)$$

wobei $\exp(-sa)$ möglichst klein sein soll! $\Rightarrow$ 2. Forderung: wähle $s = s^*$ möglichst groß (F2)
(F1) + (F2) $\Rightarrow s^* = \sup\{s > 0 : G(s) \leq 0\}$

Theorem: 14.9 Lundberg'sche Ungleichung

Im klassischen Versicherungsmodell (14.2) mit $G(s) := \lambda E e^{-sY_1} - \lambda - s\gamma$ ($N(k) \sim \text{Pois}(\lambda k)$) ist die Wahrscheinlichkeit $P(\tau_a = \infty)$ dass eine Versicherung bei Startkapital a <u>nie</u> in die roten Zahlen kommt nach unten beschränkt durch

$$1 - \exp(-s^* a) \text{ mit } s^* = \sup\{s : G(s) \leq 0\} \text{ 'Lundberg Exponent'}$$

Kapitel 15

Markov-Ketten

Betrachte Stochastischen Prozess $\{X_n\}_{n\in\mathbb{N}_0}$ der abzählbar viele Werte annimmt, im Allgemeinen $\mathbb{N}_0$.
'$X_n = i$' heißt, der Prozess ist zu Zeitpunkt n im Zustand i

Definition: 15.1

Ein Prozess $\{X_n\}_{n\in\mathbb{N}_0}$ heißt Markov-Kette, falls er die Markov-Eigenschaft erfüllt, d.h.

$$p_{ij} := P(X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, ..., X_0 = i) = P(\underbrace{X_{n+1} = j}_{\text{Zukunft}} | \underbrace{X_n = i}_{\text{Gegenwart}})$$

für alle Zustände $j, i, i_0, i_1, ..., i_{n-1}$ und $n \geq 0$.

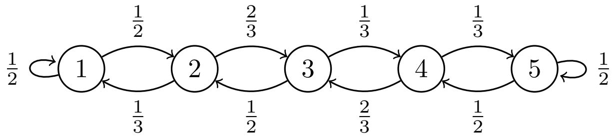
Korollar: 15.2

Es muss gelten $p_{ij} \geq 0 \ \forall \ i, j$ (da Wahrscheinlichkeit ≥ 0), $\sum_{j=0}^{\infty} p_{ij} = 1$, $i \in \mathbb{N}_0$ (man muss vom Zustand i irgendeinen Zustand erreichen)
Übergangsmatrix

$$P = \begin{pmatrix} p_{00} & p_{01} & \cdots & \cdots & \cdots \\ p_{10} & p_{11} & \cdots & \cdots & \cdots \\ \vdots & \vdots & & & \\ p_{i0} & p_{i1} & & & \\ \vdots & \vdots & & & \end{pmatrix}$$

Bsp

$$\begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\ \frac{1}{3} & 0 & \frac{2}{3} & 0 & 0 \\ 0 & \frac{1}{3} & 0 & \frac{2}{3} & 0 \\ 0 & 0 & \frac{2}{3} & 0 & \frac{1}{3} \\ 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

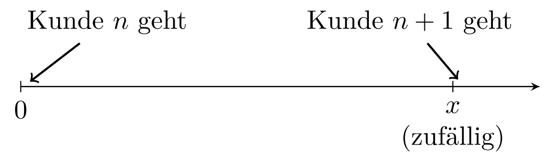


Beispiel (15.3 Warteschlange)
Ankunft von Kunden in Kundenzentrum
Modellierung mit Poisson-Prozess mit Intensität λ.
1 Mitarbeiter, 1. Kunde wird sofort bedient, der Rest kommt in die Warteschlange.
Annahme: Servicezeiten der Kunden sind iid $\sim G$
Die Zeitdauern zwischen den Ankünften sei exponentialverteilt, da es sich um einen Poisson-Prozess handelt (vgl. 14.4).
Sei $\tilde{N}(t) = \#\{\text{Anzahl Kunden zum Zeitpunkt } t\}$
$\tilde{N}(t)$ ist keine Markov-Kette, denn die bedingte Verteilung von $\tilde{N}(t+1)$ hängt nicht nur von $\tilde{N}(t)$, sondern auch von den Servicedauern ($\sim G$ unbekannt) ab.
Ausweg: Betrachte nur Zeitpunkte, an denen die Kunden fertig sind, d.h.
$X_n = \#\{\text{Kunden nachdem der } n\text{-te Kunde gegangen ist}\}$ $(n \geq 1)$
$X_n > 0$: es gibt noch X_n Kunden, einer wird bedient. $X_n - 1$ in Warteschlange
'$X_n = 0$': keine Kunden mehr da nachdem n-ter Kunde gegangen ist
d.h. wenn der nächste Kunde geht gibt es $X_n - 1 + Y_n$ Kunden (Y_n Neuzugänge; Kunden die in der Zeit gekommen sind, in der Kunde $n+1$ bedient wurde)

$$\Rightarrow X_{n+1} = \begin{cases} X_n - 1 + Y_n & X_n > 0 \\ Y_n & X_n = 0 \end{cases} \qquad (*)$$

$\{Y_n\}_{\mathbb{N}}$ ist Poisson-Prozess, nämlich $\#\{\text{ankommende Kunden in disjunkten Zeitintervallen}\}$, d.h. die Y_n sind unabhängig und

$$\begin{aligned} P(Y_n = j) &= E(P(Y_n = j) | \text{Servicezeit}) \\ &= \int_0^\infty e^{-\lambda x} \frac{(\lambda x)^j}{j!} dG(x) \quad j = 0, 1, ... \end{aligned}$$



Damit und mit $(*)$ folgt $\{X_n\}_{n\in\mathbb{N}}$ ist eine Markov-Kette mit Übergangsverteilung
Nebenrechnung

$$P(X_{n+1} = j | X_n = i) = P(X_n - 1 + Y_n = j | X_n = i) = P(Y_n = j - i + 1)$$

Nebenrechnung vorbei

$$
\begin{aligned}
p_{0j} &= \int_0^\infty e^{-\lambda x} \frac{(\lambda x)^j}{j!} dG(x) \quad j \in \mathbb{N}_0 \\
p_{ij} &= \int_0^\infty e^{-\lambda x} \frac{(\lambda x)^{j-i+1}}{(j-i+1)!} dG(x) \quad j-i+1 \geq 0 \Leftrightarrow j \geq i-1, i \geq 1 \\
p_{ij} &= 0 \text{ sonst}
\end{aligned}
$$

Beispiel (15.4 Der allgemeine Random Walk (Irrfahrt))
Seien $X_1, X_2, ...,$ iid mit

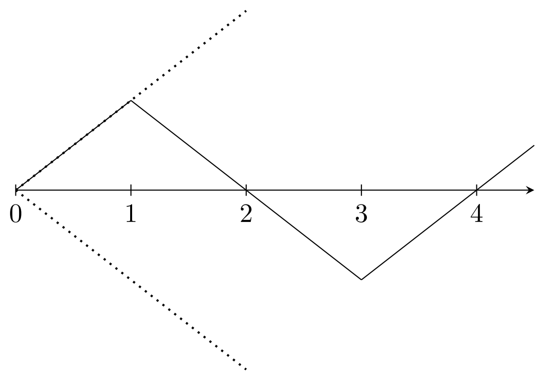
$$P(X_i = j) = a_j, \quad j = 0, \pm 1, \pm 2, ...$$

Sei

$$S_0 = 0 \quad S_n = \sum_{i=1}^n X_i$$

dann gilt S_n ist eine Markov-Kette mit $P_{ij} = a_{j-i}$ $(S_n = i \to S_{n+j} = j)$
Spezialfall: Einfacher Random Walk

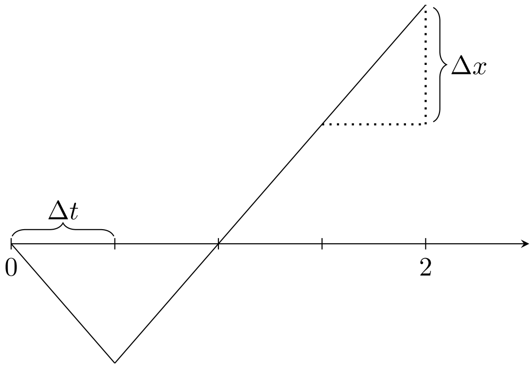
$$S_n = \sum_{i=1}^n X_n \text{ mit } P(X_i = 1) = p \text{ und } P(X_i = -1) = \underbrace{1-p}_{q} \quad 0 < p < 1$$



Kapitel 16

Brownsche Bewegung ('Wiener Prozess')

Einführung (16.1 Herleitung)
Betrachte einen symmetrischen Random Walk $X_1, X_2, ...$ (hoch/ runter mit gleicher Länge),
Δt Zeiteinheit
Δx Schrittlänge (Sprunghöhe)



$X(t)$ sei Position zum Zeitpunkt t, d.h. $X(t) = \Delta x \cdot (X_1 + X_2 + ... + X_{\frac{t}{\Delta t}})$ mit

$$X_i = \begin{cases} +1 & \text{Schritt der Länge } x \text{ nach oben} \\ -1 & \text{Schritt der Länge } x \text{ nach unten} \end{cases}$$

X_i's unabhängig mit $P(X_i = 1) = P(X_i = -1) = \frac{1}{2}$, also $EX_i = 0$, $Var X_i = EX_i^2 = 1 \cdot P(X_i^2 = 1) = 1$

$$\Rightarrow \begin{cases} E(X(t)) = 0 \\ Var(X(t)) \quad = E((X(t))^2) = (\Delta x)^2 \sum_{i=1}^{\frac{t}{\Delta t}} Var(X_i) = (\Delta x)^2 \frac{t}{\Delta t} \end{cases}$$

Grenzübergang $\Delta x \to 0$, $\Delta t \to 0$ derart sodass Grenzprozess nicht trivial (als z.B. nicht beide gleich schnell gegen 0 gehen lassen wie $\Delta x := \Delta t$, da sonst $X(t) = 0$)
Betrachte $\Delta x := c\sqrt{\Delta t}$ ($c > 0$ Konstant)
Dann gilt mit $\Delta t \to 0$:

$$E(X(t)) = 0$$
$$Var(X(t)) = (\Delta x)^2 \frac{t}{\Delta t} = c^2 \Delta t \frac{t}{\Delta t} = c^2 t$$

Eigenschaften von $X(t) = \Delta x(X_1 + X_2 + ... + X_{\frac{t}{\Delta t}})$ wenn $\Delta x = c\sqrt{\Delta t} \to 0$
(i)
$X(t) \sim N(0, c^2 t)$ (ZGW)
Begründung: Betrachte t fest, $\frac{t}{\Delta t} = n$ Teilintervalle

$$\begin{aligned}\Delta t &= \frac{t}{n} \overset{n\to\infty}{\longrightarrow} 0 \\ X(t) &= c\sqrt{\Delta t}(X_1 + X_2 + ... + X_n) \\ &= c\sqrt{t}n^{-\frac{1}{2}} \underbrace{\sum_{i=1}^{n} X_i}_{\substack{EX_i=0,\\ VarX_i=1}} \overset{\text{D}}{\to} c\sqrt{t}N(0,1) = N(0, c^2 t)\end{aligned}$$

(ii)
$\{X(t), t \geq 0\}$ hat unabhängige Zuwächse (Änderungen des Random Walk in disjunkten Zeitintervallen sind unabhängig)
Die Änderung der Position des Random Walks in jedem Zeitintervall hängt nur von der Intervalllänge ab (nicht der Position), also
(iii)
$\{X_t, t \geq 0\}$ hat stationäre Zuwächse, d.h. $X_{t+s} - X_t$ haben für jedes t dieselbe Verteilung (nur vom Abstand s abhängig)

Definition: 16.2

Ein stochastischer Prozess $(X(t))_{t\geq 0}$ heißt 'Brownscher Bewegungs-Prozess' (oder 'Brownsche Bewegung oder 'Wiener Prozess') falls gilt
(i) $X(0) = 0$
(ii) $(X(t))_{t\geq 0}$ hat unabhängige stationäre Zuwächse
(iii) Für jedes $t > 0$ gilt $X(t) \sim N(0, c^2 t)$
(Modell Partikelbewegung, entdeckt von Robert Brown, 1. Erklärung (physikalisch) durch Einstein 1905, mathematische Definition Norbert Wiener 1918)
Im Folgenden sei $c = 1$ 'standard Browninan motion'

Eigenschaften (16.3)
(a)
$X(t)$ ist P-fast sicher stetige Funktion in t ('stetige Pfade')
(b)
$X(t)$ ist P-fast sicher nirgends differenzierbar
(c)
$X(t)$ ist ein Markov-Prozess
Unabhängige Zuwächse $\Rightarrow$ Positionswechsel $X(t+s) - X(s)$ zwischen Zeitpunkt s und $t+s$ ist unabhängig

von allen Werten vor dem Zeitpunkt s. Damit gilt

$$\begin{aligned}
&P(\underbrace{X(t+s)}_{\text{Zukunft}} \leq a| \underbrace{X(s)=x}_{\text{Gegenawart}}, \underbrace{X(u), 0 \leq u}_{\text{Vergangenheit}} < s) \\
=&P(X(t+s) - X(s) \leq a - X(s)|X(s) = x, X(u), 0 \leq u < s) \\
=&P(X(t+s) - X(s) \leq a - x) \\
=&P(X(t+s) \leq a|X(s) = x)
\end{aligned}$$

Verteilung von $(X(t))_{t\geq 0}$ (16.4)
$X(t)$ hat unabhängig stationäre Zuwächse (nur vom Abstand abhängig)
Nebenbemerkung
Im diskreten Fall

$$P(X_1 = x_1, X_2 = x_2, ..., X_n = x_n) = P(X_1 = x_1, X_2 = x_2 - x_1, ..., X_n - X_{n-1} = x_n - x_{n-1})$$

Nebenbemerkung vorbei.
Damit hat $X(t_1), .., X(t_n)$ die gemeinsame Dichte

$$f(x_1, ..., x_n) = f_{t_1}(x_1) f_{t_2 - t_1}(x_2 - x_1) \cdot ... \cdot f_{t_n - t_{n-1}}(x_n - x_{n-1})$$

und die bedingte Verteilung von $X(s)$ gegeben $X(t) = B$ ist $(s < t)$:

$$\begin{aligned}
f_{s|t}(x|B) = \frac{f_s(x) f_{t-s}(B-x)}{f_t(B)} &= K_1 \exp\left(-\frac{x^2}{2s} - \frac{(B-x)^2}{2(t-s)} \right) \\
&= K_1 \exp\left(\frac{-x^2(t-s) - s(B-x)^2}{2s(t-s)} \right) \\
&= K_1 \exp\left(\frac{-t(x - \frac{s}{t}B)^2 - B^2(\frac{s}{t} - \frac{s^2}{t^2})}{2s(t-s)} \right) \\
&= K_2 \exp\left(-\frac{(x - \frac{s}{t}B)^2}{2s(t-s)/t} \right) \text{ Normalverteilung}
\end{aligned}$$

$$\begin{aligned}
&\Rightarrow E(X(s)|X(t) = B) = B\frac{s}{t}, \quad (E(X(s)|X(t)) = X(t)\frac{s}{t}) \\
&Var(X(s)|X(t) = B) = \frac{s(t-s)}{t} = \frac{s}{t}\frac{t-s}{t}t \text{ (nicht von } B \text{ abhängig)}
\end{aligned}$$

Gilt z.B. $\frac{s}{t} = \alpha \in (0,1) \Rightarrow X(s)|X(t) \sim N(\alpha X(t), \alpha(1-\alpha)t)$
Aus der Dichteformel für $X(t_1), ..., X(t_n)$ foglt auch, dass die gemeinsame Verteilung eine multivariate Normalverteilung, d.h. der Prozess ist ein Gauß-Prozess.

Definition: 16.5

Ein stochastischer Prozess heißt Gauß-Prozess, wenn die Verteilung von $X(t_1), ..., X(t_n)$ für alle $t_1, ..., t_n$ eine multivariate Normalverteilung ist.

Bemerkung (16.6 Alternative Definition der Brownschen Bewegung)
Vorbemerkung: Eine multivariate Normalverteilung ist vollständig durch die Erwartungswerte der Randverteilungen und die Kovarianzen bestimmt.

Alternative Definition: Ein Brownscher Bewegungsprozess ist ein Gaußprozess mit Erwartungswert $E(X(t)) = 0$ und $Cov(X(s), X(t)) = s \; \forall \; s < t$
Denn

$$\begin{aligned} Cov(X(s), X(t)) &= Cov(X(s), X(s) + X(t) - X(s)) \\ &= Cov(X(s), X(s)) + \underbrace{Cov(X(s), \underbrace{X(t) - X(s)}_{\text{unab. Zuwächse}})}_{0} \\ &= Cov(X(s), X(s)) = Var(X(s)) = s \quad (c^2 t \text{ mit } c = 1, t = s) \end{aligned}$$

Herleitung der Brownschen Brücke (16.7)
Sei $X(t)$ Brownsche Bewegung, betrachte

$$\{X(t), 0 \leq t \leq 1 | X(1) = 0\}$$

wie in 16.4 zeigt man: dies ist ein Gauß-Prozess.
Dort ebenfalls gezeigt:

$$E(X(s)|X(1) = 0) = 0 \cdot \frac{s}{1} = 0 \; \forall \; s < 1$$

Für $s \leq t \leq 1$ gilt:

$$\begin{aligned} Cov[(X(s), X(t))|X(1) = 0] &= E[X(s)X(t)|X(1) = 0] \\ &= E[E\{X(s)X(t)|X(t), X(1) = 0\}|X(1) = 0] \\ &= E[X(t) \underbrace{E\{X(s)|X(t)\}}_{X(t)\frac{s}{t}} |X(1) = 0] \\ &= \frac{s}{t} E(X^2(t)|X(1) = 0) \\ &= \frac{s}{t} Var(X(t)|X(1) = 0) \\ &= \frac{s}{t} t(1 - t) = s(1 - t) \end{aligned}$$

Theorem: 16.8

Sei $X(t)_{t \geq 0}$ die Brownsche Bewegung, $Z(t) = X(t) - tX(1)$, dann ist $(Z(t))_{0 \leq t \leq 1}$ die Brownsche Brücke.

Kapitel 17

Asymptotische Statistik

Definition: 17.1

$X, X_1, X_2, ..., X_n$ seien iid Zufallsvariablen mit $EX = \mu$, $VarX = \sigma^2 \in (0, \infty)$ und Verteilungsfunktion F.
Die empirische Verteilungsfunktion ist

$$\hat{F}(t) = \frac{1}{n}\sum_{i=1}^{n} 1(X_i \leq t) = \frac{\#\{X_i \leq t\}}{n}, \quad t \in (-\infty, \infty)$$

Der empirische Schätzer für $Eg(x)$ ist

$$\int g(x)d\hat{F}(x) = \frac{1}{n}\sum_{i=1}^{n} g(X_i)$$

(Spezialfall $g(x) = 1_{(-\infty,t]}(x)$ empirische Verteilungsfunktion)

Bemerkung (17.2 Eigenschaften von empirischen Schätzern)
(a)

$$\frac{1}{n}\sum_{i=1}^{n} g(X_i) \overset{n\to\infty}{\longrightarrow} Eg(X) \text{ fast sicher (SLLN)}$$

(b)

$$\sqrt{n}\left(\frac{1}{n}\sum g(X_i) - Eg(X_i)\right) \xrightarrow[n\to\infty]{D} N(0, Var\ g(x)) \text{ (ZGW)}$$

Bemerkung (17.3 Eigenschaften von der empirischen Verteilungsfunktion)
(a)
$\hat{F}(t)$ ist Zufallsvariable (für t fest)
(b)
$\hat{F}(.)$ ist eine zufällige (Verteilungs-)Funktion, $(\hat{F}(t))_{t\in\mathbb{R}}$ also stochastischer Prozess

(c)
$\hat{F}(t)$ ist erwartungstreu für $F(t)$

$$E\hat{F}(t) = \frac{1}{n}\sum \underbrace{E(1(X_i \leq t))}_{F(t)} = F(t)$$

(d)

$$n\hat{F}(t) = \sum_{i=1}^{n} 1(X_i \leq t) \sim \text{ Bin}(n, F(t))$$

(e)
Vgl. 17.2(b):

$$\hat{F}(t) = \frac{1}{n}\sum_{i=1}^{n} 1(X_i \leq t) \approx N\Big(F(t), \frac{F(t)(1-F(t))}{n}\Big)$$

$$\sqrt{n}(\hat{F}(t) - F(t)) \xrightarrow{\text{D}} N(0, F(t)(1-F(t)))$$

Theorem: 17.4 Gliwenko-Cantelli

$\hat{F}(x) = \hat{F}_n(x)$ ist stark konsistent gleichmäßig in x, d.h.

$$D_n := \sup_{t\in\mathbb{R}} |\hat{F}_n(t) - F(t)| \overset{n\to\infty}{\longrightarrow} 0 \text{ fast sicher}$$

D_n heißt 'Kolmogorov-Smirnov Abstand'.

Lemma: 17.5

Gilt $X_n \xrightarrow{\text{P}} X$ hinreichend schnell (sufficiently fast), d.h.

$$\sum_{n=1}^{\infty} P(|X_n - X| > \varepsilon) < \infty \quad \forall\, \varepsilon > 0$$

dann folgt

$$X_n \to X \text{ fast sicher}$$

Theorem: 17.6

Betrachte $H_0 : F = F_0$, $H_1 : F \neq F_0$, F, F_0 stetig
Dann ist die Kolmogorov-Smirnov-Teststatistik

$$T = \sup_{x\in\mathbb{R}} |\hat{F}(x) - F_0(x)|$$

unter H_0 verteilungsfrei, d.h. die Verteilung hängt nicht von F_0 ab.

Bemerkung (17.7)
Unter $H_0 : F = F_0$ hat $\sqrt{n}T$ für stetig F dieselbe Verteilung wie

$$S_n = \sup_{x\in[0,1]} \left| n^{-\frac{1}{2}} \sum_{j=1}^{n} 1_{[0,x]}(U_j) - x \right|$$

$U_1, ..., U_n \overset{iid}{\sim} U([0,1])$. Man kann zeigen:

$$S_n \overset{\mathrm{D}}{\to} S \sim F(x) = 1 - 2\sum_{k=1}^{\infty} (-1)^{k-1} e^{-2k^2x^2},\ x \geq 0$$

($\Rightarrow$ kritische Werte berechenbar)

Bemerkung (17.8 $\hat{F}$ und die Brownsche Brücke)
Betrachte $X_1, X_2, ...$ iid
1. Fall: $X_i \sim U(0,1)$

$$\alpha_n(s) := \sqrt{n}(\hat{F}(s) - s) = \sqrt{n}\left(\frac{1}{n}\sum_{j=1}^{n} 1_{(-\infty,s]}(X_j) - F_{U(0,1)}(s)\right)$$

Ziel: Asymptotische Eigenschaften von $\{\alpha_n(s)\}_{s\in[0,1]}$

$$N(t) := n\hat{F}(t) \sim \mathrm{Bin}(n,t), \quad N(s) \sim \mathrm{Bin}(n,s)$$

Sei $s < t$. Die bedingte Wahrscheinlichkeit von $N(t) - N(s)$ gegeben $N(s) = n_s$ ist $\mathrm{Bin}(n - n_s, \frac{t-s}{1-s})$-verteilt (Ü).
Man kann zeigen:
- $(\alpha_n(s), \alpha_n(t))$ ist asymptotisch normalverteilt
- der Grenzprozess ist ein Gauß-Prozess
Zu bestimmen: $E(\alpha_n(s)), Cov(\alpha_n(s), \alpha_n(t))$

$$E(\alpha_n(s)) = 0 \text{ (klar)}$$

Sei $0 \leq s \leq t \leq 1$

$$\begin{aligned}
Cov(\alpha_n(s), \alpha_n(t)) &= n(Cov(\hat{F}(s), \hat{F}(t)) \\
&= \frac{1}{n} Cov(N(s), N(t)) \\
&= \frac{E[E(N(s)N(t)|N(s))] - \overbrace{E(N(s))}^{ns}\overbrace{E(N(t))}^{nt}}{n} \\
&= n^{-1}\{E[N(s)E(N(t)|N(s))] - n^2st\} \\
&= n^{-1}\{E[N(s)E(N(t) - N(s) + N(s)|N(s))] - n^2st\} \\
&= n^{-1}\{E\left[N(s)(N(s) + (n - N(s))\frac{t-s}{1-s})\right] - n^2st\} \\
&= [\text{mit } N(s) \sim \mathrm{Bin}(n,s)]... \\
&= s(1-t)
\end{aligned}$$

Der Grenzprozess ist also ein Gauß-Prozess mit Erwartungswert $E\alpha_n(s) = 0 \; \forall s \in [0,1]$ und $Cov(\alpha_n(s), \alpha_n(t)) = s(1-t)$ $(0 \leq s \leq t \leq 1)$ also eine Brownsche-Brücke.
2. Fall $X_i \sim F$ beliebige stetige Verteilungsfunktion.
Gesucht: Asymptotische Verteilung von $\sqrt{n} \sup_x |\hat{F}(x) - F(x)|$
Es gilt

$$P(F(X_i) \leq x) = P(X_i \leq F^{-1}(x)) = F(F^{-1}(x)) = x$$

d.h. $F(X_i) \sim U(0,1)$.
Betrachte $\alpha_n(s)$ aus Fall 1 mit $F(X_i) \sim U(0,1)$.
Auf Fall 1 folgt:

$$\begin{aligned}
\alpha_n(s) &= \sqrt{n}\left[\frac{1}{n}\#\{X_i, i = 1, ..., n : \underbrace{F(X_i)}_{\sim U(0,1)} \leq s\} - s\right] \\
&= \sqrt{n}\left[\frac{1}{n}\sum_{i=1}^{n} 1\{X_i \leq F^{-1}(s)\} - s\right] \\
&= \sqrt{n}[\hat{F}(F^{-1}(s)) - s] \\
&= \sqrt{n}[\hat{F}(\underbrace{F^{-1}(s)}_{=:y_s}) - F(F^{-1}(s))] \\
&= \sqrt{n}(\hat{F}(y_s) - F(y_s))
\end{aligned}$$

Der Prozess $\{\alpha_n(s)\}_{s\in[0,1]}$ konvergiert gegen eine Brownsche Brücke nach Fall 1, die asymptotische Verteilung von $\sqrt{n} \sup_x |\hat{F}(x) - F(x)|$ ist das Supremum (oder Maximum da stetig) der Brownschen Brücke, also

$$\lim_{n\to\infty} P\left(\sqrt{n} \sup_x |\hat{F}(x) - F(x)| < a\right) = P\left(\max_{0\leq t\leq 1} |Z(t)| < a\right)$$

mit $\{Z(t)\}_{t\geq 0}$ der Brownsche Brücken Prozess.