Studierendenmitschrift zur Vorlesung: Einführung in die Inferenzstatistik (SoSe 2022)

Original SoSe21 hergestellt von: Jim Feller
Aktualisierung SoSe22: Jelena Popadic

The template is based on a free template from Typesetters.
The Book icon is a free icon from Icons8.

Inhaltsverzeichnis

1. Parametrische Modelle

Siehe für eine Einführung zu diesem Kapitel S. 255-260 in *Rice (2007)*.

Beispiel 1.1 (Ein Modell für den radioaktiven Zerfall) Atome eines Elements zerfallen unabhängig voneinander in scheinbar zufälliger Reihenfolge. Im Moment des Zerfalls wird ein α-Teilchen abgegeben. Im Experiment beobachtet man eine enorm große Zahl von Atomen über eine gewisse Zeitspanne. Dabei wird die (relativ kleine) Zahl der α-Teilchen beobachtet.
$S_{n_0} = \#\alpha$-Teilchen pro Zeitintervall

Modell 1

$S_{n_0} \sim B(n_0, p_0)$ ← unpraktisch zum Rechnen (n_0 unbekannt)

Atome (enorm) — Zerfallswkeit (extrem klein)

Modell 2

$\sqrt{n_0}\left(\frac{1}{n_0}S_{n_0} - p_0\right) \xrightarrow[n_0 \to \infty,\ p_0 \text{ fest}]{w} N\left(0, p_o(1-p_0)\right)$ ← unsinnig.

$\mathbb{E}(S_{n_0}) = n_0 p_0 \longrightarrow \infty$ ← passt nicht zu der Tatsache, dass im Experiment eine relativ kleine Anzahl von α-Teilchen, also ein relativ kleiner Wert von S_{n_0}, beobachtet wird.

Modell 3

$S_{n_0} \sim P(\underbrace{\lambda_0}_{n_0 p_0})$ ← praktikabel und sinnvoll

sehr groß — sehr klein, sodass λ_0 moderat

Begründung: Gesetz der kleinen Zahlen
Damit lässt sich Modell 1 durch Modell 3 approximmieren
Sei $X \sim B(n,p)$. Dann gilt für jedes $k \in \mathbb{N}_0$:

$$\mathbb{P}(X = k) \xrightarrow[n \to \infty,\ np = \lambda \text{ fest}]{} \frac{\lambda^k}{k!} e^{-\lambda}.$$

Nachrechnen.

$$\begin{aligned}
\mathbb{P}(X=k) &= \binom{n}{k} p^k (1-p)^{n-k} \\
&= \frac{\overbrace{n(n-1)\cdot\ldots\cdot(n-k+1)}^{k \text{ Faktoren}}}{k!} \underbrace{p^k}_{k \text{ Faktoren}} (1-p)^{n-k} \\
&= \frac{np(n-1)p\cdot\ldots\cdot(n-k+1)p}{k!}(1-p)^n(1-p)^{-k} \\
&= \frac{np(n-1)p\cdot\ldots\cdot(n-k+1)p}{k!}\left(1-\frac{np}{n}\right)^n\left(1-\frac{np}{n}\right)^{-k} \\
&= \frac{np\cdot np\cdot\left(1-\frac{1}{n}\right)\cdot\ldots\cdot np\cdot\left(1-\frac{k+1}{n}\right)}{k!}\left(1-\frac{np}{n}\right)^n\left(1-\frac{np}{n}\right)^{-k} \\
&\underset{\substack{n\to\infty \\ np=\lambda}}{\longrightarrow} \frac{\lambda^k}{k!}e^{-\lambda}1.
\end{aligned}$$

Bemerke:

$np=\lambda$,
$\left(1-\frac{1}{n}\right)\to 1$,
$e^{-x}=\lim_{n\to\infty}\left(1-\frac{x}{n}\right)^n$,
$\left(1-\frac{pn}{n}\right)^n \approx \left(1-\frac{\lambda}{n}\right)^n \approx e^{-\lambda}$.

□

Beispiel 1.2 (Daten von Berkson (1966)) Daten übernommen aus *Rice (2007, S. 256)*.
#α-Teilchen von Americum 241 innerhalb von 10 Sekunden.
Dieses Experiment wird 1207 mal (unabhängig) wiederholt.
$n=1207$ Beobachtung $X_1,\ldots,X_n$ *iid mit* $X_i \sim P(\lambda_0)$.
Im Poisson Modell X_i *iid* $P(\lambda_0)$ ist $\mathbb{E}(X_i)=\lambda_0 \operatorname{Var}(X_i)=\lambda_0$
Für $\bar{X}_n=\sum_{i=1}^n(X_i)$ ist $E(\bar{X}_n)=\lambda_0$, $\operatorname{Var}(\bar{X}_n)=\frac{\lambda_0}{n}$.

Schätzer für $\mathbb{E}(X_i)=\lambda_0:\ \ \hat{\lambda}_n=\bar{X}_n \leftarrow$ unverzerrte und konsistente Schätzer
Schätzer für $Var(X_i)=\lambda_0:\ \bar{X}_n$ oder $\frac{1}{n-1}\sum_{i=1}^n(X_i-\hat{\mu})^2=\hat{\sigma}_n^2$

Man findet aus den Daten: $\hat{\lambda}_n=8.37, (\hat{\sigma}_n)^2=8.37, \tilde{\sigma}_n^2=8.53$.

Konfidenzintervall für λ_0 mit nominaler Überdeckungswahrscheinlihkeit $1-\alpha$:

- Normalapproximation:

$$\hat{\lambda}_n \pm \sqrt{\frac{\hat{\sigma}_n^2}{n}}\Phi^{-1}\left(1-\frac{\alpha}{2}\right)\ldots \text{für } \alpha=0.05:[8.206, 8.533]$$

- Bootstrap:

Population	"Bootstrap Population"
$P(\lambda_0)$ Mittelwert λ_0	$X_1,\dots,X_n$ Mittelwert $\hat{\lambda}_n$
$\downarrow$	$\downarrow$
$X_1,\dots,X_n$ i.i.d. $P(\lambda_0)$	$X_1^*,\dots,X_n^*$ (ziehen mit Zurücklegen)
$\downarrow$	$\downarrow$
$\bar{X} = \hat{\lambda}_n$	$\bar{X}^* = \hat{\lambda}_n{}^*$
Fehler $\hat{\lambda} - \lambda_0$	Fehler $\hat{\lambda}^* - \hat{\lambda}$
Wiederhole (u.a.), m-mal: $\hat{\lambda}^{(1)} - \lambda_0, \dots, \hat{\lambda}^{(m)} - \lambda_0$	Wiederhole (u.a.), m-mal: $\hat{\lambda}^{*(1)} - \hat{\lambda}, \dots, \hat{\lambda}^{*(m)} - \hat{\lambda}$
Schätze die Verteilungsfunktion des Fehlers wie folgt: $\mathrm{F}(t) = \mathbb{P}(\hat{\lambda}_n - \lambda_0 \leq t)$ $\hat{F}_M(t) = \frac{1}{m}\sum_{i=1}^m \mathbb{1}\{\hat{\lambda}_n{}^{(\cdot)} - \lambda_0 \leqslant t\}$	Schätze die Verteilungsfunktion des Fehlers wie folgt: $\mathrm{F}(t)$ $\hat{F}_M^{*m}(t) = \frac{1}{m}\sum_{i=1}^m \mathbb{1}\{\hat{\lambda}_n{}^* - \hat{\lambda}_n \leqslant t\}$

95% Konfidenzintervall mit der Bootstrap-Methode:

$$\left[\hat{\lambda}_n - \hat{F}_n{}^{*m-1}(1-\frac{\alpha}{2}), \hat{\lambda}_n - \hat{F}_n{}^{*m-1}(\frac{\alpha}{2})\right] =$$

$$[8.205,\ 8.531]$$

1.1 Parametrische Modelle und Parameterschätzung

- Experiment:

Beobachtungen $X_1,\dots,X_n$

- Parametrisches Modell:

Angenommen die $X_1,\dots,X_n$ sind i.i.d. Beobachtung einer Zufallsvariable X, deren Verteilung von einer bestimmten Form ist:

$$X \sim \text{ Verteilung abhängig von einem Parameter } \theta_0 \in \Theta \subseteq \mathbb{R}^p$$

Die funktionale Form des Modells ist <u>bekannt</u>, der Parameter θ_0 ist <u>unbekannt</u>!

Beispiel 1.3 Beispiele von parametrischen Modellen:

$X \sim P(\lambda_0)$	$\lambda_0 = \theta_0$ $\Theta = (0, \infty)$
$X \sim B(p_0)$	$p_0 = \theta_0$ $\Theta = (0,1)$ bzw. $[0,1]$
$X \sim \text{Bin}(n, p)$	$p_0 = \theta_0$ $\Theta = (0,1)$ bzw. $[0,1]$
$X \sim N(\mu, \sigma_0^2)$	$\sigma_0^2 = \theta_0$ $\Theta = (0, \infty)$ bzw. $[0, \infty)$
$X \sim N(\mu_0, \sigma^2)$	$\mu_0 = \theta_0$ $\Theta = \mathbb{R}$
$X \sim N(\mu_0, \sigma_0^2)$	$\begin{pmatrix} \mu_0 \\ \sigma_0^2 \end{pmatrix} = \theta_0$ … 2-dim. $\Theta = \mathbb{R} \times (0, \infty)$

Betrachte einen Schätzer $\hat{\theta}$ für θ_0; das ist eine Größe $\hat{\theta} = \hat{\theta}(X_1, \ldots, X_n)$, die alleine aus den Daten $X_1, \ldots, X_n$ berechenbar ist (oft auch $\hat{\theta}_n$).

Beachte: Für $\theta_0 \in \mathbb{R}^p$ mit $p > 1$ ist auch $\hat{\theta} \in \mathbb{R}^p$.

Definition 1.1 (Unverzerrtheit) Der Schätzer $\hat{\theta}$ für θ_0 ist unverzerrt (bzw. erwartungstreu), wenn gilt:

$$\forall \theta \in \Theta, \forall i = 1, ..., p : \underset{\substack{\uparrow \\ \text{Erwartungstreu, wenn } \theta_n \\ \text{der wahre Parameter ist.}}}{\mathbb{E}_{\theta_n}} \left((\hat{\theta})_i \right) = (\theta)_i.$$

Beachte: Für $\hat{\theta} = \begin{pmatrix} \hat{\theta}_1 \\ \vdots \\ \hat{\theta}_p \end{pmatrix}$ ist $\mathbb{E}_{\theta_n}\left(\hat{\theta}\right) = \begin{pmatrix} \mathbb{E}_{\theta}\left(\hat{\theta}_1\right) \\ \vdots \\ \mathbb{E}_{\theta_n}\left(\hat{\theta}_p\right) \end{pmatrix}$.

Ist $\hat{\theta}$ unverzerrt, dann ist die i-te Komponente von $\hat{\theta}$ ein unverzerrter Schätzer für die i-te Komponente von θ_0.

Definition 1.2 (Konsistenz) Ein Schätzer $\underset{\substack{\uparrow \\ \text{Stichprobengröße}}}{\hat{\theta}_n}$ für θ_0 heißt konsistent, wenn gilt:

$$\forall \epsilon > 0, \forall i = 1, ..., p, \forall \theta \in \Theta : \mathbb{P}_{\theta_0}\left(\|(\hat{\theta}_n)_i - (\theta)_i\| > \epsilon \right) \overset{n \to \infty}{\longrightarrow} 0.$$

Kurz:
$$\hat{\theta}_n \xrightarrow{p} \theta_0.$$

Beachte: $\hat{\theta}_n$ ist konsistent für θ_0 genau dann, wenn $\left(\hat{\theta}_n\right)_i$ konsistent ist für $(\theta_0)_i$ für alle $i = 1, \dots, p$.

Nachrechnen. Siehe Übung. □

<u>Wunschliste für Schätzer:</u>

- Unverzerrtheit $\leftarrow$ keinen systematischen Fehler.
- Konsistenz $\leftarrow$ Genauigkeit wächst mit n.
- kleine Varianz
- Verteilung des $\hat{\theta}_n - \theta_0$ bekannt oder approximierbar.

Zwei Strategien:

- Momentenmethode
- Maximum-Likelyhood-Methode.

1.2 Die Momenten-Methode (MM)

Siehe zu diesem Kapitel S. 260-267. in *Rice (2007)*.

Betrachte ein parametrisches Modell wie in Abschnitt 1.1: $X_1, \dots, X_n$ i.i.d. Kopien von X, wobei $X \sim$ Verteilung abhängig von $\theta_0 \in \Theta \subseteq \mathbb{R}^p \quad (p \geq 1)$.
In vielen Fällen gibt es eine eindeutige Beziehung zwischen dem Parameter θ_0 und den Momenten der zugrundlegende Verteilung $X \quad (1:1 Beziehung)$.

Beispiel 1.4

	Parameter	Funktionen von $\mathbb{E}(X), \mathbb{E}(X^2), \dots$
$X \sim P_{\lambda_0}$	$\lambda_0 > 0$	$= \mathbb{E}_{\lambda_0}(X) = \lambda_0$
$X \sim Bin(n, p_0)$	$p_0 \in [0,1]$	$= \frac{1}{n}\mathbb{E}(X)_{p_0}$ weil $E_{Bin}(X) = n * p_0$
$X \sim N(\mu_0, \sigma_0^2)$	$\mu \in \mathbb{R}$	$= \mathbb{E}(X)$
	$\sigma_0^2 > 0$	$= \mathbb{E}(X^2) - (\mathbb{E}(X))^2 = Var(X)$

Bemerkung: (MM) Allgemein:
Für $\theta \leq \Theta$ und $k \geq 1$ setzt man $\mathbb{E}_{\theta_0}(X^k) = m_k(\theta_0)$ (Moment=Funktion des Parameters)

Angenommen es gibt eine Funktion $f : \mathbb{R}^p \to \mathbb{R}^p$, sodass $f\left(\begin{pmatrix} m_1(\theta_0) \\ \vdots \\ m_n(\theta_0) \end{pmatrix}\right) = \theta_0$ für jedes $\theta_0 \in \Theta$ die "glatt" ist.

Dann erhält man einen **Momenten-Methoden-Schätzer** $\hat{\theta}_{MM}$

- durch:
$$\hat{\theta} = \hat{\theta}_{MM} = f\left(\hat{m}_1, \ldots, \hat{m}_p\right)$$
- wobei
$$\hat{m}_j = \frac{1}{n}\sum_{i=1}^{n} X_i^j$$
für $1 \leq j \leq p$.

Bemerkung: Es gibt im Allgemeinem mehrere MM-Schätzer.

Beispiel 1.5 (Gleichverteilung) $X_1, \ldots, X_n$ i.i.d. $U([0,\theta])$, $\theta > 0$.

$$\begin{aligned}
\mathbb{E}(X_1) = \frac{\theta}{2} \qquad & \Rightarrow \theta = 2\mathbb{E}(X_1) \\
& \hat{\theta}_{MM} = 2\frac{1}{n}\sum_{i=1}^{n} X_i. \\
E(X_1^2) = \frac{\theta^2}{3} \qquad & \Rightarrow \theta = \sqrt{3\mathbb{E}\left(X_1^2\right)} \\
& \hat{\theta}_{MM} = \sqrt{3\frac{1}{n}\sum_{i=1}^{n} X_i^2}. \\
E(X_1^3) = \ldots \qquad & \Rightarrow \ldots \\
& \ldots
\end{aligned}$$

Erinnerung (Stetigkeit):
$g : \mathbb{R} \to \mathbb{R}$ ist **stetig** im Punkt $m \in \mathbb{R}$, wenn gilt:

$$\forall \epsilon > 0\, \exists\, \delta > 0 : |m' - m| < \delta \Rightarrow |g(m') - g(m)| < \epsilon.$$

$g : \mathbb{R}^p \to \mathbb{R}$ ist **stetig** im Punkt $m \in \mathbb{R}^p$, wenn gilt:

$$\forall \epsilon > 0\, \exists\, \delta > 0 : \|m' - m\| < \delta \Rightarrow |g(m') - g(m)| < \epsilon.$$

$g : \mathbb{R}^p \to \mathbb{R}^q$ ist **stetig** im Punkt $m \in \mathbb{R}^p$, wenn gilt:

$$\forall j = 1, ..., q \; ist \; f(\cdot)_j : \mathbb{R}^p \to \mathbb{R} \; stetig \; im \; Punkt \; m.$$

Lemma 1.1 Sind $X_1,\ldots,X_n$ i.i.d. mit $\mathbb{E}(X^k)=m_k$ für $k=1,\ldots,p$ und ist $g:\mathbb{R}^p\to\mathbb{R}$ stetig im Punkt $\left(m_1,\ldots,m_p\right)=m$, dann gilt für jedes $\epsilon>0$

$$\mathbb{P}\left(|g(\hat{m})-g(m)|>\epsilon\right)\overset{n\to\infty}{\longrightarrow}0.$$

(Hier ist $\hat{m}_k=\frac{1}{n}\sum_{i=1}^n X_i^k$ und $\hat{m}=\left(\hat{m}_1,\ldots,\hat{m}_p\right)$)

Beweis. Sei $\epsilon>0$. Weil g stetig ist im Punkt m, gibt es ein $\delta>0$, sodass gilt:

$$\|m'-m\|<\delta\Rightarrow|g(m')-g(m)|<\epsilon.$$

Setze $\hat{m}=\begin{pmatrix}\hat{m}_1\\ \vdots\\ \hat{m}_p\end{pmatrix}$. Dann gilt:

$$\begin{aligned}\{\|\hat{m}-m\|<\delta\}\quad&\subseteq\{|g(\hat{m})-g(m)|<\epsilon\}\\ \Rightarrow\{|g(\hat{m})-g(m)|\geq\epsilon\}\quad&\subseteq\{\|\hat{m}-m\|\geq\delta\}\end{aligned}$$

Also:

$$\begin{aligned}\mathbb{P}\left(|g(\hat{m})-g(m)|>\epsilon\right)&\leq\mathbb{P}\left(|g(\hat{m})-g(m)|\geq\epsilon\right)\\ &\leq\mathbb{P}\left(\|\hat{m}-m\|\geq\delta\right)\\ &\leq\mathbb{P}(\sqrt{p}*max\{|\hat{m}_k-m_k|,1\leq k\leq p\}\geq\delta)\\ &=\mathbb{P}(\bigcup_{p}^{k=1}|\hat{m}_k-m_k|>\frac{\delta}{\sqrt{p}})\\ &\leq\sum_{k=1}^{p}\underbrace{\mathbb{P}(|\hat{m}_k-m_k|>\frac{\delta}{\sqrt{p}})}_{\overset{n\to\infty}{\longrightarrow}0\,\text{(wegen LLN)}}\overset{n\to\infty}{\longrightarrow}0\end{aligned}$$

Mit dem Gesetz der großen Zahlen gilt:

$$\hat{m}_k=\frac{1}{n}\sum_{i=1}^{n}X_i^k\overset{p}{\longrightarrow}m_k\tag{$\ast\ast$}$$

für jedes $k=1,\ldots,p$.

$\square$

Satz 1.1 Betrachte X_i, $i\geq1$, i.i.d. wie in (MM).
Ist die Funktion $f:\mathbb{R}^p\to\mathbb{R}^p$ stetig in jedem Punkt der Menge $\left\{\left(m_1(\theta_0),\ldots,m_p(\theta_0)\right)'=\theta\in\Theta\right\}(\ast)$, dann ist $\hat{\theta}_{MM}$ konsistent für θ_0.

alle möglichen Werte der Momente

alle möglichen Werte des Parameters

Beweis.
Zz: Jede Komponente von $\hat{\theta}_n$ ist konsistent für die entsprechende Komponente von θ_0.

Sei $j \in \{1,...,p\}$. Laut Voraussetzung ist $f : \mathbb{R}^p \to \mathbb{R}^p$ stetig im jedem Punkt der Menge ($\ast$). Dann ist auch $f(\cdot)_j :: \mathbb{R}^p \to \mathbb{R}^p$ stetig im jedem Punkt von ($\ast$).

Sei $\theta_0 \in \Theta$ sodass $m := (m_1(\theta_0),...,m_p(\theta_0)) \in (\ast)$. Mit dem Lemma 1.1 folgt dass

$$\forall \epsilon > 0 : \mathbb{P}(|\underbrace{f(\hat{m})_j}_{=(\hat{\theta}_n)_j} - \underbrace{f(m)_j}_{=(\theta_0)_j}| > \epsilon) \to 0$$

Also ist $(\hat{\theta}_n)_j$ konsistenf für $(\theta_0)_j$.

□

Satz 1.2 Betrachte ein parametrisches Modell wo X Verteilung abh. von $\theta \leq \Theta \in \mathbb{R}$. Seien X_1, X_2 i.i.d. Kopien von X falls $\theta_0 = f(m_1(\theta_0))$ (für $\theta \in \Theta$) für eine Funktion f, die auf der Menge $\{m_1(\theta) : \theta \in \Theta\}$ stetig ist. Dann ist der MM-Schätzer

$$\hat{\theta}_n = f(\frac{1}{n}\sum_{i=1}^{n} X_i)$$

kosistent für θ_0.

Erinnerung aus GZ
Falls $X_n \xrightarrow{p} c$ und falls $f : \mathbb{R} \to \mathbb{R}$ stetig ist im Punkt c, dann folgt aus $f(X_n) \xrightarrow{p} f(c)$.
Wissen: $\bar{X}_n = \frac{1}{n} X_i$ ist ein konsistenter Schätzer für $E_{\theta_0}(X) = m_1(\theta_0)$.
D.h. $\bar{X}_n \xrightarrow{p} m_1(\theta_0)$. Laut voraussetzung ist die Funktion $f(\cdot)$ stetig im Punkt $m_1(\theta_0)$ =>

$$f(\bar{X}_n) \xrightarrow{p} f(m_1(\theta_0)), \; also \; \hat{\theta}_n \xrightarrow{p} \theta_n$$

mit andere Wörten $\hat{\theta}_n$ ist ein konsistenten Schätzer für θ_0.

Beispiel 1.6 (Poissonverteilung) $X_1, \ldots, X_n$ i.i.d. $P(\lambda_0)$, $\quad \lambda_0 \in (0, \infty)$.
Beachte $\mathbb{E}_{\lambda_0}(X_1) = \lambda_0$, dass gibt dann der MM-Schätzer

$$\hat{\lambda}_n = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Der Schätzer ist unverzerrt und konsistent. ✓

Wissen: $\lambda_0 = \mathrm{Var}_{\lambda_0}(X_1) = \mathbb{E}_{\lambda_0}(X_1^2) - (\mathbb{E}_{\lambda_0}(X_1))^2$ daraus kann man einen witeren MM-Schätzer mit λ_0 erhalten.

Beispiel 1.7 (Normalverteilung) $X_1,\dots,X_n$ i.i.d. $N\left(\mu_0,\sigma_0^2\right)$ mit $\mu_0\in\mathbb{R},\sigma_0^2\in(0,\infty)$.
$\mathbb{E}_{\mu_0,\sigma_0^2}(X_1)=\mu_0,\mathbb{E}_{\mu_0,\sigma_0^2}(X_1^2)=\sigma_0^2-\mu_0,$
$Var(X_1)=\mathbb{E}(X_1^2)-(\mathbb{E}(X_1))^2=\sigma_0^2$
Das gibt den MM-Schätzer:

$$\begin{pmatrix}\hat\mu_n\\ \hat\sigma_n^2\end{pmatrix}=\begin{pmatrix}\frac{1}{n}\sum_{i=1}^n X_i\\ \frac{1}{n}\sum_{i=1}^n X_i^2-\left(\frac{1}{n}\sum_{i=1}^n X_i\right)^2\end{pmatrix}.$$

$\mathbb{E}_{\mu_0,\sigma_0^2}(X_1)=\mu_0$
$\mathbb{E}_{\mu_0,\sigma_0^2}(\hat\sigma_n^2)=\frac{n-1}{n}\sigma_0^2\neq\sigma_0^2...verzerrt$
$\hat\sigma_n^2=\frac{1}{n}\sum_{i=1}^n\left(X_i-\bar X_n\right)^2$ Vo-GZS

Konsistenz:
Die Funktion $f\begin{pmatrix}m_1\\ m_2\end{pmatrix}\mapsto\begin{pmatrix}m_1\\ m_2-m_1^2\end{pmatrix}$ ist stetig auf ganz $\mathbb{R}^2$.
Damit ist $\begin{pmatrix}\hat\mu\\ \hat\sigma^2\end{pmatrix}$ konsistent für $\begin{pmatrix}\mu_0\\ \sigma_0^2\end{pmatrix}$.✓

Beispiel 1.8 (Geometrische Verteilung) $X_1,\dots,X_n$ i.i.d. $G(p_0)$, $\quad p_0\in(0,1)$.
Wissen: Für $X\sim G(p)$ ist
$\mathbb{E}_{p_0}(X_1)=\frac{1}{p_0},\quad\Rightarrow p_0=\frac{1}{\mathbb{E}_{p_0}(X)}$
Das erzeugt den MM-Schätzer:

$$\hat p_0=\frac{1}{\frac{1}{n}\sum_{i=1}^n X_i}.$$

Dieser Schätzer ist verzerrt! $\mathbb{E}(f(X))\neq f(\mathbb{E}(X))$!
falls $f(\cdot)$ linear ist, also $f(X)=aX+b$ ist, $\mathbb{E}(aX+b)=a\mathbb{E}(X)+b=f(\mathbb{E}(X))$;

$$\mathbb{E}_p(\hat p_n)\neq\frac{1}{\mathbb{E}_p\left(\frac{1}{n}\sum_{i=1}^n X_i\right)}=\frac{1}{\frac{1}{p}}=p.$$

Konsistenz:

Die Funktion $f(t)=\frac{1}{t}$ ist stetig auf $(0,\infty)$.
Die möglichen Werte von $\mathbb{E}_{p_0}(X_1)$ ist die Menge

$$\{\frac{1}{p}:p\in(0,1)\}=(1,\infty).(\star)$$

Damit ist die Funktion $f(\cdot)$ stetig in jedem Punkt der Menge $(\star)$ und mit dem letzten Satz ist

$$\hat p_n=f(\frac{1}{n}\sum_{i=1}^n X_i)$$

konsistent für p_0.✓

Beispiel 1.9 $X_1, \ldots, X_n$ i.i.d. $\sim \mathcal{U}([\,\theta_0, \theta_0 + 1\,]), \theta_0 \in \mathbb{R}$

Hier ist $\mathbb{E}_{\theta_0}(X_1) = \theta_0 + \frac{1}{2}$. Das gibt dem MM-Schätzer

$$\hat{\theta}_n = \frac{1}{n} \sum_{i=1}^{n} X_i - \frac{1}{2}.$$

Der Schätzer ist unverzerrt (weil $f(x) = x - \frac{1}{2}$ linear ist) und es ist konsistent (weil $f(x)$ stetig ist auf ganz $\mathbb{R}$).

Verteilung des Schätzfehlers/Konfidenzintervalls:

Beispiel 1.10 (Normalverteilung) $X_1, \ldots, X_n$ i.i.d. $N(\mu_0, \sigma_0^2), \mu_0 \in \mathbb{R}, \sigma_0^2 > 0$.

- Konfidenzintervall (KI) für μ_0:

Wissen:

$$\sqrt{n} \frac{\bar{X}_n - \mu_0}{\hat{\sigma}_n} \sim t_{n-1},$$

wobei $\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i = \hat{\mu}_{MM}$ und $\tilde{\sigma}_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2 = \frac{n}{n-1} \hat{\sigma}_{MM}^2$.
Gegeben $\alpha \in (0,1)$. Wähle $a = \mathrm{F}_{t_{n-1}}^{-1}\left(\frac{\alpha}{2}\right)$ und $b = \mathrm{F}_{t_{n-1}}^{-1}\left(1 - \frac{\alpha}{2}\right)$ (Quantile der t_{n-1}-Verteilung).
Dann ist

$$\begin{aligned}
(\star) = \mathbb{P}\left(a \leq \sqrt{n} \frac{\bar{X}_n - \mu_0}{\hat{\sigma}_n} \leq b\right) &= \underbrace{\mathrm{F}_{n-1}(b) - \mathrm{F}_{n-1}(a)}_{\text{Verteilungsfunktionen der } t_{n-1}} \\
&= 1 - \frac{\alpha}{2} - \frac{\alpha}{2} \\
&= 1 - \alpha.
\end{aligned}$$

$$\begin{aligned}
(\star) &= \mathbb{P}\left(a \frac{\hat{\sigma}_n}{\sqrt{n}} \leq \bar{X}_n - \mu_0 \leq b \frac{\hat{\sigma}_n}{\sqrt{n}}\right) \\
&= \mathbb{P}\left(-\bar{X}_n + a \frac{\hat{\sigma}_n}{\sqrt{n}} \leq -\mu_0 \leq -\bar{X}_n + b \frac{\hat{\sigma}_n}{\sqrt{n}}\right) \\
&= \mathbb{P}\left(\bar{X}_n - b \frac{\hat{\sigma}_n}{\sqrt{n}} \leq \mu_0 \leq \bar{X}_n - a \frac{\hat{\sigma}_n}{\sqrt{n}}\right) \\
&= \mathbb{P}\left(\mu_0 \in [\,\bar{X}_n - b \frac{\hat{\sigma}_n}{\sqrt{n}}, \bar{X}_n - a \frac{\hat{\sigma}_n}{\sqrt{n}}\,]\right).
\end{aligned}$$

Also: Das KI für μ_0 mit Überdeckungswahrscheinlichkeit $1 - \alpha$: $\left[\bar{X}_n - b \frac{\hat{\sigma}_n}{\sqrt{n}}, \bar{X}_n - a \frac{\hat{\sigma}_n}{\sqrt{n}}\right]$

Informell: $\left[\bar{X}_n \pm b \frac{\hat{\sigma}_n}{\sqrt{n}}\right]$, Beobachte: $a = -b$

- KI für σ_0^2:

Wissen:

$$\sum_{i=1}^{n}(X_i - \bar{X}_n)^2 \sim \sigma_0^2 \chi^2_{n-1}$$

$$\Rightarrow \hat{\sigma}_n^2 \sim \frac{\sigma_0^2 \chi^2_{n-1}}{n-1}$$

$$(n-1)\frac{\hat{\sigma}_n^2}{\sigma_0^2} \sim \chi^2_{n-1}.$$

Gegeben $\alpha \in (0,1)$ wählt man hier $a = \mathrm{F}^{-1}_{\chi^2_{n-1}}\left(\frac{\alpha}{2}\right)$ und $b = \mathrm{F}^{-1}_{\chi^2_{n-1}}\left(1-\frac{\alpha}{2}\right)$ (Quantile der χ^2_{n-1}-Verteilung).
Damit ist

$$\begin{aligned}(\star) = \mathbb{P}\left(a \leq (n-1)\frac{\hat{\sigma}_n^2}{\sigma_0^2} \leq b\right) &= \mathrm{F}_{\chi^2_{n-1}}(b) - \mathrm{F}_{\chi^2_{n-1}}(a)\\ &= 1 - \frac{\alpha}{2} - \frac{\alpha}{2}\\ &= 1-\alpha.\end{aligned}$$

$$\begin{aligned}(\star) =& \mathbb{P}\left(\frac{a}{(n-1)\hat{\sigma}_n^2} \leq \frac{1}{\sigma_0^2} \leq \frac{b}{(n-1)\hat{\sigma}_n^2}\right)\\ =& \mathbb{P}\left(\frac{(n-1)\hat{\sigma}_n^2}{b} \leq \sigma_0^2 \leq \frac{(n-1)\hat{\sigma}_n^2}{a}\right)\\ =& \mathbb{P}\left(\sigma_0^2 \in \left[\frac{(n-1)\hat{\sigma}_n^2}{b}, \frac{(n-1)\hat{\sigma}_n^2}{a}\right]\right).\end{aligned}$$

Also KI für σ_0^2 mit Überdeckungswahrscheinlichkeit $1-\alpha : \left[\frac{(n-1)\hat{\sigma}_n^2}{b}, \frac{(n-1)\hat{\sigma}_n^2}{a}\right]$.

Beispiel 1.11 (Poisson-Modell) $X_1,\dots,X_n$ i.i.d. $P(\lambda_0), \quad (\lambda_0 > 0)$.
Hier ist $\mathbb{E}(X_1) = \lambda_0, \quad \mathrm{Var}(X_1) = \lambda_0$.
($\ast$) *Erinnerung aus der Vorlesung Grundzüge der Statistik:* Für $Z_1,\dots,Z_n$ i.i.d. mit $\mathbb{E}(Z_1) = \mu_0$ und $\mathrm{Var}(Z_1) = \sigma^2$. Ist ein KI für μ mit **nominaler** Überdeckungswahrscheinlichkeit $1-\alpha$

$$\left[\hat{\mu}_n - \frac{\tilde{\sigma}_n}{\sqrt{n}}\Phi^{-1}\left(1-\frac{\alpha}{2}\right), \hat{\mu}_n - \frac{\tilde{\sigma}_n}{\sqrt{n}}\Phi^{-1}\left(\frac{\alpha}{2}\right)\right]$$

hierbei ist $\hat{\mu}_n = \frac{1}{n}\sum_{i=1}^n Z_i$ und $\hat{\sigma}_n^2 = \frac{1}{n-1}\sum_{i=1}^n (Z_i - \hat{\mu}_n)^2$.
Damit erhält man das KI für λ_0 nämlich

$$\left[\bar{X}_n - \frac{\hat{\sigma}_n}{\sqrt{n}}\Phi^{-1}\left(1-\frac{\alpha}{2}\right), \bar{X}_n - \frac{\hat{\sigma}_n}{\sqrt{n}}\Phi^{-1}\left(\frac{\alpha}{2}\right)\right] \leftarrow \text{Generisches Intervall}$$

mit nominaler Überdeckungswahrscheinlichkeit $1-\alpha$ wobei $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i, \ \hat{\sigma}_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2$

Beachte: $\bar{X}_n$ schätzt $\mathbb{E}(X_1) = \lambda_0$, $\hat{\sigma}_n^2$ schätzt $\mathrm{Var}(X_1) = \lambda_0$.

Im Kontext von $(\ast)$ gilt (Siehe VO GZ: Beispiel 8.8):

$$\sqrt{n}\frac{\hat{\mu}_n-\mu}{\sigma}\xrightarrow{w}N(0,1),$$

und auch

$$\sqrt{n}\frac{\hat{\mu}_n-\mu}{\hat{\sigma}_n}\xrightarrow{w}N(0,1).$$

Sei $\tilde{\sigma}_n^2$ ein anderer konsistenter Schätzer für σ^2

$$\sqrt{n}\frac{\hat{\mu}_n-\mu}{\tilde{\sigma}_n^2}=\sqrt{n}\underbrace{\frac{\hat{\mu}_n-\mu}{\sigma}}_{\xrightarrow{w}N(0,1)}\underbrace{\frac{\sigma}{\tilde{\sigma}_n^2}}_{f(\tilde{\sigma}_n^2)}\xrightarrow{p}1.$$

$f(\tilde{\sigma}_n^2)$ für $f(t)=\sqrt{\frac{\sigma^2}{t}}$:
f ist stetig auf $(0,\infty)$ und inbsbesondere im Punkt σ^2
Der Schätzer $\tilde{\sigma}_n^2$ ist konsistent für σ^2 : $\tilde{\sigma}_n^2\xrightarrow{p}\sigma^2$
Rechnenregel: $f(\tilde{\sigma}_n^2)\xrightarrow{p}f(\sigma^2)$
M.a.W. $\frac{\sigma}{\tilde{\sigma}_n}\xrightarrow{p}1$.

Mit Rechnenregeln aus GZ folgt: $\sqrt{n}\frac{\hat{\mu}_n-\mu}{\tilde{\sigma}_n^2}\xrightarrow{w}N(0,1)\cdot 1\equiv N(0,1)$.)

<u>Also:</u> Im Poisson-Modell ist $\bar{X}_n$ ein konsistenter Schätzer für $\mathrm{Var}(X_1)=\lambda_0$. Damit gilt:

$$\sqrt{n}\frac{\bar{X}_n-\lambda_0}{\sqrt{\bar{X}_n}}\xrightarrow{w}N(0,1). \qquad (\ast\ast)$$

Gegeben $\alpha\in(0,1)$ wähle hier $a=\Phi^{-1}\left(\frac{\alpha}{2}\right)$ und $b=\Phi^{-1}\left(1-\frac{\alpha}{2}\right)$.
Dann ist

$$\begin{aligned}\mathbb{P}\left(a\le\sqrt{n}\frac{\bar{X}_n-\lambda_0}{\sqrt{\bar{X}_n}}\le b\right)&\overset{(\ast\ast)}{=}\mathbb{P}\left(\sqrt{n}\frac{\bar{X}_n-\lambda_0}{\sqrt{\bar{X}_n}}\le b\right)-\mathbb{P}\left(\sqrt{n}\frac{\bar{X}_n-\lambda_0}{\sqrt{\bar{X}_n}}\le a\right)\\&\approx\Phi(b)-\Phi(a)\\&=1-\frac{\alpha}{2}-\frac{\alpha}{2}\\&=1-\alpha.\end{aligned}$$

$$\begin{aligned}(\ast\ast)&=\mathbb{P}\left(\sqrt{\frac{\bar{X}_n}{n}}a\le\bar{X}_n-\lambda_0\le\sqrt{\frac{\bar{X}_n}{n}}b\right)\\&=\mathbb{P}\left(-\bar{X}_n+\sqrt{\frac{\bar{X}_n}{n}}a\le-\lambda_0\le-\bar{X}_n+\sqrt{\frac{\bar{X}_n}{n}}b\right)\\&=\mathbb{P}\left(\bar{X}_n-b\sqrt{\frac{\bar{X}_n}{n}}\le\lambda_0\le\bar{X}_n-a\sqrt{\frac{\bar{X}_n}{n}}\right).\end{aligned}$$

Also ein KI für λ_0 mit nominaler Überdeckungswahrscheinlichkeit $1-\alpha$ ist $\left[\bar{X}_n-b\sqrt{\frac{\bar{X}_n}{n}},\bar{X}_n-a\sqrt{\frac{\bar{X}_n}{n}}\right]\leftarrow$ Spezielles Intervall im Poinsson-Modell
(Beachte! -a=b)

Beispiel 1.12 (Geometrische Verteilung) $X_1,\dots,X_n \sim G(p_0), \quad p_0 \in (0,1)$.
Wissen: $\mathbb{E}(X_1) = \frac{1}{p_0} = \mu_0, (p_0 = \frac{1}{\mathbb{E}(X_1)})$ $\quad \mathrm{Var}(X_1) = \frac{1-p_0}{p_0^2}$, MM-Schätzer für $p_0 : \hat{p}_n = \frac{1}{\bar{X}_n}$ (nicht linear!).
Hinweis zur geometrischen Verteilung:

- # Versuche bis zum 1. Erfolg (wird in der Vorlesung verwendet)
- # Misserfolge bis zum 1. Erfolg + 1 (wird in R verwendet).

$\frac{1-p_0}{p_0^2} = \frac{\frac{1}{p_0}-1}{p_0} = \frac{1}{p_0}(\frac{1}{p_0} - 1) = \mu_0(\mu_0 - 1)$,
$\mathbb{E}(X_1) = \mu_0,\ Var(X_1) = \mu_0(\mu_0 - 1)$

Suche KI für $\mu_0 = \mathbb{E}(X_1) = \frac{1}{p_0}$
Wie im letzten Beispiel erhält man das KI

$$\left[\bar{X}_n - \sqrt{\frac{\hat{\sigma}_n^2}{n}}\Phi^{-1}\left(1 - \frac{\alpha}{2}\right), \bar{X}_n - \sqrt{\frac{\hat{\sigma}_n^2}{n}}\Phi^{-1}\left(\frac{\alpha}{2}\right)\right],$$

mit nominaler Überdeckungswahrscheinlichkeit $1 - \alpha$, wobei

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

Im geometrisches Modell ist $\tilde{\sigma}_n^2 = \bar{X}_n(\bar{X}_n - 1)$ ein Schätzer für $Var(X_1)$ der konsistent ist.

Das gibt ein witeres KI für μ_0 nämlch:

$$\left[\bar{X}_n - \sqrt{\frac{\tilde{\sigma}_n^2}{n}}\Phi^{-1}\left(1 - \frac{\alpha}{2}\right), \bar{X}_n - \sqrt{\frac{\tilde{\sigma}_n^2}{n}}\Phi^{-1}\left(\frac{\alpha}{2}\right)\right]$$

mit nominaler Überdeckungswahrscheinlichkeit $1 - \alpha$.

Betrachte: $Z_1,...,Z_n$ i.i.d; $\mathbb{E}(Z_1) = \mu, Var(Z_1) = \sigma^2$
KI für $\mu : \bar{Z}_n \mp \sqrt{\frac{\hat{\sigma}_n^2}{n}}\Phi^{-}1(1 - \frac{\alpha}{2})$
KI für $f(m)$ =?

- <u>Fall 1:</u> $f(\cdot)$ linear, $f(t) = At + B, A > 0$, A und B bekannt;
  Wissen: $1 - \alpha \approx \mathbb{P}\left(\bar{Z}_n - \sqrt{\frac{\hat{\sigma}_n^2}{n}}\Phi^{-1}(1 - \frac{\alpha}{2}) \le \mu \le \bar{Z}_n + \sqrt{\frac{\hat{\sigma}_n^2}{n}}\Phi^{-1}(1 - \frac{\alpha}{2})\right)$
  $= \mathbb{P}\left(A\left(\bar{Z}_n - \sqrt{\frac{\hat{\sigma}_n^2}{n}}\Phi^{-1}(1 - \frac{\alpha}{2})\right) + B \le f(\mu) \le A\left(\bar{Z}_n + \sqrt{\frac{\hat{\sigma}_n^2}{n}}\Phi^{-1}(1 - \frac{\alpha}{2})\right) + B\right)$
  ...gibt KI für $f(\mu)$ mit nominaler Überdeckungswahrscheinlichkeit $1 - \alpha$
  $f(\cdot)$ linear wie oben, $A < 0$... analog.
- <u>Fall 2:</u> $f(\cdot)$ nicht linear + Skizze

$\bar{X}_n \pm \sqrt{\frac{\tilde{\sigma}_n^2}{n}}\Phi^{-1}(1-\frac{\alpha}{2})$ KI für $\frac{1}{p_0}$ für X_1 i.i.d. $G(p_0)$

<u>1.Schätzer:</u> $\frac{1}{n-1}\sum_{i=1}^{n}(X_1-\bar{X}_n)^2$
<u>2.Schätzer:</u> $\bar{X}_n(\bar{X}_n-1)$

Hinweis: $\frac{1-p_0}{p_0^2} = \frac{1}{p_0}(\frac{1}{p_0}-1)$

$Z_1,...,Z_m$ i.i.d. und

- $\mathbb{E}(Z_1) = q$... Überdeckungswahrscheinlichkeit
- $Var(Z_1) = q(q-1)$

$\bar{Z}_m$: KI für q: $\bar{Z}_m \pm \underbrace{\sqrt{\frac{\cdot}{m}}\Phi^{-}1(1-\frac{0.05}{2})}_{\text{Typische Fehler}=2\frac{\text{std. Abweichung}}{\sqrt{m}}}$

$X_1,...,X_n$ i.i.d. $G(p_0)$
? KI für $p_0 = \frac{1}{\mathbb{E}_{p_0}(X_1)}$

- $f(\cdot)$ linear: $(f(x) = ax+b)$

  $$\sqrt{n}(\bar{X}_n - \mathbb{E}(X_1)) \xrightarrow{w} N(0, Var(X_1))$$
  $$\underbrace{\sqrt{n}(f(\bar{X}_n) - f(\mathbb{E}(X_1)))}_{\sqrt{n}(a\bar{X}_n+b-(a\mathbb{E}(X_1)+b))} \xrightarrow{w} N(0, a^2Var(X_1))$$
  $$a\sqrt{n}(\bar{X}_n - \mathbb{E}(X_1)) \xrightarrow{w} \underbrace{N(0, Var(X_1))}_{\xrightarrow{w} N(0,a^2Var(X_1))}$$

- $f(\cdot)$ nicht linear: siehe nächste Satz!

Satz 1.3 **(δ-Methode)** Es seien Z_i i.i.d. Kopien von Z, mit $\mathbb{E}(Z) = \mu_Z$ und $\mathrm{Var}(Z) = \sigma_Z^2 > 0$.
Ist die Funktion $f: \mathbb{R} \to \mathbb{R}$ im Punkt μ_Z **differenzierbar**, dann gilt:

$$\sqrt{n}\Big(f(\bar{Z}_n) - f(\mu_Z)\Big) \xrightarrow{w} N\Big(0, (f'(\mu_Z))^2\,\sigma_Z^2\Big),$$

wobei $\bar{Z}_n = \frac{1}{n}\sum_{i=1}^{n} Z_i$.

Illustration: Siehe R-script vom 22.04.2021.

Beweis. Erinnerung:

$\sqrt{n}\left(\bar{Z}_n - \mu_z\right) \xrightarrow{w} N(0, \sigma_Z^2)$	Zentraler Grenzwertsatz
$\bar{Z}_n \xrightarrow{p} \mu_Z$	Gesetz der großen Zahlen
$X_n \xrightarrow{p} c, f(\cdot)$ stetig in c $\Rightarrow f(X_n) \xrightarrow{p} f(c)$ $X_n \xrightarrow{w} X, Y_n \xrightarrow{p} c$ $\Rightarrow X_n + Y_n \xrightarrow{w} X + c$ $\Rightarrow X_n \cdot Y_n \xrightarrow{w} X \cdot c$	Rechenregeln für $\xrightarrow{p}, \xrightarrow{w}$, GZS S. 80
$f'(\mu_Z) = \lim_{Z \to \mu_Z} \frac{f(Z) - f(\mu_Z)}{Z - \mu_z}$	VO Analysis

Definiere eine Funktion $R(Z)$ als

$$R(Z) = \begin{cases} f'(\mu_z) - \frac{f(Z) - f(\mu_z)}{Z - \mu_Z} & : Z \neq Z_0, \\ 0 & : Z = \mu_Z. \end{cases}$$

Dann ist $R(\cdot)$ stetig im Punkt μ_Z:

$$\lim_{Z \to \mu_Z} R(Z) \underset{\substack{\text{weil } f \text{ differenzierbar} \\ \text{ist im Punkt } \mu_Z}}{=} 0 \underset{\substack{\text{siehe} \\ \text{oben}}}{=} R(\mu_Z).$$

Ersetze Z durch $\bar{Z}_n$

$$f'(\mu_Z) = \frac{f(\bar{Z}_n) - f(\mu_Z)}{\bar{Z}_n - \mu_Z} + R(\bar{Z}_n)$$

Multiplizieren mit $\sqrt{n}(\bar{Z}_n - \mu_Z)$

$$f'(\mu_Z)\sqrt{n}(\bar{Z}_n - \mu_Z) = \sqrt{n}(f(\bar{Z}_n) - f(\mu_Z)) + \sqrt{n}(\bar{Z}_n - \mu_Z)R(\bar{Z}_n)$$

$$\sqrt{n}\left(f(\bar{Z}_n) - f(\mu_Z)\right) = \underbrace{f'(\mu_Z)\underbrace{\sqrt{n}\left(\bar{Z}_n - \mu_Z\right)}_{\xrightarrow{w} N(0,\sigma_Z^2)}}_{\xrightarrow{w} N\left(0,(f'(\mu_Z))^2\sigma_Z^2\right)} - \underbrace{\underbrace{\sqrt{n}(\bar{Z}_n - \mu_Z)}_{\xrightarrow{w} N(0,\sigma_Z^2)}\ \underbrace{R(\bar{Z}_n)}_{\xrightarrow{p} R(\mu_Z)=0}}_{\xrightarrow{w} N(0,\sigma_Z^2)\cdot 0 = 0}$$

$$\xrightarrow{w} N\left(0, (f'(\mu_Z))^2 \sigma_Z^2\right).$$

□

Beispiel 1.13 **(Fortsetzung des oberen Beispiels - Geometrische Verteilung)**
$X_i,\ i \geq 1$, i.i.d. $G(p_0),\ p_0 \in (0,1)$.
Wissen:

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{p} \mathbb{E}_{p_0}(X_1) = \frac{1}{p_0},$$

$$\underbrace{\hat{p}_0}_{\text{MM-Schätzer}} = \frac{1}{\bar{X}_n} \xrightarrow{p} p_0.$$

$$\sqrt{n}\left(\bar{X}_n - \frac{1}{p_0}\right) \xrightarrow{w} N\left(0, \frac{1-p_0}{p_0^2}\right).$$

Betrachte $f(t) = \frac{1}{t}$, sodass
$\hat{p}_n = f(\bar{X}_n)$ und $p_0 = f(\frac{1}{p_0})$
$f(\cdot)$ ist stetig an der Stelle $\frac{1}{p_0}$
$f'(t) = -\frac{1}{t^2}$
$f'(\frac{1}{p_0}) = \frac{1}{-(\frac{1}{p_0})^1} = -p_0^2$
$(f'(\frac{1}{p_0}))^2 = p_0^4$
Mit δ-Methode:

$$\sqrt{n}\Big(\underbrace{f\left(\bar{X}_n\right)}_{\hat{p}_n} - \underbrace{f\left(\frac{1}{p_0}\right)}_{p_0}\Big) \xrightarrow{w} N\Big(0, \underbrace{\left(f'\left(\frac{1}{p_0}\right)\right)^2 \frac{1-p_0}{p_0^2}}_{p_0^4 \frac{1-p_0}{p_0^2} = p_0^2(1-p_0)}\Big)$$

$$\text{Also: } \sqrt{n}(\hat{p}_n - p_0) \xrightarrow{w} N\left(0, p_0^2(1-p_0)\right).$$

$$\text{bzw.} \sqrt{n}\frac{\hat{p}_n - p_0}{\sqrt{p_0^2(1-p_0)}} \xrightarrow{w} N(0,1)$$

$$\tilde{\sigma}_n^2 = \hat{p}_n^2(1-\hat{p}_n) \xrightarrow{p} p_0^2(1-p_0)$$

(weil $\hat{p}_n \xrightarrow{p} p_0$ und weil $g(t) = t^2(1-t)$ stetig ist im Punkt p_0)

$$\text{Damit gilt auch } \sqrt{n}\frac{\hat{p}_n - p_0}{\sqrt{\hat{p}_n^2(1-\hat{p}_n)}} \xrightarrow{w} N(0,1)$$

Für $\alpha \in (0,1)$ ist ein Konfidenzintervall für p_0 mit nominaler Überdeckungswahrscheinlichkeit $1-\alpha$ damit gegeben durch

$$\left[\hat{p}_n \pm \sqrt{\frac{\hat{p}_n^2(1-\hat{p}_n)}{n}}\Phi^{-1}\left(1-\frac{\alpha}{2}\right),\right].$$

Bemerkung: Das letzte Beispiel zeigt die Konstruktion von Konfidenzintervallen mit zentralem Grenzwertsatz und δ-Methode. Eine alternative Konstruktion liefert der Bootstrap.

Bemerkung: Es gibt auch eine multivariate Version der δ-Methode.

Die tatsächliche U-Wkeit dieses Intervalls hängt von n und p_0 ab.
Wissen: für jedes $p_0 \in (0,1)$ gilt

$$\mathbb{P}_{p_0}\left(p_0 \in \hat{p}_n \pm \sqrt{\frac{\hat{p}_n^2(1-\hat{p}_n)}{n}}\Phi^{-1}\left(1-\frac{\alpha}{2}\right)\right) \xrightarrow{n \leftarrow \infty} 1-\alpha$$

Wie groß muss n sein, damit diese Wahrscheinlichkeit gleich 0.95 ± 0.01 ist.
Simulation in R:

n/p_0	5	20	80	320	640	1280	2560
0.01	✓	✓	✓	✓	✓	✓	✓
0.1	✓	✓	✓	✓	✓	✓	✓
0.5	×	✓	✓	✓	✓	✓	✓
0.9	×	×	×	✓	✓	✓	✓
0.99	×	×	×	×	×	×	✓

Die Qualität der Approximation hängt von n und p_0 ab. Beachte: p_0 ist konsistent schätzbar!
Betrachte $n, \hat{p}_n$

Diese Approximation (Überdeckungswahrscheinlichkeit $\approx 1-\alpha$) ist punktwise:

$$\forall\ p_0 : \text{Fehler } (n, p_0) \xrightarrow{n\to\infty} 0,$$

In eigenen Fällen ist auch eine gleichmäßige Approximation verfügbar.

$$\sup_{n,p_0} \quad \text{Fehler } (p_0) \xrightarrow{n\to\infty} 0.$$

In vorliegenden Fall (geom. Modell) gibt es leider keine gleichmäßige Approximation.

1.3 Die Maximum-Likelihood Methode (ML)

Siehe zu diesem Kapitel S. 267-285. in *Rice (2007)*.

Betrachte i.i.d. Zufallsvariablen $X_1,\dots,X_n$ deren jeweilige Dichte bzw. Wahrscheinlichkeitsfunktion von folgender Form ist:

$$f(x|\theta_0), \qquad \theta_0 \in \Theta \subseteq \mathbb{R}^p.$$

Die funktionale Form von $f(x|\theta_0)$ ist bekannt, aber der Parameter θ_0 ist unbekannt.

Idee: Gegeben Beobachtungen $x_1,\dots,x_n$ von $X_1,\dots,X_n$ schätzt man θ_0 durch jenen Wert von $\theta \in \Theta$, für welchen die Beobachtungen "am Wahrscheinlichsten" sind. Man wählt also jenes $\theta \in \Theta$, das die Beobachtungen "am Besten beschreibt".

Diskrete Modelle:

Betrachte i.i.d. Zufallsvariablen $X_1, \ldots, X_n$, die diskret sind mit Wahrscheinlichkeitsfunktion

$$p(x|\theta_0), \quad \theta_0 \in \Theta \leq \mathbb{R}^p, \quad x \in \mathbb{N}.$$

Betrachte Beobachtungen $x_1, \ldots, x_n$ von $X_1, \ldots, X_n$. Ist der wahre Parameter gleich $\theta \in \Theta$, dann ist die Wahrscheinlichkeit, die Werte $x_1, \ldots, x_n$ zu beobachten, gegeben durch

$$p_\theta(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) := L(\theta).$$

$L(\theta)$ nennt man die **Likelihood**.
Beachte $L(\theta) = \mathbb{P}_\theta(X_1 = x_1) \cdot \mathbb{P}_\theta(X_2 = x_2) \cdot \ldots \cdot \mathbb{P}_\theta(X_n = x_n) = p(x_1|\theta) \cdot p(x_2|\theta) \cdot \ldots \cdot p(x_n|\theta)$.
Alternativ dazu betrachtet man oft die sogenannte Log-Likelihood

$$\begin{aligned} l(\theta) &= \log(L(\theta)) \\ &= \sum_{i=1}^{n} \log p(x_i|\theta_0). \end{aligned}$$

Beachte: $L(\theta)$ und $l(\theta)$ hängen von den beobachteten Werten $x_1, \ldots, x_n$ ab.

Die Maximum-Likelihood Methode maximiert $L(\theta)$ bzw. $l(\theta)$ über $\theta \in \Theta$:

$$\hat{\theta}_{ML} = \underset{\theta \in \Theta}{\operatorname{argmax}}\, L(\theta) = \underset{\theta \in \Theta}{\operatorname{argmax}}\, l(\theta),$$

wobei $\hat{\theta}_{ML}$ ein Maximum-Likelihood-Schätzer für θ ist.

Beachte: Ein Maximierer von $L(\theta)$ ist auch ein Maximierer von $l(\theta)$, und umgekehrt.

Bemerkung: Manchmal gibt es mehr als einem Maximierer der Likelihood.

Beachte: $L(\theta)$ und $l(\theta)$ häangen von den beobachteten Werten $X_1, ..., X_n$ ab.

Beispiel 1.14 (Bernoulli-Verteilung) $X_1, \ldots, X_n$ i.i.d. $B(\theta_0)$, $\theta_0 \in (0,1)$. Die Wahrscheinlichkeitsfunktion ist hier

$$\begin{aligned} p(x|\theta_0) &= \begin{cases} \theta_0 & : x = 1, \\ 1 - \theta_0 & : x = 0 \end{cases} \\ &= \theta_0^x (1 - \theta_0)^{1-x}. \end{aligned}$$

Für Beobachtungen $x_1, \ldots, x_n$ von $X_1, \ldots, X_n$ ist die Likelihood gegeben durch

- $$\begin{aligned} L(\theta) &= p(x_1|\theta)\cdot\ldots\cdot p(x_n|\theta) \\ &= \underbrace{\left(\theta^{x_1}(1-\theta)^{1-x_1}\right)\left(\theta^{x_2}(1-\theta)^{1-x_2}\right)\cdot\ldots\cdot\left(\theta^{x_n}(1-\theta)^{1-x_n}\right)}_{n \text{ Faktoren. Jeder Faktor ist entweder gleich } \theta \text{ oder } 1-\theta} \\ &= \theta^{S_n}(1-\theta)^{n-S_n} \\ &\text{wobei } S_n = \sum_{i=1}^{n} x_i \end{aligned}$$

- $$l(\theta) = S_n \cdot \log(\theta) + (n - S_n)\log(1-\theta).$$

- $$l'(\theta) = S_n\frac{1}{\theta} + (n-S_n)\frac{-1}{1-\theta} \quad = \frac{S_n}{\theta} - \frac{n-S_n}{1-\theta}.$$

$$\begin{aligned} & f'(\theta) \overset{!}{=} 0 \\ \Leftrightarrow \quad & \frac{S_n}{\theta} = \frac{n-S_n}{1-\theta} \\ \Leftrightarrow \quad & S_n(1-\theta) = \theta(n-S_n) \\ \Leftrightarrow \quad & S_n - S_n\theta = n\theta - S_n\theta \\ \Leftrightarrow \quad & \frac{S_n}{n} = \theta_{\ast}. \end{aligned}$$

- $$\begin{aligned} f''(\theta) &= -\frac{S_n}{\theta^2} - (n-S_n)\frac{-1}{(1-\theta)^2} \\ &= -\underbrace{\left(\frac{S_n}{\theta^2} + \frac{n-S_n}{(1-\theta)^2}\right)}_{\text{Nenner alle}>0} \le 0. \end{aligned}$$

Wissen:

$0 \le S_n \le n$
$0 \le n - S_n \le n$
Weil $n \ge 1$ und $S_n(n-S_n) = n \ge 1$
ist entweder $S_n > 0$ oder $n - S_n > 0$.
Damit ist $f''(\theta) < 0$
Damit ist $\frac{S_n}{n}$ der eindeutige Maximierer von $L(\theta)$!

Also

$$\hat{\theta}_{ML} = \underset{0<\theta<1}{\operatorname{argmax}} L(\theta) = \bar{x}_n$$

Vor Durchfuhrung ders Experimentes ist

$$\hat{\theta}_{ML} = \bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

der ML-Schätzer.
Beachte: $\mathbb{E}_\theta(X_1) = \theta$.

Bemerkung: Hier stimmen ML- und MM- Schätzer überein.

Stetige Zufallsvariablen:
Betrachte i.i.d. Zufallsvariablen $X_i, i \geq 1$ mit Dichte

$$f(x|\theta_0), \quad \theta_0 \in \Theta \leq \mathbb{R}^p$$

Gegeben Beobachtungen $x_1, \ldots, x_n$ von $X_1, \ldots, X_n$ ist die **Likelihood** definiert als

$$L(\theta) = f(x_1|\theta) \cdot f(x_2|\theta) \cdot \ldots \cdot f(x_n|\theta), \theta \in \Theta.$$

Alternativ zu $L(\theta)$ betrachtet man oft die Log-Likelihood

$$l(\theta) = \log L(\theta) = \sum_{i=1}^{n} \log f(x_i|\theta).$$

Bemerkung: $L(\theta)$ und $l(\theta)$ hängen von den Daten $x_1, \ldots, x_n$ ab.

Der Maximum-Likelihood Schätzer $\hat{\theta}_{ML}$ maximiert $L(\theta)$ bzw. $l(\theta)$ über $\theta \in \Theta$:

$$\begin{aligned}\hat{\theta}_{ML} &= \underset{\theta\in\Theta}{\operatorname{argmax}}\, L(\theta) \\ &= \underset{\theta\in\Theta}{\operatorname{argmax}}\, l(\theta).\end{aligned}$$

Bemerkung: Die Likelihood einer stetigen Zufallsvariable kann als Grenzfall der Likelihoods von diskreten Zufallsvariablen betrachtet werden.

Zur Approximation einer stetigen ZV durch diskrete ZV:

Sei X eine stetige ZV mit Dichte $f(x)$.
Fur $\delta > 0$ definiert man eine diskrete ZV X_δ wie folgt:
Der Wertebereich von X_δ ist $\{k\delta : k \in \mathbb{Z}\}$, und die Wahrscheinlihkeitsfunktion $p_\delta(k\delta)$ ist:

$$\begin{aligned}p_\delta(k\delta) &:= \mathbb{P}(k\delta \leq X < (k+1)\delta) \\ &= \int_{k\delta}^{(k+1)\delta} f(x)dx.\end{aligned}$$

Vergleichen nun die Dichte $f(x)$ mit dem "Histogramm X_δ".
Bei diesem "Histogramm"wird über jedem Intervall $[k\delta,(k+1)\delta)$
ein Balken mit Höhe $\frac{1}{\delta}\mathbb{P}_\delta(k\delta)$ gezeichnet.
Die Fläche eines solcen Balkens ist damit $\mathbb{P}_\delta(k\delta)$.

Fixiert man einen Punkt $x\in\mathbb{R}$, dann liegt dieser Punkt im Intervall
$[k\delta,(k+1)\delta)$ für $k=\lfloor\frac{x}{\delta}\rfloor$(da $\lfloor\frac{x}{\delta}\rfloor\le\frac{x}{\delta}\le\frac{x}{\delta}+1$ ist).
Im Punkt x hat die Dichte den Wert $f(x)$.
Im Punkt x hat das "Histogramm"die Höhe:

$$\begin{aligned}\frac{1}{\delta}\mathbb{P}_\delta(\lfloor\frac{x}{\delta}\rfloor) &= \frac{1}{\delta}\int_{\lfloor\frac{x}{\delta}\rfloor}^{(\lfloor\frac{x}{\delta}\rfloor)+1} f(x)dx = \\ &= \frac{F(\lfloor\frac{x}{\delta}\rfloor\cdot\delta+\delta)-F(\lfloor\frac{x}{\delta}\rfloor\cdot\delta)}{\delta}.\end{aligned}$$

...dies ist ein Differentialquotient.
Für $\delta\downarrow 0$ kovergiet $\lfloor\frac{x}{\delta}\rfloor\cdot\delta+\delta$ von oben gegn x, und $\lfloor\frac{x}{\delta}\rfloor\cdot\delta$ konvergiert von unten gegen x, sodass der Differentialquotient gegen die Ableitung von F an der stelle x konvergiert.
M.a.W:

$$\lim_{\delta\downarrow 0}\frac{1}{\delta}\mathbb{P}_\delta(\lfloor\frac{x}{\delta}\rfloor)=f(x)$$

Beispiel 1.15 (Exponentialverteilung) $X_1,\dots,X_n$ i.i.d. $Exp(\lambda_0)$, $\quad\lambda_0>0$.
D.h. $\mathbb{E}(X_1)=\frac{1}{\lambda_0}$, und Dichte $f(x|\lambda_0)=\lambda_0 e^{-\lambda_0 x}\quad(x\ge 0)$.
Für Beobachtungen $x_1,\dots,x_n$ ist

$$\begin{aligned}L(\lambda) &= \prod_{i=1}^n f(x_i|\lambda)\\ &= \prod_{i=1}^n \lambda e^{-\lambda x_i}\\ &= \lambda^n e^{-\lambda S_n}\end{aligned}$$

wobei $S_n=\sum_{i=1}^n x_i$.

$$l(\lambda)=n\log\lambda-\lambda S_n.$$

$$\begin{aligned}l'(\lambda)=\frac{n}{\lambda}-S &\overset{!}{=}0\\ \Leftrightarrow\frac{n}{\lambda} &=S_n\\ \Leftrightarrow\lambda &=\frac{1}{\frac{n}{S_n}}\end{aligned}$$

Setze:

$$\hat{\lambda}_M L=\frac{n}{S_n}=\frac{1}{\frac{1}{n}S_n}=\frac{1}{\bar{X}_n}.$$

$$l''(\lambda)=\frac{d}{d\lambda}\left(\frac{n}{\lambda}-S\right)=-\frac{n}{\lambda^2}<0.$$

Damit ist der Maximum-Likelihood-Schätzer für λ_0 gegeben durch

$$\hat{\lambda}_{ML} = \frac{1}{\bar{X}_n}.$$

Beachte: Da $\mathbb{E}(X_1) = \frac{1}{\lambda_0}$ ist, also $\lambda_0 = \frac{1}{\mathbb{E}(X_1)}$, ist der Maximum-Likelihood-Schätzer für λ_0 auch ein Momentenmethoden-Schätzer.

Beispiel 1.16 (Gleichverteiltes Skalenmodell) X_i i.i.d. $U([0,\theta_0])$, $\quad \theta_0 > 0$.
Hier ist

$$f(x_1|\theta_0) = \begin{cases} \frac{1}{\theta_0} & : 0 \le x \le \theta_0, \\ 0 & : \text{sonst.} \end{cases}$$

und

$$\mathbb{E}_{\theta_0}(X_i) = \frac{\theta_0}{2}$$

$$L(\theta) = \prod_{i=1}^{n} f(x_i|\theta_0) = \begin{cases} \theta^{-n} & : \text{alle } x_i \text{ zwichen } 0 \text{ und } \theta \\ & \quad\; 0 \le min(x_i) \wedge max(x_i) = \theta \\ 0 & : \text{sonst.} \end{cases}$$

Achtung: $L(\theta)$ ist nicht differenzierbar an der Stelle $\theta = x_{(n)}$.
$L(\theta)$ wird maximal für $\theta = max\{x_i : 1 \le i \le n\}$.
Damit ist der Max-Likelihood Schätzer $\hat{\theta}_M L$ gegeben durch

$$\hat{\theta}_{ML} = max\{x_i : 1 \le i \le n\}$$

$\mathbb{E}(X_1) = \frac{\theta_0}{2}$, also $\theta_0 = 2\mathbb{E}(X_1)$, ist ein MM-Schätzer für θ_0 gegeben durch $\hat{\theta}_{MM} = 2\bar{X}_n$.
Hier ist $\hat{\theta}_{ML} \neq \hat{\theta}_{MM}$.
Bemerkungen:

- In dem Fall, wo $2\bar{X}_n < x_{(n)}$ liefert $\hat{\theta}_{MM}$ einen unsinnigen Wert $\hat{\theta}_{MM} < x_{(n)}$, also ist $\hat{\theta}_M M$ mit dem Daten nicht konsistent.

  Mit $\hat{\theta}_{ML}$ kann genau das nicht passieren.

- Dagegen ist $\hat{\theta}_{MM}$ unverzerrt, $\hat{\theta}_{ML}$ dagegen nicht.

Definition 1.3 Betrachte ein Parametrisches Modell mit Parameter $\theta_0 \in \Theta \le \mathbb{R}$. Für einen Schätzer $\hat{\theta}_n$ ist der **mittlere quadratische Fehler (MSE)** definiert als

$$MSE(\hat{\theta}_n, \theta) = \mathbb{E}_\theta\left[\left(\hat{\theta}_n - \theta\right)^2\right].$$

Bemerkung:

- Ist $\hat{\theta}_n$ ein unverzerrter Schätzer dann ist::

$$MSE(\hat{\theta}_n, \theta) = Var_\theta(\hat{\theta}_n).$$

- Im Allgemeinem ist
$$MSE(\hat\theta_n,\theta) = Var_\theta(\hat\theta_n) + \left(\mathbb{E}_\theta\left(\hat{\theta_n}\right)-\theta\right)^2$$
$$\text{MSE} = \text{Varianz} + \text{Bias}^2.$$

Nachrechnen.
$$\begin{aligned}
E_\theta\left(\left(\hat\theta_n-\theta\right)^2\right) &= \mathbb{E}_\theta\left(\left(\left(\hat\theta_n-\mathbb{E}_\theta\left(\hat\theta_n\right)\right)+\left(\mathbb{E}_\theta\left(\hat\theta_n\right)-\theta\right)\right)^2\right)\\
&= \mathbb{E}_\theta\left(\left(\hat\theta_n-\mathbb{E}_\theta\left(\hat\theta_n\right)\right)^2\right)+2\mathbb{E}_\theta\left(\hat\theta_n-\mathbb{E}_\theta\left(\hat\theta_n\right)\right)\left(\mathbb{E}_\theta\left(\hat\theta_n-\theta\right)\right)+\mathbb{E}_\theta\left(\left(\mathbb{E}_\theta\left(\hat\theta_n\right)-\theta\right)^2\right)\\
&= \underbrace{Var_\theta(\hat\theta_n)}_{\text{Varianz}}+2\left(\mathbb{E}_\theta\left(\hat\theta_n-\theta\right)\underbrace{\mathbb{E}_\theta\left(\hat\theta_n-\mathbb{E}_\theta\left(\hat\theta_n\right)\right)}_{=\mathbb{E}(\hat\theta)-\mathbb{E}(\hat\theta)=0}\right)+\underbrace{\left(\mathbb{E}_\theta\left(\hat\theta_n\right)-\theta\right)^2}_{\text{Bias}^2}.
\end{aligned}$$

□

Beispiel 1.17 $X_1,\dots,X_n$ i.i.d. $U([0,\theta_0])$, $\quad\theta_0>0$.
2 Schätzer:
$$\hat\theta_{MM} = 2\bar X_n \text{ und } \hat\theta_{ML} = \max_{1\le i\le n} X_i = X(n)$$

<u>MSE von $\hat\theta_{MM}$:</u> $\hat\theta_{MM}$ ist unverzerrt, sodass:
$$\begin{aligned}
MSE(\hat\theta_{MM},\theta) = \operatorname{Var}_\theta\left(\hat\theta_{MM}\right) &= \operatorname{Var}_\theta(2\bar X_n)\\
&= 4\operatorname{Var}_\theta(\bar X_n)\\
&= 4\frac{\operatorname{Var}_\theta(X_1)}{n}\\
&= 4\frac{1}{n}\frac{\theta^2}{12}\\
&= \frac{\theta^2}{3n} \xrightarrow{wie\ \frac{1}{n}} 0.
\end{aligned}$$

Für $Z\sim U([0,1])$ ist $\mathbb{E}(Z)=\frac12$ und $\operatorname{Var}(Z)=\frac{1}{12}$.
Nun ist $\theta_0 Z\sim U([0,\theta_0])$, sodass
$$\mathbb{E}(X_1) = \mathbb{E}(\theta_0 Z) = \theta_0\mathbb{E}(Z) = \theta_0\frac12$$
$$\operatorname{Var}(X_1) = \operatorname{Var}(\theta_0 Z) = \theta_0^2\operatorname{Var}(Z) = \theta_0^2\frac{1}{12}.$$

<u>MSE von $\hat\theta_{ML}$:</u> Die ZV $X_{(n)}$ hat die Vertielingsfunktion
$$\begin{aligned}
F_\theta(t) &= \mathbb{P}_\theta\left(X_{(n)}\le t\right)\\
&= \mathbb{P}_\theta\left(X_1\le t,\dots,X_n\le t\right)\\
&= \underbrace{\mathbb{P}_\theta\left(X_1\le t\right)}_{\frac{t}{\theta}}\cdot\ldots\cdot\underbrace{\mathbb{P}\left(X_n\le t\right)}_{\frac{t}{\theta}}\\
&= \left(\frac{t}{\theta}\right)^n \text{ für } 0\le t\le\theta_0.
\end{aligned}$$

$\Rightarrow$ Dichte von $\hat{\theta}_{ML}$:

$$f_\theta(t) = F'(t) = \frac{n \cdot t^{n-1}}{\theta^n}.$$

- $\mathbb{E}_\theta(\hat{\theta}_{ML})$

$$\mathbb{E}_\theta(\hat{\theta}_{ML}) = \mathbb{E}_\theta(X_{(n)}) = \int_0^\theta t f_\theta(t) dt = \int_0^\theta t \frac{nt^{n-1}}{\theta^n} dt = \int_0^\theta \frac{n}{\theta^n} t^n dt = \frac{n}{\theta^n}\Big|_0^\theta =$$

$$= \frac{n}{\theta^n} \frac{\theta^{n+1}}{n+1} = \frac{n}{n+1}\theta < \theta$$

- $Var_\theta(\hat{\theta}_{ML})$

$$Var_\theta\left(\hat{\theta}_{ML}\right) = \mathbb{E}_\theta\left(\left(X_{(n)} - \frac{n}{n+1}\theta\right)^2\right) = \mathbb{E}_\theta\left(X_{(n)}\right)^2 - \left(\frac{n}{n+1}\theta\right)^2 =$$

$$= \int_0^\theta t^2 \frac{nt^{n-1}}{\theta^n} - \left(\frac{n}{n+1}\theta\right)^2 = \frac{n}{\theta^n} \frac{t^{n+2}}{n+2}\Big|_0^\theta - \left(\frac{n}{n+1}\theta\right)^2 = \frac{n}{\theta^n} \frac{\theta^{n+2}}{n+2} - \left(\frac{n^2}{(n+1)^2}\theta^2\right) =$$

$$\frac{n}{n+2}\theta^2 - \left(\frac{n^2}{(n+1)^2}\theta^2\right) = \ldots = \theta^2 \frac{n}{(n+1)^2(n+2)}$$

- Also $MSE\left(\hat{\theta}_{ML}, \theta\right)$

$$MSE\left(\hat{\theta}_{ML}, \theta\right) = \mathbb{E}\left(\left(\hat{\theta}_{ML} - \theta\right)^2\right) = \theta^2 \frac{n}{(n+1)^2(n+2)} + \left(\frac{n}{n+1}\theta - \theta\right)^2$$

$$= \ldots = \frac{2\theta^2}{(n+1)(n+2)} \xrightarrow{wie\ \frac{1}{n^2}} 0.$$

Beispiel 1.18 (Beispiel aus Vorjahr)
$X_1, \ldots, X_n$ i.i.d. $N(\mu_0, \sigma_0^2)$, $\quad \mu_0 \in \mathbb{R}, \sigma_0^2 > 0$.
Hier ist der unbekannte Parameter $\theta_0 = \begin{pmatrix} \mu_0 \\ \sigma_0^2 \end{pmatrix} \in \Theta = \mathbb{R} \times (0, \infty)$.
Die Dichte von X_1 an der Stelle x ist

$$\Phi_{\mu_0, \sigma_0^2}(x) = \frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-\frac{1}{2}\frac{(x-\mu_0)^2}{\sigma_0^2}}.$$

Für Beobachtungen $x_1, \ldots, x_n$ von $X_1, \ldots, X_n$ ist

$$\begin{aligned} L\left(\mu, \sigma^2\right) &= \prod_{i=1}^n \Phi_{\mu, \sigma^2}(x_i) \\ &= \prod_{i=1}^n (2\pi\sigma^2)^{-\frac{1}{2}} e^{-\frac{1}{2}\frac{(x_i-\mu)^2}{\sigma^2}} \\ &= (2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i-\mu)^2}. \end{aligned}$$

$$l\left(\mu,\sigma^2\right) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2.$$

Man sieht:
Für jeden festen Wert von σ^2 wird $l\left(\mu,\sigma^2\right)$ als Funktion von μ minimiert im Punkt $\mu = \bar{X}_n$. (Alternativ: Berechne $\frac{d}{d\mu}l\left(\mu,\sigma^2\right)$, Null setzen, nach μ auflösen.)
$\mu = \bar{X}_n$ einsetzen ergibt

$$l\left(\bar{X}_n,\sigma^2\right) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\bar{X}_n)^2.$$

Um dies in σ^2 zu maximieren, setzt man die 1. Ableitung gleich 0:

$$\begin{aligned}\frac{dl\left(\bar{X}_n,\sigma^2\right)}{d\sigma^2} &= 0 - \frac{n}{2}\frac{1}{\sigma^2} + \frac{1}{2}\frac{1}{\sigma^4}\sum_{i=1}^{n}(x_i-\bar{X}_n)^2 \\ &= -\frac{n}{2}\frac{1}{\sigma^2} + \frac{1}{2}\frac{1}{\sigma^4}\sum_{i=1}^{n}(x_i-\bar{X}_n)^2.\end{aligned}$$

$$\begin{aligned} & & \frac{dl\left(\bar{X}_n,\sigma^2\right)}{d\sigma^2} &\overset{!}{=} 0 \\ \Leftrightarrow & & -\frac{n}{2}\frac{1}{\sigma^2} + \frac{1}{2}\frac{1}{\sigma^4}\sum_{i=1}^{n}(x_i-\bar{X}_n)^2 &= 0 \\ \Leftrightarrow & & \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\bar{X}_n)^2 &= n \\ \Leftrightarrow & & \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{X}_n)^2 &= \sigma^2 \\ & & &:= \tilde{\sigma}^2.\end{aligned}$$

Ist $\tilde{\sigma}^2$ ein Maximierer?

$$\begin{aligned}\frac{d^2l\left(\bar{X}_n,\sigma^2\right)}{d\left(\sigma^2\right)^2} &= \frac{d}{d\sigma^2}\left(-\frac{n}{2}\frac{1}{\sigma^2} + \frac{1}{2}\frac{1}{\sigma^4}\sum_{i=1}^{n}(x_i-\bar{X}_n)^2\right) \\ &= \frac{n}{2}\frac{1}{\left(\sigma^2\right)^2} - \frac{1}{2}2\frac{1}{\left(\sigma^2\right)^3}\sum_{i=1}^{n}(x_i-\bar{X}_n)^2 \\ &= \frac{n}{2}\frac{1}{\left(\sigma^2\right)^2} - \frac{n}{\left(\sigma^2\right)^3}\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{X}_n)^2\end{aligned}$$

Setzt man hier für σ^2 den Wert $\tilde{\sigma}^2$ ein, dann erhält man

$$\frac{n}{2}\frac{1}{\left(\tilde{\sigma}^2\right)^2} - \frac{n}{\left(\tilde{\sigma}^2\right)^3}\tilde{\sigma}^2 = \frac{n}{2}\frac{1}{\left(\tilde{\sigma}^2\right)^2} - \frac{n}{\left(\tilde{\sigma}^2\right)^2} = \frac{n}{\tilde{\sigma}^4}\left(\frac{1}{2}-1\right) < 0.$$

Also Maximum bei $\sigma^2 = \tilde{\sigma}^2$.

Damit ist der Maximum-Likelihood-Schätzer für $\begin{pmatrix}\mu_0\\ \sigma_0^2\end{pmatrix}$ gegeben durch

$$\begin{pmatrix}\hat{\mu}_{ML}\\ \hat{\sigma}_{ML}^2\end{pmatrix} = \begin{pmatrix}\bar{X}_n\\ \frac{1}{n}\sum_{i=1}^n (x_i - \bar{X}_n)^2\end{pmatrix}.$$

Beachte: Das ist auch ein Momentenmethoden-Schätzer für $\begin{pmatrix}\mu_0\\ \sigma_0^2\end{pmatrix}$.

Beispiel 1.19 (Capture/Recapture-Methode) (Beispiel aus Vorjahr)
Schätzung der Populationsgröße mit der Capture/Recapture Methode.
Gegeben eine Population von N Individuen.
Zur Vorbereitung des Experiments werden davon m Individuen markiert (Capture).
Im Experiment selbst werden n Individuen zufällig ausgewählt, und die Anzahl der davon markierten Individuen wird ermittelt (Recapture).
Modell:

$$X \sim HG(n, N_0, m)$$

Anzahl der gezogenen markierten Kugeln

n Kugeln aus Urne ohne zurücklegen ziehen

Anzahl der Kugeln in Urne

Anzahl der markierten Kugeln

N_0 ist unbekannt. $N_0 \geq \max\{m, n\}$, $N_0 \in \mathbb{N}$.
Es ist $\mathbb{E}(X) = \frac{m\cdot n}{N_0}$, also $N_0 = \frac{m\cdot n}{\mathbb{E}(X)}$.
Damit ist ein MM-Schätzer für N_0 gegeben durch

$$\hat{N}_{MM} = \frac{m\cdot n}{X}.$$

Beachte: Im Allgemeinem ist $\hat{N}_{MM} \notin \mathbb{N}$.
Suche nun ML-Schätzer.
Gegeben eine Beobachtung x von X ist

$$\begin{aligned} L(N) &= \mathbb{P}_N(X = x)\\ &= \frac{\binom{m}{x}\binom{N-m}{n-x}}{\binom{N}{n}}. \end{aligned}$$

$$\begin{aligned} D(N) &= \frac{L(N)}{L(N-1)}\\ &= \frac{\binom{m}{x}\binom{N-m}{n-x}}{\binom{N}{n}} \cdot \frac{\binom{N-1}{n}}{\binom{m}{x}\binom{N-1-m}{n-x}}\\ &= \frac{(N-m)!}{(n-x)!(N-m-n+x)!} \frac{n!(N-n)!}{N!} \frac{(N-1)!}{n!(N-1-n)!} \frac{(n-x)!(N-1-m-n+x)!}{(N-1-m)!}\\ &= \frac{(N-n)(N-m)}{N(N-m-n+x)}. \end{aligned}$$

Es ist

$$\begin{aligned} & D(N) > 1 \\ \Leftrightarrow \quad & \frac{(N-n)(N-m)}{N(N-m-n+x)} > 1 \\ \Leftrightarrow \quad & (N-n)(N-m) > N(N-m-n+x) \\ \Leftrightarrow \quad & N^2 - nN - mN + mn > N^2 - mN - nN + xN \\ \Leftrightarrow \quad & N < \frac{m \cdot n}{x}. \end{aligned}$$

Also:

- Für $N < \frac{m \cdot n}{x}$ ist $D(N) > 1$
- Für $N > \frac{m \cdot n}{x}$ ist $D(N) > 1$
- Für $N = \frac{m \cdot n}{x}$ ist $D(N) = 1$.

Zwei Fälle:

- Fall 1: $\frac{m \cdot n}{x} \notin \mathbb{N}$.
  Wähle $N \in \mathbb{N}$, sodass $\frac{m \cdot n}{x} \in (N, N+1)$.
  Damit ist $D(N) > 1$ und auch $D(\tilde{N}) > 1$ für alle $\tilde{N} < \frac{m \cdot n}{x}$.
  Analog ist $D(N+1) < 1$ und auch $D(\tilde{N}) < 1$ für alle $\tilde{N} > \frac{m \cdot n}{x}$.
  Schließlich ist $D(N+1) = \frac{L(N+1)}{L(N)} < 1$, also $L(N+1) < L(N)$.
  Damit wird das Maximum von $L(\cdot)$ angenommen an der Stelle $N = \lfloor \frac{m \cdot n}{x} \rfloor$.
- Fall 2: $\frac{m \cdot n}{x} \in \mathbb{N}$.
  Setze $N = \frac{m \cdot n}{x}$.
  Wie im Fall 1 ist $D(\tilde{N}) > 1$ für $\tilde{N} < N$,
  $D(\tilde{N}) < 1$ für $\tilde{N} > N$.
  Schließlich ist $1 = D(N) = \frac{L(N)}{L(N-1)}$.
  $\Rightarrow L(N) = L(N-1)$.
  Hier wird $L(\cdot)$ maximiert an den Stellen $\frac{mn}{x}$ sowie $\frac{mn}{x} - 1$.

Zusammenfassend:

$$\hat{N}_{ML} = \lfloor \frac{mn}{x} \rfloor,$$

wobei $\hat{N}_{ML}$ nicht eindeutig ist wenn $\frac{mn}{x} \in \mathbb{N}$.

1.4 Zulässigkeit, Effizienz und die Cramér-Rao-Schranke

Siehe zu diesem Kapitel S. 298-306. in *Rice (2007)*.

Was zeichnet einen "guten" Schätzer aus?

Betrachte durchwegs ein parametrisches Modell: $X_1,\dots,X_n$ i.i.d. mit Dichte $f(x|\theta_0)$ (bzw. Wahrscheinlichkeitsfunktion $p(x|\theta_0)$ für $\theta \in \Theta \in \mathbb{R}^k$).
Gegeben einen Schätzer $\hat{\theta}_n = \hat{\theta}_n(X_1,\dots,X_n) \in \mathbb{R}^k$ für θ_0 betrachtet man den **mittleren quadratischen Fehler**

$$MSE\left(\hat{\theta}_n,\theta_0\right) := \mathbb{E}_{\theta_0}\|\hat{\theta}_n - \theta_0\|^2 = \left(\sum_{i=1}^{k}\mathbb{E}_{\theta_0}\left(\hat{\theta}_n - \theta_0\right)_i^2\right) = \sum_{i=1}^{k} MSE_{\theta_0}\left(\left((\hat{\theta}_n)_i,(\theta_0)_i\right)^2\right)$$

Beachte: $MSE(\hat{\theta}_n,\theta)$ ist eine Funktion von θ.

Beim Vergleich zweier Schätzer $\hat{\theta}_n$ und $\tilde{\theta}_n$ muss man also die Funktionen $MSE(\hat{\theta}_n,\cdot)$ und $MSE(\tilde{\theta}_n,\cdot)$ vergleichen.

Beispiel 1.20

A	B	C	D
4	2	1	2
2	3	1	1
1	1	1	1
3	3	1	2

D ist besser als A.
A und B sind nicht vergleichbar.
C ist klassenbeste.

Im Folgenden sei κ eine Klasse von Schätzern für θ_0, also Funktionen von $X_1,\dots,X_n$ die Werte in Θ einnehmen. Oft ist κ die Klasse aller Schätzer bzw. die Klasse aller unverzerrter Schätzer.

Definition 1.4 Sei $\hat{\theta}$ ein Schätzer aus der Klasse κ.

- $\hat{\theta}$ heißt **unzulässig** (in der Klasse κ), wenn es einen Schätzer $\tilde{\theta}$ aus κ gibt sodass: $\forall\theta \in \Theta$ ist $MSE(\tilde{\theta},\theta) \le MSE(\hat{\theta},\theta)$, und $\exists\theta \in \Theta$ sodass $MSE(\tilde{\theta},\theta) < MSE(\hat{\theta},\theta)$.
- $\hat{\theta}$ heißt **zulässig** (in der Klasse κ), wenn der vorherige Punkt nicht gilt.

Ist $\kappa = \{A,B,C,D\}$ dann ist C zulässig.
Ist $\kappa = \{A,D\}$ dann ist A unzulässig und D ist zulässig.
Ist $\kappa = \{A,B\}$ dann sind A und B beide zulässig.

Definition 1.5 Sei $\hat{\theta}$ ein Schätzer aus der Klasse κ. $\hat{\theta}$ heißt **effzient** (in der Klasse κ), wenn gilt:

$$\forall\theta \in \Theta \text{ ist } MSE\left(\hat{\theta},\theta\right) = \min_{\tilde{\theta}\in\kappa} MSE\left(\tilde{\theta},\theta\right).$$

Ist $\kappa = \{A,B,C,D\}$ dann ist C effizient.
Ist $\kappa = \{A,B,D\}$ dann ist D effizient.
Ist $\kappa = \{A,B\}$ dann sind niemand effizient.

Bemerkung: Effiziente Schätzer sind selten bekannt und müssen nicht unbedingt existieren. Aber es gibt einige Ausnahmen, die Konzeptionell wichtig sind.
Oft kann man aber ”asymptotisch effiziente” Schätzer finden.(für $n \to \infty$)

Bemerkung:

- effizient $\Rightarrow$ zulässig; aber nicht umgekehrt.
- Diese beiden Begriffe hängen von κ und Θ ab.

Betrachte bis auf weiteres die Klasse κ derr unverzerrten Schätzer.

Definition 1.6 Sei X eine Zufallsvariable X mit Dichte $f(x|\theta_0)$, $\theta_0 \in \Theta \subseteq \mathbb{R}^1$.
Die Größe

$$I(\theta_0) := \mathbb{E}_{\theta_0}\left[\left(\frac{d}{d\theta}\log f(x|\theta_0)\right)^2\right]$$

ist die Fisher-Information (von X über θ_0), sofern diese wohldefiniert ist.

Satz 1.4 (Cramér-Rao-Schranke) Seien $X_1,\dots,X_n$ i.i.d mit Dichte $f(x|\theta_0)$, $\theta_0 \in \Theta \subseteq \mathbb{R}^1$.
Weiters sei $\hat{\theta}$ ein unverzerrter Schätzer für θ_0.
Dann gilt:

$$\mathrm{Var}_{\theta_0}(\hat{\theta}) \geq \frac{1}{n_0 I(\theta_0)},$$

unter geeigneten Glattheitsbedingungen an $f(\cdot|\cdot)$, sodass die Schritte $(\ast)$ und $(\ast\ast)$ im Beweis gültig sind.

Korollar 1.1 Ist $\hat{\theta}$ ein unverzerrter Schätzer für θ_0, sodass $\forall \theta \in \Theta$, $\mathrm{Var}_\theta(\hat{\theta}) = \frac{1}{nI(\theta)}$, dann ist $\hat{\theta}$ effizient in der Klasse der unverzerrten Schätzer.

Es gibt auch eine multivariate Version der Cramér-Rao-Schranke, also für unverzerrte Schätzer, wo θ_0 hochdimensional sein kann.

Beweis. Setze $Z = \sum_{i=1}^n \frac{d}{d\theta}\log f(x_i|\theta_0)$.
Es gilt $|\mathrm{Corr}_{\theta_0}(Z,\hat{\theta})| \leq 1$, also $\frac{\mathrm{Cov}_{\theta_0}(Z,\hat{\theta})^2}{\mathrm{Var}_{\theta_0}(Z)\mathrm{Var}_{\theta_0}(\hat{\theta})} \leq 1$, bzw. $\frac{\mathrm{Cov}_{\theta_0}(Z,\hat{\theta})^2}{\mathrm{Var}_{\theta_0}(Z)} \leq \mathrm{Var}_{\theta_0}(\hat{\theta})$.
Zu zeigen:
(i) $\mathrm{Cov}_{\theta_0}(Z,\hat{\theta}) = 1$
(ii) $\mathrm{Var}_{\theta_0}(Z) = nI(\theta_0)$.

$$
\begin{aligned}
\mathbb{E}_{\theta_0}(Z_i) &\overset{i.i.d.}{=} n\mathbb{E}_{\theta_0}\left(\frac{d}{d\theta}\log f(x_1|\theta_0)\right) \\
&= n\mathbb{E}_{\theta_0}\left(\frac{d}{d\theta}\Big|_{\theta=\theta_0}\log f(x_1|\theta_0)\right) \\
&= n\int \frac{d}{d\theta}\Big|_{\theta=\theta_0}\log f(x|\theta) f(x|\theta_0)dx \\
&= n\int \frac{\frac{d}{d\theta}\Big|_{\theta=\theta_0} f(x|\theta)}{f(x|\theta_0)} f(x|\theta_0)dx \\
&= n\int \frac{d}{d\theta}\Big|_{\theta=\theta_0} f(x|\theta)dx \\
&\overset{(\divideontimes)}{=} \underbrace{n\frac{d}{d\theta}\Big|_{\theta=\theta_0}\underbrace{\int f(x|\theta)dx}_{=1}}_{=0} = n\cdot 0 = 0.
\end{aligned}
$$

Man sieht damit auch, dass $\mathbb{E}_{\theta_0}\left(\frac{d}{d\theta}\Big|_{\theta=\theta_0}\log f(x_1|\theta)\right) = 0$ ist.

$$
\begin{aligned}
\mathrm{Var}_{\theta_0}(Z) &= \sum_{i=1}^{n}\mathrm{Var}_{\theta_0}\left(\frac{d}{d\theta}\log f(x_1|\theta_0)\right) \\
&= n\,\mathrm{Var}_{\theta_0}\left(\frac{d}{d\theta}\log f(x_1|\theta_0)\right) \\
&= n\mathbb{E}_{\theta_0}\left(\left(\frac{d}{d\theta}\log f(x_1|\theta_0)\right)^2\right) \\
&= nI(\theta_0).
\end{aligned}
$$

Damit ist (ii) gezeigt. ✓

$$
\begin{aligned}
\operatorname{Cov}_{\theta_0}(Z,\hat\theta) &= \mathbb{E}_{\theta_0}(Z\hat\theta) - \overbrace{\mathbb{E}_{\theta_0}(Z)}^{=0}\mathbb{E}_{\theta_0}(\hat\theta) \\
&= \underbrace{\int \dots \int}_{n-\text{mal}} \hat\theta(x_1\dots,x_n)\left(\sum_{i=1}^n \frac{d}{d\theta}\log f(x_i|\theta_0)\right)\prod_{j=1}^n f(x_j|\theta_0)dx_1,\dots,dx_n \\
&= \int \dots \int \hat\theta(x_1\dots,x_n)\left(\sum_{i=1}^n \frac{\frac{d}{d\theta}f(x_i|\theta_0)}{f(x_i|\theta_0)}\right)\prod_{j=1}^n f(x_j|\theta_0)dx_1,\dots,dx_n \\
&= \int \dots \int \hat\theta(x_1\dots,x_n)\left(\sum_{i=1}^n \frac{d}{d\theta}f(x_i|\theta_0)\right)\prod_{j=1,i\neq i}^n f(x_j|\theta_0)dx_1,\dots,dx_n \\
&= \int \dots \int \hat\theta(x_1\dots,x_n)\frac{d}{d\theta}\Big|_{\theta=\theta_0}\prod_{j=1}^n f(x_j|\theta_0)dx_1,\dots,dx_n \\
&\overset{(\ast\ast)}{=} \frac{d}{d\theta}\Big|_{\theta=\theta_0}\int \dots \int \hat\theta(x_1\dots,x_n)\prod_{j=1}^n f(x_j|\theta_0)dx_1,\dots,dx_n \\
&= \frac{d}{d\theta}\Big|_{\theta=\theta_0}\mathbb{E}_{\theta_0}(\hat\theta) \\
&= \frac{d}{d\theta}\Big|_{\theta=\theta_0}\theta \\
&= 1.
\end{aligned}
$$

Also gilt auch (i). ✓ □

Eine alternative Formel für $I(\theta_0)$:

Lemma 1.2 Die **Fisher-Information** lässt sich auch berechnen als

$$I(\theta) = -\mathbb{E}(\theta)\left(\frac{d^2}{d\theta^2}\log f(x|\theta)\right)$$

unter geeigneten Glattheitsbedingungen, sodass Schritte (a) und (b) im Beweis zulässig sind.

Beweis. Für jedes $\theta \in \Theta$ ist $\int f(x|\theta)dx = 1$.

$$\begin{aligned}
0 &= \frac{d}{d\theta}\int f(x|\theta)dx \\
&\overset{(a)}{=} \int \frac{d}{d\theta} f(x|\theta)dx \\
&= \int \underbrace{\frac{\frac{d}{d\theta}f(x|\theta)}{f(x|\theta)}}_{\left(\frac{d}{d\theta}\log f(x|\theta)\right)} f(x|\theta)dx \\
&= \int \left(\frac{d}{d\theta}\log f(x|\theta)\right) f(x|\theta)dx = (c).
\end{aligned}$$

$$\begin{aligned}
\Rightarrow 0 &= \frac{d}{d\theta}(c) \\
&= \frac{d}{d\theta}\int \left(\frac{d}{d\theta}\log f(x|\theta)\right) f(x|\theta)dx \\
&\overset{(b)}{=} \int \left(\frac{d^2}{d\theta^2}\log f(x|\theta)\right) f(x|\theta) + \left(\frac{d}{d\theta}\log f(x|\theta)\right) \underbrace{\frac{d}{d\theta}f(x|\theta)}_{\overset{s.o}{=}\left(\frac{d}{d\theta}\log f(x|\theta)\right)f(x|\theta)} dx \\
&= \int \left(\frac{d^2}{d\theta^2}\log f(x|\theta)\right) f(x|\theta)dx + \int \left(\frac{d}{d\theta}\log f(x|\theta)\right)^2 f(x|\theta)dx \\
&= \mathbb{E}_\theta\left(\frac{d^2}{d\theta^2}\log f(X|\theta)\right) + \underbrace{\mathbb{E}_\theta\left(\left(\frac{d}{d\theta}\log f(X|\theta)\right)^2\right)}_{=I(\theta)}
\end{aligned}$$

$$\Rightarrow I(\theta) = -\mathbb{E}_{\theta_0}\left(\frac{d^2}{d\theta^2}\log f(X|\theta)\right).$$

□

Bemerkung: Die Cramér-Rao-Schranke gilt auch für diskrete Zufallsvariablen:
Sind $X_1, \ldots, X_n$ i.i.d. diskrete Zufallsvariablen, verteilt wie X, mit Wahrscheinlichkeitsfunktion $p(x|\theta)$, $\theta \in \Theta \subseteq \mathbb{R}$, und ist $\hat{\theta_n} = \hat{\theta_n}(X_1, ..., X_n)$ ein unverzerrter Schätzer für θ, dann gilt

$$\operatorname{Var}\hat{\theta_n} \geq \frac{1}{nI(\theta)},$$

unter analogen Glattheitsbedingungen wie im letzten Satz.
Details: Siehe Übung.

Beispiel 1.21 (Bernoulli-Verteilung) Seien $X_1, \dots, X_n$ i.i.d $B(p_0)$, $0 < p_0 < 1$.
Wissen: $\hat{p}_{ML} = \bar{X}_n = \hat{p}_{MM} = \frac{1}{n}\sum_{i=1}^n x_i$
ist unverzerrt für p_0 mit Varianz $\mathrm{Var}_{p_0}(\bar{X}_n) = \frac{\mathrm{Var}_{p_0}(X_1)}{n} = \frac{p_0(1-p_0)}{n}$.

$$\begin{aligned} p(x|p_0) &= \begin{cases} p_0 & : x = 1, \\ 1 - p_0 & : x = 0 \end{cases} \\ &= p_0^x (1-p_0)^{1-x},\ x \in \{0,1\},\ 0 < p_0 < 1. \end{aligned}$$

$$\begin{aligned} \frac{d}{dp_0} \log p(x|p_0) &= \frac{d}{dp_0}(x \log p_0 + (1-x)\log(1-p_0)) \\ &= x\frac{1}{p_0} + (1-x)\frac{-1}{1-p_0} = \frac{x}{p_0} + \frac{1-x}{1-p_0} \end{aligned}$$

$$\begin{aligned} I(p_0) &= \mathbb{E}_{p_0}\left(\left(\frac{X}{p_0} + \frac{1-X}{1-p_0})\right)^2\right) \\ &= \frac{1}{p_0^2}p_0 + \frac{1}{(1-p_0)^2}(1-p_0) \\ &= \frac{1}{p_0} + \frac{1}{1-p_0} = \frac{1-p_0+p_0}{p_0(1-p_0)} \\ &= \frac{1}{p_0(1-p_0)}. \end{aligned}$$

Die Cramér-Rao-Schranke in diesem Modell ist also

$$\frac{1}{nI(p_0)} = \frac{p_0(1-p_0)}{n} = Var_{p_0}(\bar{X}_n)$$

Der Schätzer $\bar{X}_n$ für p_0 ist also effizient (in der Klasse der unverzerrten Schätzer) !

Beispiel 1.22 (Exponentialverteilung) $X_1, \dots, X_n$ i.i.d. verteilt wie X, wobei X eine stetige Zufallsvariable ist mit $X\ E_\lambda(\lambda_0)$, $\lambda_0 > 0$.
Dichte $f(x|\lambda_0) = \lambda_0 e^{-\lambda_0 x}$, $(x > 0)$.
$\log f(x|\lambda_0) = \log\lambda_0 - \lambda_0 x$.
$\frac{d}{d\lambda}\log f(x|\lambda_0) = \frac{1}{\lambda_0} - x$.
Erinnerung: $E_{\lambda_0}(X) = \frac{1}{\lambda_0}$, $Var_{\lambda_0}(X) = \frac{1}{\lambda_0^2}$

$$\begin{aligned} I(\lambda_0) &= \mathbb{E}_{\lambda_0}\left(\left(\frac{d}{d\lambda}\log f(x|\lambda_0)\right)^2\right) \\ &= \mathbb{E}_{\lambda_0}\left(\left(\frac{1}{\lambda_0} - X\right)^2\right) = \mathrm{Var}_{\lambda_0}(X) = \frac{1}{\lambda_0^2}. \end{aligned}$$

Cramér-Rao-Schranke:

$$\frac{1}{(nI(\lambda_0))} = \frac{\lambda_0^2}{n}.$$

Wissen: Der Schätzer $\frac{1}{\bar{X}_n} = \frac{1}{\frac{1}{n}\sum_{i=1}^n X_i} = \hat{\lambda}_{MM} = \hat{\lambda}_{ML}$
Aber dieser Schätzer ist verzerrt! Cramer-Rao-Schranke nicht anwendbar!

Beispiel 1.23 (Umparametrisierung der Exponentialverteilung) $X_1,\dots,X_n$ i.i.d. $Exp\left(\frac{1}{\theta_0}\right)$, $\theta_0 > 0$.

Wissen: $\mathbb{E}_{\theta_0}(X) = \frac{1}{\frac{1}{\theta_0}} = \theta_0$, $\mathrm{Var}_{\theta_0}(X) = \left(\frac{1}{\frac{1}{\theta_0}}\right)^2 = \theta_0^2$.
Dichte: $f(x|\theta_0) = \frac{1}{\theta_0} e^{-\frac{x}{\theta_0}}$, $(x > 0)$.
$\log f(x|\theta_0) = -\log\theta_0 - \frac{x}{\theta}$.
$\frac{d}{d\theta_0}\log f(x|\theta_0) = -\frac{1}{\theta_0} + \frac{x}{\theta_0^2}$.

$$\begin{aligned} I(\theta_0) &= \mathbb{E}_{\theta_0}\left(\left(\frac{d}{d\theta_0}\log f(x|\theta_0)\right)^2\right) \\ &= \mathbb{E}_{\theta_0}\left(\left(\frac{X-\theta_0}{\theta_0^2}\right)^2\right) \\ &= \frac{1}{\theta_0^4}\mathbb{E}_{\theta_0}\left((X-\theta_0)^2\right) \\ &= \frac{1}{\theta_0^4}\mathrm{Var}_{\theta_0}(X) \\ &= \frac{1}{\theta_0^4}\theta_0^2 = \frac{1}{\theta_0^2}. \end{aligned}$$

Cramér-Rao-Schranke:

$$\frac{1}{(nI(\theta_0))} = \frac{\theta_0^2}{n}.$$

Beachte $\bar{X}_n$: $\mathbb{E}_{\theta_0}(\bar{X}_n) = \mathbb{E}_{\theta_0}(X) = \theta_0$,
$\mathrm{Var}_{\theta_0}(\bar{X}_n) = \frac{\mathrm{Var}_{\theta_0}(X_1)}{n} = \frac{\theta_0^2}{n}$.
Der Schätzer $\bar{X}_n$ für θ_0 ist also unverzerrt und erreicht die Cramer-Rao-Schranke. $\bar{X}_n$ ist effizient (in der Klasse der unverzerrten Schätzer).

Bemerkung: Die Effizienz von Schätzern kann von der Parametrisierung abhängen.

Beispiel 1.24 (Normalverteilung - Varianz bekannt) $X_1,\dots,X_n$ i.i.d. $N(\mu_0,\sigma^2)$, $\mu_0 \in \mathbb{R}$, $\sigma^2 > 0$ bekannt.
Wissen: $\bar{X}_n = \hat{\mu}_{ML} = \hat{\mu}_{MM}$ ist unverzerrt mit $\mathrm{Var}_\mu(\bar{X}_n) = \frac{\sigma^2}{n}$.
Dichte: $f(x|\mu_0) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu_0)^2}$.
$\log f(x|\mu_0) = -\frac{1}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}(x-\mu_0)^2$.
$\frac{d}{d\mu_0}\log f(x|\mu_0) = \frac{1}{2\sigma^2}2(x-\mu_0) = \frac{x-\mu_0}{\sigma^2}$.

$$\begin{aligned} I(\mu_0) &= \mathbb{E}_{\mu_0}\left(\left(\frac{d}{d\mu_0}\log f(X|\mu_0)\right)^2\right) \\ &= \mathbb{E}_{\mu_0}\left(\left(\frac{X-\mu_0}{\sigma^2}\right)^2\right) \\ &= \frac{1}{\sigma^4}\operatorname{Var}_{\mu_0}(X) \\ &= \frac{\sigma^2}{\sigma^4} = \frac{1}{\sigma^2}. \end{aligned}$$

Cramér-Rao-Schranke:

$$\frac{1}{nI(\mu_0)} = \frac{\sigma^2}{n} = \operatorname{Var}_{\mu_0}(\bar{X}_n).$$

$\bar{X}_n$ ist damit effizient (in der Klasse der unverzerrten Schätzer).

Bemerkung: $\bar{X}_n$ ist effizient in der Klasse

$$\kappa_+ = \left\{{}_n\hat{\mu} = \hat{\mu}_n(X_1,\ldots,X_n,\sigma^2) \colon \hat{\mu}_n \text{ unverzerrt}\right\}$$

$\bar{X}_n$ hängt nicht von σ^2 ab.

Beispiel 1.25 (Normalverteilung - Beide unbekannt) $X_1,\ldots,X_n$ i.i.d. $N(\mu_0,\sigma_0^2)$, $\mu_0 \in \mathbb{R}$, $\sigma_0^2 > 0$.
Sei κ die Klasse aller unverzerrten Schätzer für μ_0:

$$\kappa = \{\hat{\mu} = \hat{\mu}_n(X_1,\ldots,X_n) \colon \hat{\mu}_n \text{ unverzerrt}\}$$

Wissen:
(1) $\bar{X}_n \in \kappa$ (unverzerrt),
(2) $\operatorname{Var}_{\mu_0}(\bar{X}_n) = \min\limits_{\hat{\mu}\in\kappa_+} \operatorname{Var}_{\mu_0}(\hat{\mu}_n)$,
(3) $\kappa \subseteq \kappa_+$.
Also:

$$\begin{aligned} \operatorname{Var}_{\mu_0,\sigma_0^2}(\bar{X}_n) &\underset{(2)}{=} \min_{\hat{\mu}_n\in\kappa_+} \operatorname{Var}_{\mu_0,\sigma_0^2}(\hat{\mu}_n) \\ &\underset{(3)}{\leq} \min_{\hat{\mu}_n\in\kappa} \operatorname{Var}_{\mu_0,\sigma_0^2}(\hat{\mu}) \\ &\underset{(1)}{\leq} \operatorname{Var}_{\mu_0,\sigma_0^2}(\bar{X}_n). \end{aligned}$$

$\Rightarrow \operatorname{Var}_{\mu_0,\sigma_0^2}(\bar{X}_n) = \min\limits_{\hat{\mu}_n\in\kappa} \operatorname{Var}_{\mu_0,\sigma_0^2}(\hat{\mu}_n)$
Also ist $\bar{X}_n$ effizient für μ_0 auch im Fall unbekannter Varianz. (in der Klasse κ).

Bemerkung: Es gibt eines Version der Cramér-Rao-Schranke für mehr-dimensionale Parameter θ_0 und für den Fall wo die $X_1,...,X_n$ Zufallsvektoren sind.
Insbesondere gilt:
Sei $X_1,\ldots,X_n$ i.i.d. Zufallsvektoren mit X_1 $N(\mu,\sigma^2 I_p)$ mit $\mu_0 \in \mathbb{R}^p$, $\sigma_0^2 > 0$ (p-dimensional), dann ist $\bar{X}_n$ effizient.(in der Klasse der unverzerrten Schätzer für μ_0).

Betrachte nun auch möglicherweise verzerrte Schätzer.

Proposition 1.1 Betrachte $X \sim N(\mu_0, I_p)$ mit $\mu_0 = \mathbb{R}^p$. Der Maximum-Likelihood-Schätzer $\hat{\mu}_{ML} = X$ für μ_0 ist zulässig in der Klasse aller Schätzer für μ_0, wenn $p = 1, 2$.

(Ohne Beweis).
? Was passiert für $p \geq 3$?

Überraschung: (James, Stein, 1961)
Für $X \sim N(\mu_0, I_p)$, $\mu_0 \in \mathbb{R}^p$, $p \geq 3$, gibt es einen Schätzer $\hat{\mu}_{JS}$ für μ_0, sodass gilt:

$$\forall \mu_0 \in \mathbb{R}^p : MSE(\underset{\text{J-S-Schätzer}}{\hat{\mu}_{JS}}, \mu_0) < MSE(\underset{\text{ML-Schätzer}}{X}, \mu_0)$$

$$MSE(\hat{\mu}_{JS}, 0) = 2, \ MSE(X, 0) = p.$$

Dieser Schätzer ist gegeben durch

$$\hat{\mu}_{JS} = \underbrace{\left(1 - \frac{p-2}{X'X}\right)}_{\text{Kontraktionsfaktor ¡1}} \underset{\text{ML-Schätzer}}{X}.$$

Insbesondere ist der Maximum-Likelihood-Schätzer in diesem Modell unzulässig wenn $p \geq 3$.

Bemerkung:
In der Praxis wird der James-Stein-Schätzer fast nie verwendet. Aber als so genannter Shrinkage-Schätzer lieferte $\hat{\mu}_{JS}$ die Inspiration für zahlreiche moderne Methoden wie LASSO, SVM, LARS, Dantzig-Selektor, ...

Zur Motivation des James-Stein-Schätzers:
$X \sim N(\mu, I_p)$, $\hat{\mu}_{JS} = \left(1 - \frac{p-2}{X'X}\right)X$, $(p \geq 3)$.
Betrachte eine Approximation, wo $p \longrightarrow \infty$ geht, wobei

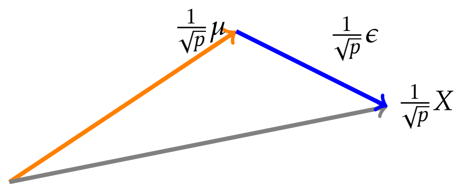
$$\underbrace{\frac{\overset{\text{Signal}}{\|\mu\|^2}}{\underset{\text{Noise}}{p}}}_{singal-to-noice\ ratio} \xrightarrow{p\to\infty} \rho^2 > 0 \in [0, 1].$$

Zerlege X als $X = \mu + \epsilon$ (für $\epsilon = X - \mu$), sodass $\epsilon \sim N(0, I_p)$.

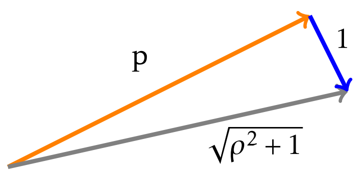
Beachte: Als Schätzer für μ ist X (im Mittel) ”zu lang”:

$$\begin{aligned}
\text{Länge von } \mu : \|\mu\|^2 & \\
\text{Länge von } X : \mathbb{E}(\|X\|^2) &= \mathbb{E}(X'X) \\
&= \mathbb{E}\left((\mu+\epsilon)'(\mu+\epsilon)\right) \\
&= \mathbb{E}\left(\mu'\mu + 2\mu'\epsilon + \epsilon'\epsilon\right) \\
&= \mu'\mu + 0 + \underbrace{\mathbb{E}(\epsilon'\epsilon)}_{\mathbb{E}(\sum_{i=1}^p \epsilon_i^2)=p} \\
&= \mu'\mu + p > \mu'\mu\,!
\end{aligned}$$

Betrachte das Dreieck $0, \mu, X$ bzw- $0, \frac{1}{\sqrt{p}}\mu, \frac{1}{\sqrt{p}}X$:



Approximatives Dreieck ($p \to \infty$) (rechtwinklig):



$\|\frac{1}{\sqrt{p}}\mu\|^2 = \frac{\mu'\mu}{p} \to \rho^2$.
$\|\frac{1}{\sqrt{p}}\epsilon\|^2 = \frac{1}{p}\epsilon'\epsilon = \frac{1}{p}\sum_{i=1}^p \epsilon_i^2 \xrightarrow{p} 1$ (LLN).
$\|\frac{1}{\sqrt{p}}X\|^2 = \frac{1}{p}(\mu+\epsilon)'(\mu+\epsilon) = \underbrace{\frac{1}{p}\mu'\mu}_{\to\rho^2} + \underbrace{\frac{2}{p}\mu'\epsilon}_{(\ast)} + \underbrace{\frac{1}{p}\epsilon'\epsilon}_{\xrightarrow{p}1} \xrightarrow{p} \rho^2 + 1.$
$(\ast) \sim N\left(0, \frac{4}{p^2}\mu'\mu\right)$, da $\operatorname{Var}(\mu'\epsilon) = \mu'\operatorname{VC}(\epsilon)\mu = \mu'\mu$.
Damit gilt: Für $c > 0$ ist

$$\begin{aligned}
\mathbb{P}\left(|\frac{2}{p}\mu'\epsilon| > c\right) &\underset{Chebyshev}{\leq} \frac{1}{c^2}\operatorname{Var}\left(\frac{2}{p}\mu'\epsilon\right) \\
&= \frac{1}{c^2}\frac{3}{p^2}\mu'\mu \\
&= \underbrace{\frac{4}{c^2}\frac{\mu'\mu}{p}}_{\to\frac{4}{c^2}\rho^2}\underbrace{\frac{1}{p}}_{\to 0} \\
&\xrightarrow{p\to\infty} 0.
\end{aligned}$$

Also ist $(\ast) = \frac{2}{p}\mu'\epsilon \xrightarrow{p} 0$.

Das legt nahe, dass $\sphericalangle \epsilon, \mu \to 0$ strebt.
Nachrechnen: Betrachte den Cosinus des Winkels:
$\frac{\mu'\epsilon}{\|\mu\|\cdot\|\epsilon\|} = \frac{\frac{1}{p}\mu'\epsilon}{\|\frac{1}{\sqrt{p}}\mu\|\cdot\|\frac{1}{\sqrt{p}}\epsilon\|} = \frac{\frac{1}{p}\mu'\epsilon}{\left(\frac{\mu'\mu}{p}\right)^{\frac{1}{2}}\left(\frac{\epsilon'\epsilon}{p}\right)^{\frac{1}{2}}} \to \frac{0}{(\rho^2)^{\frac{1}{2}}(1)^{\frac{1}{2}}} = 0.$

$$
\begin{aligned}
a^2 - (\alpha c)^2 &= b^2 - ((1-\alpha)c)^2 \\
a^2 - \alpha^2 c^2 &= b^2 - (c^2 - 2\alpha c^2 + \alpha^2 c^2) \\
a^2 - \alpha^2 c^2 &= b^2 - c^2 + 2\alpha c^2 - \alpha^2 c^2 \\
\frac{a^2 - b^2 + c^2}{2c^2} &= \alpha. \\
\Rightarrow \alpha = \frac{a^2 - b^2 + c^2}{2c^2} &= \frac{2a^2}{2c^2} = \frac{c^2 - b^2}{c^2} = 1 - \frac{b^2}{c^2}.
\end{aligned}
$$

Im Dreieck $0, \mu, X$ ist

$$
\begin{aligned}
\alpha &= 1 - \frac{\epsilon'\epsilon}{X'X} \\
&= 1 - \frac{p}{X'X}\underbrace{\frac{1}{p}\epsilon'\epsilon}_{\xrightarrow{p} 1} \\
&\approx 1 - \frac{p}{X'X}.
\end{aligned}
$$

Das legt folgenden Schätzer für μ nahe:

$$
\tilde{\mu} = \left(1 - \frac{p}{X'X}\right)X
$$

- Das ist (fast) der James-Stein-Schätzer.
Fehler von X: $\|\mu - X\|^2 = \epsilon'\epsilon$.
Fehler von $\tilde{\mu}$: $\|\tilde{\mu} - X\|^2$.
Relativer Fehler:

$$
\frac{\|\tilde{\mu} - X\|^2}{\|X - \mu\|^2} = \frac{\frac{1}{p}\|\tilde{\mu} - X\|^2}{\frac{1}{p}\|X - \mu\|^2} \xrightarrow{p} 1 - \frac{1}{1+\rho^2} < 1
$$

... hängt nur von $\rho^2 \approx \frac{\|\mu\|^2}{p}$ ab.
Es gilt sogar:
$\frac{MSE(\tilde{\mu},\mu)}{MSE(X,\mu)}$ bzw. $\frac{MSE(\hat{\mu}_{JS},\mu)}{MSE(X,\mu)}$ hängen von μ nur über $\|\mu\|^2$ bzw. $\frac{\|\mu\|^2}{p}$ ab.

1.5 Das Verhalten von ML-Schätzer in großen Stichproben

Siehe zu diesem Kapitel S. 274-279 in *Rice (2007)*.

Kurz gesagt: Unter geeigneten Voraussetzungen sind Maximum-Likelihood-Schätzer konsistent, asymptotisch normalverteilt, und "asymptotisch effizient".
Betrachte durchwegs X_i, $i \geq 1$, i.i.d. mit Dichte (bzw. Wahrscheinlichkeitsfunktion)

$$
f(x|\theta_0),\ \theta_0 \in \Theta \subseteq \mathbb{R}, 1-dim
$$

und für die Stichprobe der Größe n den Maximimum-Likelihood-Schätzer

$$\hat{\theta}_n = \hat{\theta}_n(X_1, \dots, X_n).$$

Satz 1.5 **(Konsistenz)** Unter geeigneten Voraussetzungen an $f(\cdot|\cdot)$ ist der Maximum-Likelihood-Schätzer $\hat{\theta}_n$ konsistent:

$$\hat{\theta}_n \xrightarrow{p} \theta_0,$$

für jeden Wert des wahren Parameters $\theta_0 \in \Theta$.

Beweisidee. $\hat{\theta}_n$ maximiert $l_n(\theta) = \prod_{i=1}^n f(X_i|\theta)$ bzw. $l_n(\theta) = \frac{1}{n}\sum_{i=1}^n \underbrace{\log f(X_i|\theta)}_{iid\ mit\ \mathbb{E}_{\theta_0}(\log f(X_i|\theta))}$ bzw. $\frac{1}{n}l_n(\theta)$ für jedes $\theta \in \Theta$.
Mit dem Gesetz der großen Zahlen gilt

$$l_n(\theta) \xrightarrow{p} \mathbb{E}_{\theta_0}(\log f(X_1|\theta)) := l_\infty(\theta).$$

für jedes $\theta \in \Theta$.
Idee:

$$\underbrace{\text{Maximierer von } \frac{1}{n}l_n(\theta)}_{\hat{\theta}_n} \approx \underbrace{\text{Maximierer von } l_\infty(\theta)}_{?=\theta_0\,\text{(siehe unten)}}.$$

? Maximierer von $l_\infty(\theta) = ?$

$$\begin{aligned}
\frac{d}{d\theta}l_\infty(\theta) &= \frac{d}{d\theta}\int \log f(x|\theta) \underbrace{f(x|\theta_0)}_{\text{dichte an der Stelle}\theta_0} dx \\
&= \int \frac{d}{d\theta}\log f(x|\theta)\cdot f(x|\theta_0)dx \\
&= \int \frac{\frac{d}{d\theta}f(x|\theta)}{f(x|\theta)}\cdot f(x|\theta_0)dx = (\ast).
\end{aligned}$$

Für $\theta = \theta_0$ ist also

$$\frac{d}{d\theta}l_\infty(\theta_0) = \frac{d}{d\theta}l_\infty(\theta)\Big|_{\theta=\theta_0}.$$

$$\begin{aligned}
(\ast) &= \int \frac{\frac{d}{d\theta}f(x|\theta_0)}{\cancel{f(x|\theta_0)}}\cancel{f(x|\theta_0)}dx \\
&= \int \frac{d}{d\theta}f(x|\theta_0)dx \\
&= \frac{d}{d\theta}\underbrace{\int f(x|\theta_0)dx}_{=1}\Big|_{\theta=\theta_0} \\
&= 0.
\end{aligned}$$

? Ist θ_0 ein Maximierer für $l_\infty(\theta)$?

$$\begin{aligned}
&\frac{d^2}{d\theta^2}\frac{d}{d\theta}\int \log f(x|\theta) f(x|\theta_0)dx = \\
&= \int \frac{d^2}{d\theta^2}\log f(x|\theta)\cdot f(x|\theta_0)dx\Big|_{\theta=\theta_0} \\
&= \int \frac{d^2}{d\theta^2}\log f(x|\theta_0)\cdot f(x|\theta_0)dx \\
&= \mathbb{E}_{\theta_0}\left(\frac{d^2}{d\theta^2}\log f(X_1|\theta_0)\right) \\
&= -I(\theta_0) \leq 0.
\end{aligned}$$

Im nicht-trivialen Fall (stetige Zufallsvariablen), wo $I(\theta_0) > 0$ ist also θ_0 ein Maximierer von $l_\infty(\theta)$.
Weitere Details zum Beweis folgen im Masterstudium.
Für diskrete Zufallsvariablen funktioniert analog. □

Satz 1.6 (Asymptotische Normalität) Unter geeigneten Voraussetzungen an $f(\cdot|\cdot)$, gilt:

$$\sqrt{n}\left(\hat{\theta}_n - \theta_0\right) \xrightarrow{w} N\left(0, \frac{1}{I(\theta_0)}\right);$$

Beweisidee. (Für Stetige Zufallsvariablen, diskrete analog) $\hat{\theta}_n$ maximiert $l_n(\theta) = \sum_{i=1}^n \log f(X_i|\theta)$.

$$\begin{aligned}
\Rightarrow 0 &= \frac{d}{d\theta} l_n(\theta)\Big|_{\theta=\hat{\theta}_n} \\
&= \frac{d}{d\theta} l_n(\hat{\theta}_n) = l_n'(\hat{\theta}_n) \\
&\approx l_n'(\theta_0) + (\hat{\theta}_n - \theta_0) l_n''(\theta_0) + \cancel{Rest}
\end{aligned}$$

Taylor-Entwicklung der
Ordnung 1 im Punkt θ_0

$$\Rightarrow \hat{\theta}_n - \theta_0 \approx \frac{l_n'(\theta_0)}{-l_n''(\theta_0)} = \frac{\frac{1}{n} l_n'(\theta_0)}{-\frac{1}{n} l_n''(\theta_0)}$$

$$\Rightarrow \sqrt{n}(\hat{\theta}_n - \theta_0) \approx \frac{\sqrt{n}\, l_n'(\theta_0)\frac{1}{n}}{-l_n''(\theta_0)\frac{1}{n}} = \frac{\frac{1}{\sqrt{n}} l_n'(\theta_0)}{-\frac{1}{n} l_n''(\theta_0)} \qquad (\ast)$$

$$\frac{1}{\sqrt{n}} l_n'(\theta_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \underbrace{\frac{d}{d\theta}\log f(X_i|\theta)\Big|_{\theta=\theta_0}}_{\substack{\frac{d}{d\theta}\log f(X_i|\theta_0) \\ \text{i.i.d. mit E-Wert=0 und Varianz} \\ \mathbb{E}_{\theta_0}\left(\left(\frac{d}{d\theta}\log f(X_1|\theta_0)\right)^2\right) = I(\theta_0)}}$$

$\xrightarrow{w} N(0, I(\theta_0))$ wegen dem zentralen Grenzwertsatz.

$$
\begin{aligned}
-\frac{1}{n} l_n''(\theta_0) &= \frac{1}{n} l_n''(\theta_0) \\
&= \frac{1}{n} \sum_{i=1}^{n} \underbrace{\left(-\frac{d^2}{d\theta^2} \log f(X_i|\theta_0)\right)}_{\substack{\text{i.i.d. mit E-Wert} \\ -\mathbb{E}_{\theta_0}\left(\frac{d^2}{d\theta^2}\log f(X_1|\theta_0)\right)=I(\theta_0)}} \\
&\xrightarrow{p} I(\theta_0) \text{ wegen dem Gesetz der großen Zahlen.}
\end{aligned}
$$

Falls $I(\theta_0)$ folgt auch dass

$$
\frac{1}{-\frac{1}{n} l_n''(\theta_0)} \xrightarrow{p} \frac{1}{I(\theta_0)}.
$$

Also:

$$
\frac{\frac{1}{\sqrt{n}} l_n'(\theta_0)}{-\frac{1}{n} l_n''(\theta_0)} = \underbrace{\frac{1}{\sqrt{n}} l_n'(\theta_0)}_{\xrightarrow{w} N(0, I(\theta_0))} \cdot \underbrace{\frac{1}{-\frac{1}{n} l_n''(\theta_0)}}_{\xrightarrow{p} \frac{1}{I(\theta_0)}}
$$

Mit den Rechenregeln für $\xrightarrow{p}$ und $\xrightarrow{w}$ ergibt sich, dass

$$
\frac{\frac{1}{\sqrt{n}} l_n'(\theta_0)}{-\frac{1}{n} l_n''(\theta_0)} \xrightarrow{w} N\left(0, \frac{1}{I(\theta_0)}\right).
$$

Wenn der Approximationsfehler in $(\ast)$ für $n \to \infty$ vernachlässigbar ist, dann folgt daraus auch, dass

$$
\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{w} N\left(0, \frac{1}{I(\theta_0)}\right).
$$

Weitere Details folgen im Masterstudium.

□

Bemerkung: Für Inferenz über θ_0 benötigt man noch einen konsistenten Schätzer für $\frac{1}{I(\theta_0)}$ bzw. $I(\theta_0)$, der für $I(\theta_0)$ (unter geeigneten Voraussetzungen) gegeben ist durch

$$
\hat{I} = -\frac{1}{n} \sum_{i=1}^{n} \frac{d^2}{d\theta^2} \log f(X_i|\hat{\theta}_n)
$$

oder durch

$$
\tilde{I} = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{d}{d\theta} \log f(X_i|\hat{\theta}_n)\right)^2.
$$

Bemerkung: Falls $\hat{\theta}_n$ ein unverzerrter Maximum-Likelihood-Schätzer für θ_0 ist, dann gilt die Cramér-Rao-Schranke:

$$\begin{aligned}
\mathrm{Var}_{\theta_0}(\hat{\theta}_n) &\geq \frac{1}{nI(\theta_0)} \\
\Rightarrow \mathrm{Var}_{\theta}(\sqrt{n}(\hat{\theta}_n - \theta_0)) &= \mathrm{Var}_{\theta_0}(\sqrt{n}\hat{\theta}_n) \\
&= n\,\mathrm{Var}_{\theta_0}(\hat{\theta}_n) \\
&\geq n\frac{1}{nI(\theta_0)} = \frac{1}{I(\theta_0)}.
\end{aligned}$$

Für allgemeine Maximum-Likelihood-Schätzer besagt der letzte Satz, dass

$$\mathrm{Var}(\text{Grenzverteilung von } \sqrt{n}(\hat{\theta}_n - \theta_0)) = \frac{1}{I(\theta_0)}.$$

In diesem Sinne ist der Maximum-Likelihood-Schätzer asymptotisch effizient.

2. Testen von Hypothesen

2.1 Neyman-Pearson Paradigma

Siehe zu diesem Kapitel S. 329-341 in *Rice (2007)*.

... ein allgemeines Schema zum Testen von Hypothesen.

Gegeben:

Zufallsvariable (Daten) $X_1, \ldots, X_n$;
Nullhypothese H_0 (über die Verteilung der X_i);
Alternativ-Hypothese H_1 (über die Verteilung der X_i);
Signifikanzniveau α, $0 < \alpha < 1$.

Wahl:

Test-Statistik $T = T(X_1, \ldots, X_n)$, sodass die **Verteilung von** T **unter** H_0 **bekannt** ist;
Verwerfungsbereich R, sodass $\mathbb{P}(T \in R | H_0) = \alpha$ (bzw. $\leq \alpha$).

Test:

$\cancel{H_0}$ falls $T \in R$;
$\rightsquigarrow H_0$ falls $T \notin R$.

Dieser Test kann zwei Arten von Fehlern begehen:

- $\cancel{H_0}$, aber H_0 trifft zu - Fehler 1. Art. (gut kontrollierbar)
- $\rightsquigarrow H_0$ aber H_1 trifft zu - Fehler 2. Art. (schwierig zu kontrollieren)

Laut Konstruktion ist die Wahrscheinlichkeit 1. Art gleich

$$\mathbb{P}(\cancel{H_0} | H_0) = \mathbb{P}(T \in R | H_0) = \alpha,$$

also das Signifikanzniveau des Tests.
Die Wahrscheinlichkeit eines Fehlers 2. Art wird von der Wahl von T und R beeinflusst.

Bezeichnungen:

- Für einen Test mit Signifikanzniveau α nennt man α auch die **"Size"** des Tests:

$$\text{Size} = \mathbb{P}(\cancel{H_0} | H_0) = \alpha \text{ (je kleiner, desto besser).}$$

- Als **"Power"** des Tests bezeichnet man die Wahrscheinlichkeit, H_0 korrekterweise zu verwerfen:

$$\text{Power} = \mathbb{P}(\cancel{H_0} | H_1) \text{ (je größer, desto besser).}$$

Bemerkung:

$$\text{Power} = 1 - \mathbb{P}(\rightsquigarrow H_0 | H_1) = 1 - \text{Wahrscheinlichkei eines Fehlers 2. Art.}$$

Bemerkung: Vergrößert man den Ablehnungsbereich R eines Tests, dann …
… steigt die Size ☹,
… steigt die Power ☺.
Power und Size verhalten sich ”antagonistisch”.

Beispiel 2.1 (Futschik, 2002, Ist der Euro fair?) **Kreiselexperiment:**

Land	n	$T = \#\text{Kopf}$	c	$T - \frac{n}{2}$	
AUT	100	50	11	0	$\rightsquigarrow H_0$
GE	100	52	11	2	$\rightsquigarrow H_0$
IT	250	103	16	-22	$\cancel{H_0}$
FR	250	158	16	33	$\cancel{H_0}$

Für jedes Land ist $X_1, \dots, X_n$ i.i.d. $B(p)$.

- $H_0 \colon p = \frac{1}{2}$,
- $H_1 \colon P \neq \frac{1}{2}$,
- $\alpha = 0.05$.

Wähle $T = \sum_{i=1}^{n} X_i \sim B(n, p)$. (Unter H_0 ist $T \sim B\left(n, \frac{1}{2}\right)$).
Große Werte von $|T - \frac{n}{2}|$ sprechen gegen H_0!
Verwerfe H_0, wenn $|T - \frac{n}{2}| \geq c$ ist. Also

$$R = \left(-\infty, \frac{n}{2} - c\right] \cup \left[\frac{n}{2} + c, \infty\right),$$

und verwerfe H_0 falls $T \in R$.
Für $n = 100$ ist $c = 11$,
für $n = 250$ ist $c = 16$.
Bemerkung: Weil T diskret ist, kann das Signifikanzniveau α nicht exakt erreicht werden (im Allgemeinem).
Hier ist c so gewählt, dass die Wahrscheinlichkeit eines Fehlers 1. Art $\leq \alpha$ ist, und so, dass c möglichst klein (also R möglichst groß) ist.
Katapultexperiment:

Land	n	$T = \#\text{Kopf}$	c	$T - \frac{n}{2}$	
AUT	400	196	21	-4	$\rightsquigarrow H_0$
IT	150	71	13	-4	$\rightsquigarrow H_0$
FR	150	79	13	4	$\rightsquigarrow H_0$

Man sieht: Die Nullhypothese kann nur verworfen werden aber nicht bestätigt werden.
Was ist die Power dieses Tests:

$$\text{Power} = \mathbb{P}\Big(|T - \frac{n}{2}| \geq c|H_1\Big)\ldots \text{hängt von } p\Big(\neq \frac{1}{2}\Big) \text{ ab.}$$

Setze

$$\Pi(p) = \mathbb{P}\Big(|T - \frac{n}{2}| \geq c|p\Big) = \mathbb{P}\Big(|B(n,p) - \frac{n}{2}| \geq c\Big),$$

falls X_i i.i.d. $B(p)$.
Beachte:

- Für $p \neq \frac{1}{2}$ ist $\Pi(p)$ die Power des Tests.
- Für $p = \frac{1}{2}$ ist $\Pi(p)$ die Size des Tests ($\leq \alpha$).

Bemerkung: Viele Tests sind von der Form $\cancel{H_0}$ falls $|S| \geq c$, oder $S \geq c$, oder $S \leq c$.
Für solche Tests nennt man c den **kritischen Wert** des Tests (zum Signifikanzniveau α des Tests).
Beachte: $\mathbb{P}(\cancel{H_0}|H_0) = \alpha$.

Bezeichnung: Für einen Test der oberen Form ist der **p-Wert** des Signifikanzniveaus jenes Test, bei dem der kritische Wert c ersetzt wird durch den beobachteten Wert der Test-Statistik.
Beachte: $0 \leq$ p-Wert ≤ 1.

Test	beobachteter Wert	p-Wert	
$\cancel{H_0}$ falls $\lvert S\rvert \geq c$	$\lvert s\rvert$	$\mathbb{P}(\lvert S\rvert \geq \lvert s\rvert\,\lvert H_0)$	… fällt in $\lvert s\rvert$
$\cancel{H_0}$ falls $S \geq c$	s	$\mathbb{P}(S \geq s\,\lvert H_0)$	… fällt in s
$\cancel{H_0}$ falls $S \leq c$	s	$\mathbb{P}(S \leq s\,\lvert H_0)$	… steigt in s

Bemerkung: Der p-Wert misst, wie stark die Daten der Nullhypothese wiedersprechen.

Beispiel 2.2 (Futschik, 2002, Ist der Euro fair?) $S = T - \frac{n}{2}$, $\cancel{H_0}$ falls $|S| \geq c$. $\alpha = 0.05$.
Kreiselexperiment:

Land	n	$T = \#$Kopf	c	$\lvert s\rvert$		p-Wert
AUT	100	50	11	0	$\rightsquigarrow H_0$	1
GE	100	52	11	2	$\rightsquigarrow H_0$	0.69
IT	250	103	16	22	$\cancel{H_0}$	0.0054
FR	250	158	16	33	$\cancel{H_0}$	0.000028

Katapultexperiment:

Land	n	$T = \#\text{Kopf}$	c	$\lvert s\rvert$		p-Wert
AUT	400	196	21	4	$\rightsquigarrow H_0$	0.69
IT	150	71	13	4	$\rightsquigarrow H_0$	0.52
FR	150	79	13	4	$\rightsquigarrow H_0$	0.52

Berechnung des p-Wertes in diesem Beispiel:
$T \sim Bin\left(n, \frac{1}{2}\right)$ unter H_0.

$$\begin{aligned}
\mathbb{P}(|S| \geq |s| \,|H_0) &= \mathbb{P}\left(|T - \frac{n}{2}| \geq |s| \,|H_0\right) \\
&= \mathbb{P}\left(|Bin\left(n, \frac{1}{2}\right) - \frac{n}{2}| \geq |s|\right) \\
&= \mathbb{P}\left(Bin\left(n, \frac{1}{2}\right) \leq \frac{n}{2} - |s|\right) + \mathbb{P}\left(Bin\left(n, \frac{1}{2}\right) \geq \frac{n}{2} + |s|\right).
\end{aligned}$$

Bemerkung: Liefert ein Test einen p-Wert von $a \in (0,1)$, dann wird in dem Test die Null-Hypothese verworfen auf jedem Signifikanzniveau $\alpha \geq a$.

Nachrechnen. Betrachte Test, wo $\cancel{H_0}$ falls $S \geq c$ mit Signifikanzniveau α.
Nach Durchführung des Tests erhält man einen p-Wert von a und einen Wert s der Test-Statistik.
Es gilt:

$$\begin{aligned}
\mathbb{P}(S \geq \tilde{c} \,|H_0) = \tilde{\alpha} \geq a = \mathbb{P}(S \geq s \,|H_0) \\
\Leftrightarrow \tilde{c} \leq s \\
\equiv s \geq \tilde{c}
\end{aligned}$$

Beobachtbarer Wert der Test-Statistik Kritische Wert des Tests

$$\Leftrightarrow \cancel{H_0}.$$

□

Bemerkung: Falls die Verteilungsfunktion F der Test-Statistik S (bzw. $|S|$) invertierbar ist, dann ist der p-Wert vor der Durchführung des Experiments/Tests eine Zufallsvariable, die $U([0,1])$-verteilt ist.

Nachrechnen. Betrachte Test, wo $\cancel{H_0}$ falls $S \leq c$:
Nach der Durchführung des Experiments ist der p-Wert gegeben durch $\mathbb{P}(S \leq s \,|H_0) = F(s)$.
Vor der Durchführung des Experiments ist der p-Wert gegeben durch $F(S)$.
Betrachte die Verteilungsfunktion der Zufallsvariable $Z = F(S)$:

$$\begin{aligned}
\mathbb{P}(Z \leq x) &= \mathbb{P}(F(S) \leq x) \\
&= \mathbb{P}(S \leq F^{-1}(x)) \\
&= F(F^{-1}(x)) = x
\end{aligned}$$

für $0 \leq x \leq 1$.

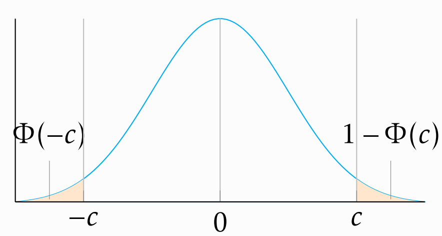
Für $x < 0$ ist $\mathbb{P}(Z \leq x) = 0$ da $Z \in [0,1]$,
für $x > 1$ ist $\mathbb{P}(Z \leq x) = 1$ da $Z \in [0,1]$.
Das heißt die Verteilungsfunktion von $Z = F(S)$, also die Verteilungsfunktion des p-Werts vor der Durchführung, ist gerade die Verteilungsfunktion der $U([0,1])$; $Z \sim U([0,1])$. $\square$

Beispiel 2.3 (Der z-Test) (Testen des Mittelwerts im Gauß'schem Modell mit bekannter Varianz).
Betrachte $X_1,\dots,X_n$ i.i.d. $N(\mu_0,\sigma^2)$, $\mu_0 \in \mathbb{R}$, $\sigma^2 > 0$ bekannt.
$H_0\colon \mu = \mu_z$ (Für einen bestimmten Wert μ_0).
Die sogenannte z-Statistik ist gegeben durch

$$\begin{aligned} Z &= \underbrace{\frac{1}{\sqrt{n}}\underbrace{\sum_{i=1}^{n} \underbrace{\frac{X_i - \mu_0}{\sigma}}_{\sim \text{ i.i.d. } N(0,1) \text{ unter } H_0}}_{\sim N(0,n) \text{ unter } H_0}}_{\sim N(0,1) \text{ unter } H_0} \\ &= \sqrt{n}\frac{\bar{X}_n - \mu_0}{\sigma}. \end{aligned}$$

Zweiseitige Alternative:
$H_1\colon \mu_0 \neq \mu_z$.
$\cancel{H_0}$ falls $|Z|$ groß ist. Für Signifikanzniveau $\alpha \in (0.1)$ wählt man $c = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)$ und verwirft, wenn $|Z| \geq c$:

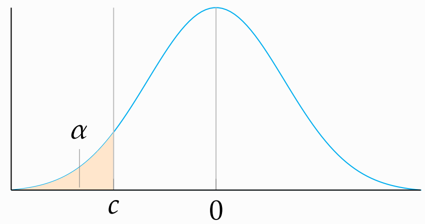
$$\begin{aligned} \mathbb{P}(\cancel{H_0} | H_0) &= \mathbb{P}(|Z| \geq c \,|H_0) \\ &= \mathbb{P}(|N(0,1)| \geq c) \\ &= \mathbb{P}(N(0,1) \leq -c) + 1 - \mathbb{P}(N(0,1) \leq c) \\ &= \Phi(-c) + 1 - \Phi(c) \\ &= 2\,(1 - \Phi(c)) \qquad\qquad = 2\left(1 - \Phi\left(\Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\right)\right) \\ &= 2\left(1 - 1 + \frac{\alpha}{2}\right) = \alpha. \checkmark \end{aligned}$$



Linksseitige Alternative:
$H_1\colon \mu_0 < \mu_z$.
$\cancel{H_0}$ falls Z links von 0 liegt. Für Signifikanzniveau $\alpha \in (0,1)$ wählt man $c = \Phi^{-1}(\alpha)$ und verwirft H_0, wenn $Z \leq c$.

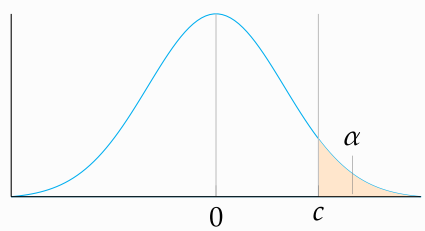
Signifikanzniveau dieses Tests:

$$\begin{aligned}\mathbb{P}(\cancel{H_0}|H_0) &= \mathbb{P}(Z \leq c|H_0)\\ &= \mathbb{P}(N(0,1) \leq c)\\ &= \Phi(c) = \Phi(\Phi^{-1}(\alpha))\\ &= \alpha\end{aligned}$$



Rechtsseitige Alternative:
$H_1: \mu_0 > \mu_z$.
$\cancel{H_0}$ falls Z rechts von 0 liegt. Für Signifikanzniveau $\alpha \in (0,1)$ wählt man $c = \Phi^{-1}(1-\alpha)$ und verwirft H_0, wenn $Z \geq c$ ist.
Signifikanzniveau dieses Tests:

$$\begin{aligned}\mathbb{P}(\cancel{H_0}|H_0) &= \mathbb{P}(z \geq c|H_0)\\ &= \mathbb{P}(N(0,1) \geq c)\\ &= 1 - \mathbb{P}(N(0,1) < c)\\ &= 1 - \Phi(c) = 1 - \Phi(\Phi^{-1}(1-\alpha))\\ &= 1 - (1-\alpha) = \alpha\end{aligned}$$



Beispiel 2.4 (Der t-Test) (Testen des Mittelwerts im Gauß'schen Modell mit unbekannter Varianz).
Betrachte $X_1, \dots, X_n$ i.i.d. $N(\mu_0, \sigma_0^2)$, $\mu_0 \in \mathbb{R}$, $\sigma_0^2 > 0$.
$H_0: \mu_0 = \mu_z$.
Setze

$$T = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{X_i - \mu_z}{\hat{\sigma}_n} = \sqrt{n}\frac{\bar{X} - \mu_z}{\hat{\sigma}_n}$$

mit $\sqrt{n}(\bar{x}_n - \mu_z) \sim N(0, \sigma^2)$; $\hat{\sigma}_n \sim \sqrt{\frac{\sigma^2 \chi_{n-1}}{n-1}}$

Beachte: Unter H_0 ist $T \sim t_{n-1}$ denn

$$
\begin{aligned}
T &\sim \frac{N(0,\sigma^2)}{\sigma\sqrt{\chi^2_{n-1}/(n-1)}} \\
&\sim \frac{N(0,1)}{\sqrt{\chi^2_{n-1}/(n-1)}} \\
&\sim t_{n-1}.
\end{aligned}
$$

Zweiseitige Alternative:
$H_1: \mu_0 \neq \mu_z$.
Für ein Signifikanzniveau $\alpha \in (0,1)$ wählt man $c = F^{-1}_{n-1}\left(1-\frac{\alpha}{2}\right)$ und verwirft, wenn $|T| \geq c$.
Linksseitige Alternative:
$H_1: \mu_0 < \mu_z$.
Für Signifikanzniveau $\alpha \in (0,1)$ wählt man $c = F^{-1}_{n-1}(\alpha)$ und verwirft, wenn $T \leq c$.
Rechtsseitige Alternative:
$H_1: \mu_0 > \mu_z$.
Für ein Signifikanzniveau $\alpha \in (0,1)$ wählt man $c = F^{-1}_{n-1}(1-\alpha)$ und verwirft, wenn $T \geq c$.

Beispiel 2.5 (2-Stichproben t-Test) Betrachte 2 Stichproben:

- $X_1,\ldots,X_n$ i.i.d. $N\left(\mu_x,\sigma_0^2\right)$
- $Y_1,\ldots,Y_m$ i.i.d. $N\left(\mu_y,\sigma_0^2\right)$.

Beide Stichproben sind voneinander unabhängig und haben die gleiche Varianz.
H_0: $\mu_x = \mu_y$ und $\bar{X}_n = \bar{Y}_n$.
Unter H_0 gilt:

$$
\begin{aligned}
\bar{X}_n &=\sim N\left(\mu_x, \frac{\sigma_0^2}{n}\right), \\
\bar{Y}_m &=\sim N\left(\mu_x, \frac{\sigma_0^2}{m}\right), \\
\bar{X}_n - \bar{Y}_m &\sim N\left(0, \sigma_0^2\left(\frac{1}{n}+\frac{1}{m}\right)\right) \\
\frac{\bar{X}_n - \bar{Y}_n}{\sigma_0\sqrt{\frac{1}{n}-\frac{1}{m}}} &\sim N(0,1).
\end{aligned}
$$

$\bar{X}_n$ und $\bar{Y}_m$ sind unabhängig voneinander.
Setze

$$
\hat{\mu} = \frac{1}{n+m}\left(\sum_{i=1}^{n} X_i + \sum_{i=1}^{m} Y_i\right)
$$

$$\begin{aligned}\hat{\sigma}^2 &= \frac{1}{m+n-1}\left(\sum_{i=1}^{n}(X_i-\hat{\mu})^2+\sum_{i=1}^{m}(Y_i-\hat{\mu})^2\right)\\ &\sim \frac{\sigma_0^2\chi^2_{n+m-1}}{n+m-1} \text{ unter } H_0.\end{aligned}$$

Betrachte die Test-Statistik

$$T=\frac{\bar{X}_n-\bar{Y}_m}{\hat{\sigma}\sqrt{\frac{1}{m}+\frac{1}{n}}}.$$

Beachte:

$$T=\underbrace{\frac{\bar{X}_n-\bar{Y}_m}{\sigma_0\sqrt{\frac{1}{m}+\frac{1}{n}}}}_{\sim N(0,1)}\cdot\overbrace{\frac{\sigma_0}{\hat{\sigma}}}^{\frac{1}{\sqrt{\frac{\chi^2_{m+n-1}}{m+n-1}}}}\sim t_{m+n-1}.$$

$\hat{\sigma}^2\sim\frac{\sigma_0^2}{m+n-1}\chi^2_{m+n-1}$,
$\frac{\hat{\sigma}^2}{\sigma_0^2}\sim\frac{1}{m+n-1}\chi^2_{m+n-1}$.

Linksseitige Alternative: $H_1: \mu_x<\mu_y \quad (\sigma_x^2,\sigma_y^2$ egal).
$\cancel{H_0}$ falls $T\le c$ für $c=F^{-1}_{m+n-1}(\alpha)$.

Rechtsseitge Alternative: $H_1: \mu_x>\mu_y \quad (\sigma_x^2,\sigma_y^2$ egal).
$\cancel{H_0}$ falls $T\ge c$ für $c=F^{-1}_{m+n-1}(1-\alpha)$.

Zweiseitige Alternative: $H_1: \mu_x\neq\mu_y \quad (\sigma_x^2,\sigma_y^2$ egal).
$\cancel{H_0}$ falls $|T|\ge c$ für $c=F^{-1}_{m+n-1}\left(1-\frac{\alpha}{2}\right)$.

Beispiel 2.6 (Der F-Test) (Testen auf Gleichheit der Varianzen zweier normalverteilter Populationen).
Betrachte zwei Stichproben:

- $X_1,\dots,X_n$ i.i.d. $N\left(\mu_x,\sigma_x^2\right)$
- $Y_1,\dots,Y_m$ i.i.d. $N\left(\mu_y,\sigma_y^2\right)$.

Beide Stichproben sind unabhängig voneinander.
$H_0: \sigma_x^2=\sigma_y^2 \quad (\mu_x,\mu_y$ egal).
Unter H_0 ist

$$\hat{\sigma}_x^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X}_n)^2\sim\frac{\sigma_x^2}{n-1}\chi^2_{n-1},$$

$$\hat{\sigma}_y^2=\frac{1}{m-1}\sum_{i=1}^{m}(Y_i-\bar{Y}_m)^2\sim\frac{\sigma_x^2}{m-1}\chi^2_{m-1},$$

wobei $\hat{\sigma}_x^2$ und $\hat{\sigma}_y^2$ unabhängig voneinander sind.
Damit gilt unter H_0:

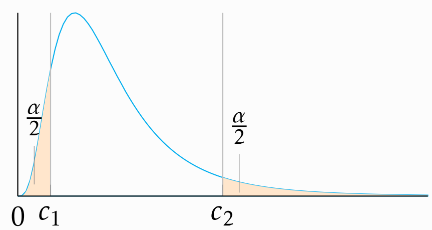
$$\begin{aligned} F = \frac{\hat{\sigma}_x^2}{\hat{\sigma}_y^2} &\sim \frac{\frac{\sigma_x^2}{n-1}\chi_{n-1}^2}{\frac{\sigma_x^2}{m-1}\chi_{m-1}^2} \\ &\sim \frac{\frac{\chi_{n-1}^2}{n-1}}{\frac{\chi_{m-1}^2}{m-1}} \\ &\sim F_{n-1,m-1}, \end{aligned}$$

wobei $F_{n-1,m-1}$ die F-Verteilung mit $n-1$ und $m-1$ Freiheitsgraden ist.

Linksseitige Alternative: $\sigma_x^2 < \sigma_y^2$.
$\cancel{H_0}$ falls $F \le c$ für $c = F_{n-1,m-1}^{-1}(\alpha)$.

Rechtsseitige Alternative: $\sigma_x^2 > \sigma_y^2$.
$\cancel{H_0}$ falls $F \ge c$ für $c = F_{n-1,m-1}^{-1}(1-\alpha)$.

Zweiseitige Alternative: $\sigma_x^2 \ne \sigma_y^2$.
$\cancel{H_0}$ falls $F \le c_1$ oder $F \ge c_2$ für $c_1 = F_{n-1,m-1}^{-1}\left(\frac{\alpha}{2}\right)$ und $c_2 = F_{n-1,m-1}^{-1}\left(1-\frac{\alpha}{2}\right)$.



2.2 Likelihood-Ratio-Tests und das Neyman-Pearson Lemma

Betrachte ein parametrisches Modell $X_1, \ldots, X_n$ i.i.d. mit Dichte (Wahrscheinlichkeitsfunktion) $f(x|\theta), \quad \theta \in \Theta$.
Testproblem:
$H_0: \theta \in \Theta_0$ für $\Theta_0 \subseteq \Theta$,
$H_1: \theta \in \Theta \setminus \Theta_0$.

Bisher hatten wir z.B.:

- $\Theta = \mathbb{R}, \quad \Theta_0 = \{\mu_\star\}; \quad H_0 = \mu = \mu_\star \ vs \ H_1 : \mu \ne \mu_\star$ (z-Test 2-Seitig),
- $\Theta = (-\infty, \mu_\star) \quad \Theta_0 = \{\mu_\star\}; \quad H_0 = \mu = \mu_\star \ vs \ H_1 : \mu \ne \mu_\star$ (z-Test 1-Seitig),
- $\Theta = \mathbb{R} \times (0,\infty), \quad \Theta_0 = \{\mu_0\} \times (0,\infty); \quad H_0 = \mu = \mu_\star \ vs \ H_1 : \mu \ne \mu_\star$ (t-Test),
- $\Theta = \mathbb{R} \times (0,\infty) \times \mathbb{R} \times (0,\infty), \quad \Theta_0 = \{(\mu_x, \sigma_x^2, \mu_y, \sigma_y^2) \in \Theta : \sigma_x^2 = \sigma_y^2\};$
  $H_0 = \sigma_x^2 = \sigma_y^2 \ vs \ H_1 : \sigma_x^2 \ne \sigma_y^2$ (F-Test).

Allgemeine Modell: Hier ist die Likelihood-Ratio-Statistik (L-R-Statistik) ist definiert als

$$\Lambda = \frac{L(\hat{\theta}_0)}{L(\hat{\theta})},$$

wobei $\hat{\theta}_0$ der ML-Schätzer für θ unter H_0 ist, also $\hat{\theta}_0 = \underset{\theta\in\Theta_0}{\operatorname{argmax}}\, L(\theta)$,

und wobei $\hat{\theta}$ der (unrestringierte) ML-Schätzer für θ ist, also $\hat{\theta} = \underset{\theta\in\Theta}{\operatorname{argmax}}\, L(\theta)$.

Beachte: Wird $\Theta_0 \subseteq \Theta$, ist
$\underbrace{\underset{\theta\in\Theta_0}{\operatorname{argmax}}\, L(\theta)}_{L(\hat{\theta}_0)} \leq \underbrace{\underset{\theta\in\Theta}{\operatorname{argmax}}\, L(\theta)}_{L(\hat{\theta})}$
Sodass $0 \leq \Lambda \leq 1$ liegt.

Der L-R-Test verwirft H_0, wenn $\Lambda \leq c$ für einen geeigneten kritischen Wert $c < 1$.

Bemerkung: Λ kann interpretiert werden als

$$\frac{\text{Beste Beschreibung der Beobachtungen durch ein } \theta \in \Theta_0}{\text{Beste Beschreibung der Beobachtungen durch ein } \theta \in \Theta}.$$

Bemerkung: Einige der bisher vorgestellten Tests sind tatsächlich L-R-Tests.

Beispiel 2.7 (z-Test mit zweiseitiger Alternative) $X_1, \ldots, X_n$ i.i.d. $N(\mu, \sigma^2)$, $\mu \in \mathbb{R}$, σ^2 bekannt.
$H_0: \mu_0 = \mu_\star$, $H_1: \mu_0 \neq \mu_\star$.
$\Theta = \mathbb{R}, \Theta_0 = \{\mu_\star\}$

Likelihood für $\mu \in \Theta$ ist:

$$\begin{aligned} L(\mu) &= \prod_{i=1}^{n} \Phi_{\mu,\sigma^2}(X_i) \\ &= \prod_{i=1}^{n} (2\pi\sigma^2)^{-\frac{1}{2}} e^{-\frac{1}{2\sigma^2}(X_i-\mu)^2} \\ &= (2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2}. \end{aligned}$$

$\hat{\theta}_0 = \underset{\mu\in\{\theta_0\}}{\operatorname{argmax}}\, L(\mu) = \mu_\star$.
$\hat{\theta} = \underset{\mu\in\theta}{\operatorname{argmax}}\, L(\mu) = \bar{X}_n$.

$$
\begin{aligned}
\Lambda &= \frac{L(\mu_\star)}{L(\bar{X}_n)} \\
&= \frac{(2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n (X_i-\mu_\star)^2}}{(2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n (X_i-\bar{X}_n)^2}} \\
&= \exp\left(-\frac{1}{2\sigma^2} \underbrace{\left(\sum_{i=1}^n (X_i-\mu_\star)^2 - \sum_{i=1}^n (X_i-\bar{X}_n)^2 \right)}_{(\circledast)} \right).
\end{aligned}
$$

L-R-Test verwirft H_0, wenn Λ klein ist, also wenn $(\circledast)$ groß ist.

$$
\begin{aligned}
(\circledast) &= \sum_{i=1}^n \left(X_i^2 - 2\mu_\star X_i + \mu_\star^2 - Xi^2 + 2\bar{X}_n X_i - \bar{X}_n^2 \right) \\
&= -2\mu_\star \bar{X}_n \cdot n + \mu_\star^2 \cdot n + 2\bar{X}_n \bar{X}_n \cdot n - \bar{X}_n^2 \cdot n \\
&= n\left(\mu_\star^2 - 2\mu_\star \bar{X}_n + \bar{X}_n^2 \right) \\
&= n\left(\bar{X}_n - \mu_\star \right)^2.
\end{aligned}
$$

Damit ist $(\circledast)$ genau dann groß, wenn $|\bar{X}_n - \mu_0|$ groß ist. Das ist dieselbe Entscheidungsregel wie beim z-Test.
Also: L-R-Test ist hier der z-Test (bei zweiseitiger Alternative).

Beispiel 2.8 (z-Test mit linksseitiger Alternative) $X_1, \ldots, X_n$ wie zuvor.

$$
\begin{matrix}
H_0 : \mu = \mu_\star & & \Theta_0 = \{\mu_\star\} \\
 & \equiv & \\
H_1 : \mu < \mu_\star & & \Theta = (-\infty, \mu_\star]
\end{matrix}
$$

Hier ist

$$
\begin{aligned}
\hat{\theta}_0 &= \mu_\star \\
\hat{\theta} &= \min\{\bar{X}_n, \mu_0\}
\end{aligned}
$$

Hier verwirft der L-R-Test, wenn

$$
\begin{aligned}
& \frac{L(\mu_\star)}{L(\min\{\bar{X}_n, \mu_\star\})} \text{ klein ist } (\le c < 1) \\
\Leftrightarrow \quad & \text{Nun ist } \frac{L(\mu_\star)}{L(\min\{\bar{X}_n, \mu_\star\})} \le c < 1 \\
\Leftrightarrow \quad & \frac{L(\mu_\star)}{L(\bar{X}_n)} \le c < 1 \quad \text{und } \bar{X}_n < \mu_\star \\
\Leftrightarrow \quad & (\bar{X}_n - \mu_0)^2 \quad \text{groß und } \bar{X}_n < \mu_\star \\
\Leftrightarrow \quad & \bar{X}_n - \mu_\star \quad \text{Weit links von 0 (klein und negativ)}
\end{aligned}
$$

Der Z-Test verwirft wenn $\frac{\sqrt{n}}{\sigma}(\bar{X}_n - \mu_0)$ klein ist ($\le \phi^{-1}(\alpha)$). Also ist auch hier der L-R-Test das selbe wie die Z-Test.

Bemerkung: Für den t-Test gelten die Resultate analog zu den letzten beiden Beispielen.

Im folgenden **Neyman-Pearson Lemma** betrachtet man sogenannte **simple Hypothesen**, unter denen die Verteilung der Daten jeweils komplett spezifiziert ist:
$X_1, \ldots, X_n$ i.i.d. mit Dichte $f_\theta(x)$ (bzw. Wahrscheinlichkeitsfunktion $p_\theta(x)$), wobei $\theta \in \Theta = \{0,1\}$. Teste:

$$\begin{array}{ll} H_0 : \theta = 0, & \\ \qquad\quad \nwarrow & \\ & \text{simple Hypothesen} \\ \qquad\quad \swarrow & \\ H_1 : \theta = 1. & \end{array}$$

Satz 2.1 (Neyman-Pearson Lemma) Unter allen Tests zwischen zwei simplen Hypothesen mit Signifikanzniveau $\alpha \in (0,1)$ hat der entsprechende L-R-Test die größtmögliche Power. (Fehler 2. Art möglichst klein)

Bemerkung: Zwischen 2 einfache Hypothesen ist die entsprechende L-R-Statistik gegeben durch:

$$\Lambda = \frac{L(0)}{\max\{L(0), L(1)\}} (< 1).$$

Also wenn $\frac{L(0)}{L(1)} \leq c(< 1)$ ist .

Beweis. Zunächst für $n = 1$.
Betrachte $X = X_1$.
$L(\theta) = f_\theta(x), \quad \theta \in \{0,1\}$.
Sei $\alpha \in (0,1)$ und sei c der Kritische Wert der L-R-Test zum Signifikanzniveau α hat:

$$\alpha = \mathbb{P}(\Lambda \leq c \,|H_0) = \mathbb{P}\left(\frac{f_0(x)}{f_1(x)} \leq c \,|H_0\right).$$

Diesen Test entspricht eine $0-1$-wertige Zufallsvariable d_{LR}:

$$d_{LR} = d_{LR}(X_1) = \begin{cases} 0 & : (\frac{f_0(x)}{f_1(x)} > c, \\ 1 & : (\frac{f_0(x)}{f_1(x)} \leq c, \end{cases}$$

Beachte $c < 1$,
Verteilung von d_{LR} unter H_0:
$\mathbb{P}(d_{LR} = 1 \,|H_0) = \mathbb{P}(\Lambda \leq c \,|H_0) = \mathbb{P}\left(\frac{f_0(x)}{f_1(x)} \leq c \,|H_0\right) = \alpha$
Also: $d_{LR}\ B(p)$ unter H_0.

Betrachte nun einen weiteren Test mit Signifikanzniveau α, diesen Test entspricht eine Weitere $0-1$-wertige Zufallsvariable $d = d(X)$, wobei

$$d(X) = \begin{cases} 0 & : \text{Test verwirft nicht} \\ 1 & : \text{Test verwirft} \end{cases}$$

$$\mathbb{P}(d = 1 \,|H_0) = \mathbb{P}(\text{Test verwirft } |H_0) = \alpha$$

Also: d $B(p)$ unter H_0.
Zu zeigen:

$$\text{Power von des neuen Tests } (d) \leq \text{Power des L-R-Tests } (d_{LR})$$
$$\equiv \mathbb{P}(d = 1 \,|H_1) \leq \mathbb{P}(d_{LR} = 1 \,|H_1).$$

Hilfsmittel:

$$\mathbb{P}(A) = \mathbb{P}(A \cap B) + \mathbb{P}(A \cap B^{\complement});\ A = d = 1;\ B = d_{LR} = 1.$$

$$\begin{aligned}
\mathbb{P}(d = 1 \,|H_1) &= \mathbb{P}(d = 1, d_{LR} = 1 \,|H_1) + \mathbb{P}(d = 1, d_{LR} = 0 \,|H_1) \\
&= \mathbb{P}(d_{LR} = 1 \,|H_1) - \mathbb{P}(d = 0, d_{LR} = 1 \,|H_1) + \mathbb{P}(d = 1, d_{LR} = 0 \,|H_1) \\
&= \mathbb{P}(d_{LR} = 1 \,|H_1) - \int_{\substack{x:\ d(x)=0,\\ d_{LR}(x)=1}} f_1(x)dx + \int_{\substack{x:\ d(x)=1,\\ d_{LR}(x)=0}} f_1(x)dx \\
&\leq \mathbb{P}(d_{LR} = 1 \,|H_1) - \frac{1}{c}\int_{\substack{x:\ d(x)=0,\\ d_{LR}(x)=1}} f_0(x)dx + \frac{1}{c}\int_{\substack{x:\ d(x)=1,\\ d_{LR}(x)=0}} f_0(x)dx \\
&= \mathbb{P}(d_{LR} = 1 \,|H_1) - \frac{1}{c}(\mathbb{P}(d = 0, d_{LR} = 1 \,|H_0) - \mathbb{P}(d = 1, d_{LR} = 0 \,|H_0)) \\
&= \mathbb{P}(d_{LR} = 1 \,|H_1) - \frac{1}{c}\left(\alpha - \underbrace{\mathbb{P}(d = 1, d_{LR} = 1 \,|H_0) - \mathbb{P}(d = 1, d_{LR} = 0 \,|H_0)}_{=\alpha}\right) \\
&= \mathbb{P}(d_{LR} = 1 \,|H_1) - \frac{1}{c}(\alpha - \alpha) \\
&= \mathbb{P}(d_{LR} = 1 \,|H_1).
\end{aligned}$$

Also: $\mathbb{P}(d = 1 \,|H_1) \leq \mathbb{P}(d_{LR} = 1 \,|H_1)$ ✓.
Der Fall für diskrete Zufallsvariable (mit W-Fkt) geht analog.
Für den Fall $n \geq 1$ argumentiert man ganz genauso (siehe Übung).

□

Korollar 2.1 Seien $X_1, \ldots, X_n$ i.i.d. $N(\mu, \sigma^2)$, $\quad \mu \in (-\infty, \mu_0]$, $\quad \sigma^2 > 0$ bekannt.
Teste $H_0: \mu = \mu_0$ gegen $H_1: \mu < \mu_0$, für eine feste Zahl μ.
Hier hat die L-R-Test mit Signifikanzniveau α die größtmögliche Power.

Power einen Test wenn $\mu = \mu_1 \leq$ Power den L-R-Test wenn $\mu \neq \mu_1$
$\Pi(\mu_1) \leq \Pi_{LR}(\mu_1)$

Einen Test mit diesem Eigenschaft nennt man Uniformly Most Powerful (UMP)

Beweis. Sie $\mu_1 < \mu_0$. Zu zeigen: $\Pi(\mu_1) \leq \Pi_{LR}(\mu_1)$
Der beiden Tests entsprechen wieder $0-1$–Wertige Zufallsvariable d und d_{LR}.
Beachte: $\mathbb{P}(d_{LR} = 1 \,|H_0) = \mathbb{P}(d = 1 \,|H_0) = \alpha$.
Beide Tests haben die Signifikanzniveau α.
Betrachte dazu ein neues Testproblem:

$$\begin{aligned}&\tilde{H}_0:\ \mu=\mu_0, \\ &\qquad\qquad \nwarrow \\ &\qquad\qquad\qquad \text{simple Hypothesen} \\ &\qquad\qquad \swarrow \\ &\tilde{H}_1:\ \mu=\mu_1.\end{aligned}$$

Beachte:

- $\mathbb{P}(d_{LR}=1\,|\tilde{H}_0)=\mathbb{P}(d_{LR}=1\,|H_0)=\alpha.$
  $\mathbb{P}(d=1\,|\tilde{H}_0)=\mathbb{P}(d=1\,|H_0)=\alpha,$
  D.h. beim Testen von $\tilde{H}_0$ gegen $\tilde{H}_1$ haben d_{LR} und d beide Signifikanzniveau α.
- Der L-R-Test zum Testen von $\tilde{H}_0$ gegen $\tilde{H}_1$, verwirft H_0, wenn

$$\begin{aligned}
&\tilde{\Lambda}=\frac{L(\mu_0)}{L(\mu_1)} && \leq \tilde{c}<1\\
\Leftrightarrow\quad & \frac{(2\pi\sigma^2)^{-\frac{n}{2}}e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n(X_i-\mu_0)^2}}{(2\pi\sigma^2)^{-\frac{n}{2}}e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n(X_i-\mu_1)^2}} && \text{Soll klein sein}\\
\Leftrightarrow\quad & \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^n((X_i-\mu_0)^2)+\frac{1}{2\sigma^2}\sum_{i=1}^n((X_i-\mu_1)^2)\right) && \text{Soll klein sein}\\
\Leftrightarrow\quad & \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^n(X_i^2-2\mu_0X_i+\mu_0^2-X_i^2+2\mu_1X_i-\mu_1^2)\right) && \text{Soll klein sein}\\
\Leftrightarrow\quad & \exp\left(-\frac{n}{2\sigma^2}((\mu_0^2-\mu_1^2)+2(\mu_1-\mu_0)\bar{X}_n)\right) && \text{Klein}\\
\Leftrightarrow\quad & n((\mu_0^2-\mu_1^2)+2\bar{X}_n\underbrace{(\mu_1-\mu_0)}_{<0})\ \text{groß}\\
\Leftrightarrow\quad & \bar{X}_n\ \text{klein}\\
\Leftrightarrow\quad & \text{Genauer: } \bar{X}_n \text{ sodass } \mathbb{P}(\bar{X}_n\leq c|H_0)=\alpha
\end{aligned}$$

- Der L-R-Test von H_0 gegen H_1 ist der Z-Test. Dieser Test verwirft H_0(bzw. $\tilde{H}_0$) wenn

$$\begin{aligned}
&\frac{\sqrt{n}}{\sigma}(\bar{X}_n-\mu_0)\ \text{klein}\\
\Leftrightarrow\quad & \bar{X}_n-\mu_0\ \text{klein}\\
\Leftrightarrow\quad & \bar{X}_n\ \text{klein.}
\end{aligned}$$

Also: Der L-R-Test zwischen H_0 und H_1, und der L-R-Test zwischen $\tilde{H}_0$ und $\tilde{H}_1$ stimmen überein.

Mit N-P-Lemma: Unter allen Tests von $\tilde{H}_0$ gegen $\tilde{H}_1$ mit Signifikanzniveau α hat der L-R-Test die Maximale Power.

Also:

$$\begin{aligned}
\mathbb{P}(d=1\,|\tilde{H}_1)&\leq\mathbb{P}(d_{LR}=1\,|\tilde{H}_1)\\
\mathbb{P}(d=1\,|\mu_1)&\leq\mathbb{P}(d_{LR}=1\,|\mu_1)\\
\Pi(\mu_1)&\leq\Pi_{LR}(\mu_1)
\end{aligned}$$

□

Bemerkung: Die obige Aussage gilt auch für rechtsseitige Alternativen ($\mu > \mu_0$), sowie im Fall unbekannter Varianz, d.h. für den t-Test. Bei 2-Seitige Alternative gibt es keinen UMP.

Beispiel 2.9 Seien $X_1,\dots,X_n$ i.i.d. $N(\mu,\sigma^2)$, $\mu \in \mathbb{R}$, $\sigma^2 > 0$ bekannt. Teste $H_0\colon \mu = \mu_0$ gegen $H_1 : \mu \neq \mu_0$.
Hierfür existiert kein UMP-Test (Siehe hierzu auch R-Code vom 17.06.21).

Beweisidee. Fixiere $\alpha \in (0,1)$. Angenommen es gibt einen UMP-Test. Sei $d_{\ast}$ die diesem entsprechende $0-1$-wertige Zufallsvariable.

- Betrachte zunächst dem Fall $\mu_1 < \mu_0$, sowie das asymptotische Testproblem

  $$\tilde{H}_0\colon \mu < \mu_0$$
  $$\tilde{H}_1\colon \mu < \mu_1.$$

  Sei $\tilde{d}$ der L-R-Test zum Testen von $\tilde{H}_0$ gegen $\tilde{H}_1$. Mit dem Neyman-Pearson-Lemma ist

  $$\mathbb{P}\left(\tilde{d} = 1 \,|\mu = \mu_1\right) \geq \mathbb{P}(d_{\ast} = 1 \,|\mu = \mu_1).$$

  Weil $d_{\ast}$ UMP ist, gilt auch

  $$\mathbb{P}\left(\tilde{d} = 1 \,|\mu = \mu_1\right) \leq \mathbb{P}(d_{\ast} = 1 \,|\mu = \mu_1).$$

  Die Power von $d_{\ast}$ und von $\tilde{d}$ ist gleich.

  $\underset{\text{informell}}{\Rightarrow}$ Die Tests $d_{\ast}$ und d_{LR^-} stimmen überein.

  Nun ist $\tilde{d}$ den Z-Test (siehe letzte Beispiel):

  $$d_{\ast} = \begin{cases} 1 & : \tilde{d} = 1, \\ 0 & : \tilde{d} = 0, \end{cases}$$
  $$= \begin{cases} 1 & : \frac{\sqrt{n}}{\sigma}(\bar{X}_n - \mu_0) \leq \Phi^{-1}(\alpha) \\ 0 & : \text{sonst.} \end{cases}$$

- Betrachte nun den Fall $\mu_1 > \mu_0$ und argumentiere wie zuvor.

  Syntetisches Testproblem:

  $$\tilde{\tilde{H}}_0\colon \mu = \mu_0$$
  $$\tilde{\tilde{H}}_1\colon \mu = \mu_1.$$

  Ist $\tilde{\tilde{d}}$ der entsprechende L-R-Test, dann stimmen wie zuvor die Tests $\tilde{\tilde{d}}$ und $d_{\ast}$ überein. Wieder ist $\tilde{\tilde{d}}$ ein Z-Test:

$$d_{\divideontimes} = \begin{cases} 1 & : \tilde{\tilde{d}} = 1, \\ 0 & : \tilde{\tilde{d}} = 0, \end{cases}$$
$$= \begin{cases} 1 & : \frac{\sqrt{n}}{\sigma}(\bar{X}_n - \mu_0) \geq \Phi^{-1}(1-\alpha) \\ 0 & : \text{sonst.} \end{cases}$$

Die beiden Formeln für $d_{\divideontimes}$ widersprechen einander.

Damit kann $d_{\divideontimes}$ nicht einem UMP-Test entsprechen.

□

Bemerkung: Das oben beobachtete Phänomen tritt auch im Fall unbekannter Varianz (t-Test) auf. Darüber hinaus ist in vielen der hier beobachteten Modellen ($B(p), F_{m-1,n-1}$, etc.) bei einseitigen Alternativen der entsprechende L-R-Test UMP, und für zweiseitige Alternativen existiert kein UMP.

2.3 L-R-Tests in großen Stichproben

Betrachte: $X_1, \ldots, X_n$ i.i.d. mit Dichte $f(x|\theta)$ (bzw. Wahrscheinlichkeitsfunktion $p(x|\theta)$), $\theta \in \Theta \subseteq \mathbb{R}^p$.
Teste:

$$\begin{aligned} H_0 &: \theta \in \Theta_0 \quad \subseteq \Theta_1 \\ H_1 &: \theta \in \Theta \setminus \Theta_0. \end{aligned}$$

Betrachte der L-R-Statistik: $\Lambda = \frac{L(\hat{\theta}_0)}{L(\hat{\theta})}$
wobei: $\hat{\theta}_0 = \underset{\theta \in \Theta_0}{\operatorname{argmax}}\, L(\theta), \hat{\theta} = \underset{\theta \in \Theta}{\operatorname{argmax}}\, L(\theta),$.

Satz 2.2 (Wilks' Theorem) Unter geeigneten Voraussetzungen gilt unter H_0:

$$-2 \log \Lambda_n \xrightarrow{w} \chi^2_{k-k_0}$$

wobei k (bzw. k_0) die Anzahl der freien Parameter in Θ (bzw. Θ_0) ist. Oft $k = \dim \Theta$, $k_0 = \dim \Theta_0$.

$$H_0 : \mu = \mu \quad , k_0 = 0$$

$$>$$
$$H_1 : \mu < \mu_0 \quad , k = 1.$$
$$\neq$$

(Ohne Beweis)
Damit erhält man einen Test (H_0 gegen H_1) mit nominaler Signifikanzniveu α wie folgt:

$$\cancel{H_0} \text{ falls } -2 \log \Lambda \geq c$$

Wähle c sodaß,

$$\begin{aligned}
\alpha &\approx \mathbb{P}(-2\log\Lambda_n \geq c|H_0)\\
&= 1 - \mathbb{P}(-2\log\Lambda_n \leq c|H_0)\\
&\approx 1 - \mathbb{P}(\chi^2_{k-k_0} < c) = \%\\
&\uparrow \text{Wilks}
\end{aligned}$$

Setze $c = \mathrm{F}^{-1}_{k-k_0}(1-\alpha) \to$ Quantil der $\chi^2_{k-k_0}$

$$\begin{aligned}
\Rightarrow \% &= 1 - \mathbb{P}(\chi^2_{k-k_0} \leq c)\\
&= 1 - (1-\alpha) = \alpha\\
&\Rightarrow \mathbb{P}(\cancel{H_0}|H_0) \approx \alpha
\end{aligned}$$

Für $n- > \infty$ näher sich das Signifikanzniveu dieses Test dem Wert α an:

$$\begin{aligned}
&\lim_{n\to\infty} \mathbb{P}(-2\log\Lambda_n \geq c|H_0)\\
&= \lim_{n\to\infty} 1 - \mathbb{P}(-2\log\Lambda_n < c|H_0)\\
&= 1 - \lim_{n\to\infty} \mathbb{P}(-2\log\Lambda_n < c|H_0)\\
&= 1 - \underbrace{\mathbb{P}(\chi^2_{k-k_0} < c)}_{1-\alpha}\\
&= \alpha \checkmark
\end{aligned}$$

Beispiele für $k, k_0 : \lambda_i$, iid, $\mathbb{N}(\mu, \sigma^2)$

$$\begin{aligned}
&H_0 : \mu = \mu_0\\
&H_1 : \mu \neq \mu_0\\
&\Theta = \mathbb{R} \times (0,\infty) ... k = 2\\
&\Theta_0 = \{\mu_0\} \times (0,\infty) ... k = 1
\end{aligned}$$

Beispiel 2.10 (Kategoriale Daten, einfache Nullhypothese) Betrachte:
$X_1, ..., X_n$ iid diskret mit Werten in $\{1, .., m\}$.
Wahrscheinlichkeitsfunktion $\underset{i \leq j \leq m}{p(j) = \mathbb{P}(X_i = j)}, \theta = (p(j))^m_{j=1}$ unbekannt.
Weiters sei $p_0(j)$, $1 \leq j \leq m$, eine bestimmte Wahrscheinlichkeitsfunktion. Teste:

$H_0 : p(j) = p_0(j)$ für alle $j = 1,..,m$
$H_1 : p(j) \neq p_0(j)$ für mindestens ein $j \in \{1,..,m\}$

$$\Theta = \left\{(p(j))_{j=1}^m \in [0,1]^m : \sum_{j=1}^m p(j) = 1\right\} ...k = 1$$

$$\Theta_0 = \left\{(p(j))_{j=1}^m\right\} ...k_0 = 0$$

$$\hat{\theta}_0 = \underset{\theta=\Theta_0}{\operatorname{argmax}} L(\theta) = (p_0(j))_{j=1}^m$$

$$\hat{\theta} = \underset{\theta=\Theta}{\operatorname{argmax}} L(\theta) = \left(\frac{O_j}{n}\right)_{j=1}^m \text{, wobei } = O_j = \#\{i = 1, ..., n : \ X_i = j\},\ 1 \leq j \leq m$$

O_j " Observed number of cases in class j "

Allgemein ist:

$$L((p(j))_{j=1}^m) = p(x_1) \cdot p(x_2) \cdot ... \cdot p(x_n) =$$
$$p(1)^{O_1} \cdot p(2)^{O_2} \cdot ... \cdot p(m)^{O_m}$$

$$-2\log\Lambda = -2\log\frac{L(\hat{\theta}_0)}{L(\hat{\theta})} = 2\log\frac{L(\hat{\theta})}{L(\hat{\theta}_0)}$$

$$2\log\frac{\prod_{j=1}^m (\frac{O_j}{n})^{O_j}}{\prod_{j=1}^m (p_0(j))^{O_j}} = 2\log\prod_{j=1}^m (\frac{O_j}{n \cdot p_0(j)})^{O_j} =$$

$$2\sum_{j=1}^m O_j \log\frac{O_j}{np_0(j)} = 2\sum_{j=1}^m O_j \log\frac{O_j}{E_j},$$

Wobei $E_j = n \cdot p_0(j),\ 1 \leq j \leq m$"Expected number of cases in class j under H_0"

Grenzverteilung von $2\sum_{j=1}^m O_j \log\frac{O_j}{E_j}$ ist die χ^2_{m-1}-Verteilung.

Bemerkung: Im Kontext des letzten Beispiels wird oft Pearson's χ^2-Anpassungstest verwendet.
Die Statistik ist: $\chi^2 := \sum_{j=1}^m \frac{(E_j - O_j)^2}{E_j} (\chi^2 \leftarrow$ nicht eine Verteilung) .
Unter H_0 gilt $\chi^2 \xrightarrow{w} \chi^2_{m-1}$.
Tatsächlich sind hier der L-R-Test und der χ^2-Anpassungstest äquivalent.

Bemerkung: Die Approximation
$\mathbb{P}(-2\log\Lambda \leq t|H_0) \approx \mathbb{P}(\chi^2_{m-1} \leq t)$
$\mathbb{P}(\chi^2 \leq t|H_0) \approx \mathbb{P}(\chi^2_{m-1} \leq t)$
werden klein, wenn für jedes $j = 1,..,m$ die größte $E_j = n \cdot p_0(j)$ größ ist. Daumenregel: $E_j \geq S^j$

Falls die Daumenregeln verletzt ist, kann man "Klassen zusammenlegen".

Beispiel 2.11 $\lambda_1, ...\lambda_n$ iid $\mathbb{N}(\mu, \sigma^2)$, $\sigma^2 > 0$ bekannt. Teste:

$$\begin{aligned} H_0 : \mu = \mu_0 \quad & \Theta_0 = \{\mu_0\} \\ H_1 : \mu \neq \mu_0 \quad & \Theta = \mathbb{R} \end{aligned}$$

$\hat{\mu}_0$...ML Schätzer für μ_0 unter $H_0 = \mu_0$
$\hat{\mu}$...ML Schätzer "unrestringiert" $= \bar{X}_n$

$$\begin{aligned}
\Lambda &= \frac{\prod_{i=1}^n (2\pi\sigma^2)^{-\frac{1}{2}} e^{-\frac{1}{2\sigma^2}(X_i-\mu_0)^2}}{\prod_{i=1}^n (2\pi\sigma^2)^{-\frac{1}{2}} e^{-\frac{1}{2\sigma^2}(X_i-\bar{X}_n)^2}} = e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n\left((X_i-\mu_0)^2-(X_i-\bar{X}_n)^2\right)} \\
&= -2\log\Lambda = -2\frac{(-1)}{2\sigma^2}\sum_{i=1}^n\left(X_i^2 - 2X_i\mu_0 + \mu_0^2 - X_i^2 + 2X_i\bar{X}_n - \bar{X}_n^2\right) \\
&= \frac{1}{\sigma^2}(-2\mu_0 n\bar{X}_n + n\mu_0^2 + \underbrace{2\bar{X}_n n\bar{X}_n - n\bar{X}_n^2}_{n\bar{X}_n^2}) \\
&= \frac{n}{\sigma^2}(\mu_0^2 - 2\mu_0\bar{X}_n + \bar{X}_n^2) \\
&= \frac{n}{\sigma^2}(\bar{X}_n - \mu_0)^2 \\
&= \left(\frac{\sqrt{n}(\bar{X}_n - \mu_0)}{\sigma}\right)^2 = (\ast)
\end{aligned}$$

Unter H_0 ist

$$\begin{aligned}
&\bar{X}_n \ \mathbb{N}\left(\mu_0, \frac{\sigma^2}{n}\right) \\
&\bar{X}_n - \mu_0 \ \mathbb{N}\left(0, \frac{\sigma^2}{n}\right) \\
&\frac{\sqrt{n}}{\sigma}\left(\bar{X}_n - \mu_0\right) \ \mathbb{N}(0,1) \\
&\Rightarrow -2\log\Lambda = (\ast) = \chi^2.
\end{aligned}$$

3. Zufallsvektoren und Zufallsmatrizen

Siehe zu diesem Kapitel S. 564-574. in *Rice, J. A. (2007). Mathematical statistics and data analysis. Belmont, CA: Thomson/Brooks/Cole.*

Dieses Kapitel haben wir im SS 2022 nicht gemacht, es bleibt aber vom Vorjahr hier!

Wissen: Für $X_1,\dots,X_n$ i.i.d. $N(\mu,\sigma^2)$, $\mu\in\mathbb{R}$, $\sigma^2>0$, ist

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \sim N\left(\mu, \frac{\sigma^2}{n}\right),$$

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

und weiters sind $\bar{X}_n$ und $\hat{\sigma}_n^2$ unabhängig. Damit ist insbesondere

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

3.1 Erwartungswerte und Varianz/Kovarianz-Matrizen

Definition 3.1 (Erwartungswert) Für zufällige Vektoren bzw. Matrizen wird der Erwartungswert komponentenweise definiert. Also:
(i) Ist $\mathbf{X}=(X_1,\dots,X_n)'$ ein Zufallsvektor, dann ist

$$\mathbb{E}(\mathbf{X}) = \begin{pmatrix}\mathbb{E}(X_1)\\ \mathbb{E}(X_2)\\ \vdots \\ \mathbb{E}(X_n)\end{pmatrix}.$$

(ii) Ist $\mathbf{M}=(M_{ij})_{i=1,j=1}^{n\ \ m}$ eine zufällige Matrix, dann ist

$$\mathbb{E}(\mathbf{M}) = \mathbb{E}\Big((M_{ij})\Big)_{i=1,j=1}^{n\ \ m}$$

Bemerkung: (i) ist ein Spezialfall von (ii), da man den n-Vektor $\mathbf{X}$ auch als $n\times 1$ Matrix betrachten kann.

Für Zufallsvariable Z ist der Erwartungswert linear für $a, b \in \mathbb{R}$,

$$\mathbb{E}(aZ + b) = a\mathbb{E}(Z) + b.$$

Proposition 3.1 (Linearität des Erwartungswertes) (i) Sei $\mathbf{X} = (X_1, \ldots, X_n)'$ ein Zufallsvektor, $\mathbf{A}$ eine $m \times n$-Matrix und $\mathbf{b}$ ein m-dimensionaler Vektor, dann ist

$$\mathbb{E}(\mathbf{AX} + \mathbf{b}) = \mathbf{A}\mathbb{E}(\mathbf{X}) + \mathbf{b}.$$

(ii) Sei $\mathbf{M}$ eine zufällige $(n \times m)$ Matrix, $\mathbf{A}$ eine $k \times n$ Matrix und $\mathbf{B}$ eine $k \times m$ Matrix, dann ist

$$\mathbb{E}(\mathbf{AM} + \mathbf{B}) = \mathbf{A}\mathbb{E}(\mathbf{M}) + \mathbf{B}.$$

Ist $\mathbf{C}$ eine $m \times k$ Matrix und $\mathbf{D}$ eine $n \times k$ Matrix, dann ist

$$\mathbb{E}(\mathbf{MC} + \mathbf{D}) = \mathbb{E}(\mathbf{M})\mathbf{C} + \mathbf{D}.$$

Beweis. Es genügt, (ii) zu zeigen, da (i) ein Spezialfall von (ii) ist.
Zeige: $\mathbb{E}(\mathbf{AM} + \mathbf{B}) = \mathbf{A}\mathbb{E}(\mathbf{M}) + \mathbf{B}$.
Vergleiche Eintrag in Zeile i und Spalte j links und rechts.

$$\begin{aligned}
(\mathbb{E}(AM + B))_{ij} &= \mathbb{E}\left((AM + B)_{ij}\right) \\
&= \mathbb{E}\left((AM)_{ij} + B_{ij}\right) \\
&= \mathbb{E}\left(\sum_{l=1}^{n} A_{il} M_{lj} + B_{ij}\right) \\
&= \sum_{l=1}^{n} A_{il} \mathbb{E}\left(M_{lj}\right) + B_{ij} \\
&= (A\mathbb{E}(M))_{ij} + B_{ij} \\
&= (A\mathbb{E}(M) + B)_{ij}.
\end{aligned}$$

Zeige: $\mathbb{E}(\mathbf{MC} + \mathbf{D}) = \mathbb{E}(\mathbf{M})\mathbf{C} + \mathbf{D}$.

$$\begin{aligned}
(\mathbb{E}(MC + D))_{ij} &= \mathbb{E}\left((MC + D)_{ij}\right) \\
&= \mathbb{E}\left((MC)_{ij} + D_{ij}\right) \\
&= \mathbb{E}\left(\sum_{l=1}^{m} M_{il} C_{lj} + D_{ij}\right) \\
&= \sum_{l=1}^{m} \mathbb{E}(M_{il}) C_{lj} + D_{ij} \\
&= (\mathbb{E}(M)C)_{ij} + D_{ij} \\
&= (\mathbb{E}(M)C + D)_{ij}.
\end{aligned}$$

□

Beispiel 3.1 $X_1, \ldots, X_n$ i.i.d. mit $\mathbb{E}(X_1) = \mu$.

Betrachte $\mathbf{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}$.

$$\mathbb{E}(\mathbf{X}) = \begin{pmatrix} \mathbb{E}(X_1) \\ \vdots \\ \mathbb{E}(X_n) \end{pmatrix} = \begin{pmatrix} \mu \\ \vdots \\ \mu \end{pmatrix} = \mu\iota$$

für $\iota = \begin{pmatrix} 1, & \ldots, & 1 \end{pmatrix}' \in \mathbb{R}^n$.

$$\mathbb{E}\left(\frac{1}{n}\iota'\mathbf{X}\right) = \frac{1}{n}\iota'\mathbb{E}(\mathbf{X}) = \frac{1}{n}\iota'\mu\iota = \frac{\mu}{n}\underbrace{\iota'\iota}_{=n} = \mu.$$

$$\mathbf{X} = \iota\bar{X} + (\mathbf{X} - \iota\bar{X})$$

orthogonal!

Beispiel 3.2

$$\begin{aligned} \mathbb{E}(\underbrace{\iota\bar{X}}_{(n\times 1)}) &= \mathbb{E}\left(\iota\frac{1}{n}\iota'\mathbf{X}\right) \\ &= \frac{1}{n}\iota\iota'\mathbb{E}(\mathbf{X}) \\ &= \frac{1}{n}\iota\iota'\mu\iota \\ &= \mu\frac{1}{n}\iota\iota'\iota \\ &= \mu\iota. \end{aligned}$$

Beispiel 3.3

$$\begin{aligned}
\mathbb{E}(\mathbf{X} - \iota\bar{X}) &= \mathbb{E}\left(\mathbf{X} - \iota\frac{1}{n}\iota'\mathbf{X}\right) \\
&= \mathbb{E}\left(\left(\boldsymbol{I}_n - \iota\frac{1}{n}\iota'\right)\mathbf{X}\right) \\
&= \left(\boldsymbol{I}_n - \frac{1}{n}\iota\iota'\right)\mathbb{E}(\mathbf{X}) \\
&= \left(\boldsymbol{I}_n - \frac{1}{n}\iota\iota'\right)\mu\iota \\
&= \mu\left(\iota - \frac{1}{n}\iota\iota'\iota\right) \\
&= \mu(\iota - \iota) \\
&= 0.
\end{aligned}$$

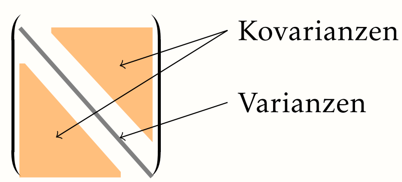
Definition 3.2 (Varianz/Kovarianz-Matrix) Ist $\mathbf{X}_{(n\times 1)}$ ein Zufallsvektor, dann ist die so genannte Varianz/Kovarianz-Matrix von $\mathbf{X}$ die $n \times n$ Matrix

$$\mathrm{VC}(\mathbf{X}) = \left(\mathrm{VC}(X)_{ij}\right)_{i=1}^{n} {}_{j=1}^{n},$$

wobei

$$\mathrm{VC}(\mathbf{X}) = \begin{cases} \mathrm{Var}(X_i) & : i = j \\ \mathrm{Cov}(X_i, X_j) & : i \neq j \end{cases}$$

mit $1 \leq i, j \leq n$.

VC(**X**) =


Bemerkung: Weil $\mathrm{Cov}\left(X_i, X_j)\right) = \mathrm{Cov}\left(X_j, X_i\right)$ ist, ist $\mathrm{VC}(\mathbf{X})$ immer symmetrisch:

$$(\mathrm{VC}(\mathbf{X}))' = \mathrm{VC}(\mathbf{X})$$

Bemerkung: Ist $\mathbf{X}$ ein 1-dimensionaler Vektor, dann ist $\mathrm{VC}(\mathbf{X})$ die 1×1 Matrix $\mathrm{Var}(\mathbf{X})$.

Proposition 3.2

$$\mathrm{VC}(\mathbf{X}) = \mathbb{E}\left((\mathbf{X} - \mathbb{E}(\mathbf{X}))(\mathbf{X} - \mathbb{E}(\mathbf{X}))'\right).$$

Beweis. Sei $i, j \in \{1, \ldots, n\}$.

$$
\begin{aligned}
(\mathbb{E}(X-\mathbb{E}(X))(X-\mathbb{E}(X))')_{ij} &= \left(\mathbb{E}(X-\mathbb{E}(X))_i(X-\mathbb{E}(X))_j\right) \\
&= \left(\mathbb{E}(X_i-\mathbb{E}(X_i))(X_j-\mathbb{E}(X_j))\right) \\
&= \operatorname{Cov}(X_i, X_j) \\
&= (\operatorname{VC}(X))_{ij}.
\end{aligned}
$$

Beachte: Für $i = j$ ist $\operatorname{Cov}(X_i, X_j) = \operatorname{Var}(X_i)$.

□

Proposition 3.3 Ist $\mathbf{X}$ ein n-dimensionaler Zufallsvektor, $\mathbf{A}$ eine $m \times n$ Matrix und $\mathbf{b} \in \mathbb{R}^m$, dann ist $\operatorname{VC}(\mathbf{AX}+\mathbf{b}) = \mathbf{A}\operatorname{VC}(\mathbf{X})\mathbf{A}'$

Beweis.

$$
\begin{aligned}
\operatorname{VC}(\mathbf{AX}+\mathbf{b}) &= \mathbb{E}((\mathbf{AX}+\mathbf{b}-\mathbb{E}(\mathbf{AX}+\mathbf{b}))(\mathbf{AX}+\mathbf{b}-\mathbb{E}(\mathbf{AX}+\mathbf{b}))')) \\
&= \mathbb{E}((\mathbf{AX}+\mathbf{b}-(\mathbf{A}\mathbb{E}(\mathbf{X})+\mathbf{b})(\mathbf{AX}+\mathbf{b}-(\mathbf{A}\mathbb{E}(\mathbf{X})+\mathbf{b})')) \\
&= \mathbb{E}((\mathbf{AX}-\mathbf{A}\mathbb{E}(\mathbf{X}))(\mathbf{AX}-(\mathbf{A}\mathbb{E}(\mathbf{X}))')) \\
&= \mathbb{E}(\mathbf{A}(\mathbf{X}-\mathbb{E}(\mathbf{X}))(\mathbf{X}-(\mathbb{E}(\mathbf{X}))')\mathbf{A}') \\
&= \mathbf{A}\mathbb{E}((\mathbf{X}-\mathbb{E}(\mathbf{X}))(\mathbf{X}-(\mathbb{E}(\mathbf{X}))')\mathbf{A}') \\
&= \mathbf{A}\mathbb{E}((\mathbf{X}-\mathbb{E}(\mathbf{X}))(\mathbf{X}-(\mathbb{E}(\mathbf{X}))')\mathbf{A}' \\
&= \mathbf{A}\operatorname{VC}(\mathbf{X})\mathbf{A}'.
\end{aligned}
$$

□

Beispiel 3.4 $X_1, \ldots, X_n$ i.i.d., $\mathbb{E}(X_1) = \mu$, $\operatorname{Var}(X_1) = \sigma^2$. $\mathbf{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}$

$$
\mathbb{E}(\mathbf{X}) = \begin{pmatrix} \mu \\ \vdots \\ \mu \end{pmatrix} = \mu\iota.
$$

$$
\operatorname{VC}(\mathbf{X}) = \begin{pmatrix} \sigma^2 & & & \\ & \ddots & & \mathbf{0} \\ & & \ddots & \\ & \mathbf{0} & & \ddots & \\ & & & & \sigma^2 \end{pmatrix} = \sigma^2 \mathbf{I}_n
$$

Betrachte $\bar{X}$, $\mathbf{X} - \iota\bar{X}$.
$\begin{pmatrix} \bar{\mathbf{X}} \\ \mathbf{X} - \iota\bar{X} \end{pmatrix} \ldots$ ein $(n+1)$-dimensionaler Vektor.

$$
\begin{pmatrix} \mathbb{E}\left(\bar{X}\right) \\ \mathbb{E}\left(\mathbf{X} - \iota\bar{X}\right) \end{pmatrix} = \begin{pmatrix} \mu \\ \mathbf{0} \end{pmatrix}.
$$

Varianzkovarianzmatrix:

$$
\begin{aligned}
\mathrm{VC}(\mathbf{Z}) &= \mathbb{E}\left[\begin{pmatrix} \bar{X}-\mu \\ \mathbf{X}-\iota\bar{X}-\mathbf{0}\end{pmatrix}\begin{pmatrix} \bar{X}-\mu & \left(\mathbf{X}-\iota\bar{X}-\mathbf{0}\right)\end{pmatrix}'\right] \\
&= \mathbb{E}\left[\begin{pmatrix} \bar{X}-\mu \\ \mathbf{X}-\iota\bar{X}\end{pmatrix}\begin{pmatrix} \left(\bar{X}-\mu\right) & \left(\mathbf{X}-\iota\bar{X}\right)\end{pmatrix}'\right] \\
&= \mathbb{E}\left[\begin{pmatrix} \left(\bar{X}-\mu\right)\left(\bar{X}-\mu\right) & \left(\bar{X}-\mu\right)\left(\mathbf{X}-\iota\bar{X}\right)' \\ \left(\mathbf{X}-\iota\bar{X}\right)\left(\bar{X}-\mu\right) & \left(\mathbf{X}-\iota\bar{X}\right)\left(\mathbf{X}-\iota\bar{X}\right)'\end{pmatrix}\right] \\
&= \begin{pmatrix} \mathbb{E}\left(\bar{X}-\mu\right)^2 & \mathbb{E}\left(\left(\bar{X}-\mu\right)\left(\mathbf{X}-\iota\bar{X}\right)'\right) \\ \mathbb{E}\left(\mathbf{X}-\iota\bar{X}\right)\left(\bar{X}-\mu\right) & \mathbb{E}\left(\left(\mathbf{X}-\iota\bar{X}\right)\left(\mathbf{X}-\iota\bar{X}\right)'\right)\end{pmatrix} \\
&= \begin{pmatrix} \mathrm{Var}(\bar{X}) & \mathbb{E}\left(\left(\bar{X}-\mu\right)\left(\mathbf{X}-\iota\bar{X}\right)'\right) \\ \mathbb{E}\left(\mathbf{X}-\iota\bar{X}\right)\left(\bar{X}-\mu\right) & \mathbb{E}\left(\left(\mathbf{X}-\iota\bar{X}\right)\left(\mathbf{X}-\iota\bar{X}\right)'\right)\end{pmatrix}.
\end{aligned}
$$

$$
\mathrm{Var}(\bar{X}) = \frac{\sigma^2}{n}.
$$

$$
\mathbb{E}\left(\left(\bar{X}-\mu\right)\left(\mathbf{X}-\iota\bar{X}\right)'\right) = ?
$$

Zwei Zwischenschritte:

$$
\begin{aligned}
\bar{X}-\mu &= \left(\frac{1}{n}\iota'\mathbf{X}\right) - \left(\frac{1}{n}\iota'\mu\iota\right) \\
&= \left(\frac{1}{n}\iota'\mathbf{X}\right) - \left(\frac{1}{n}\iota'\mathbb{E}(\mathbf{X})\right) \\
&= \frac{1}{n}\iota'\left(\mathbf{X}-\mathbb{E}(\mathbf{X})\right),
\end{aligned}
$$

und

$$
\begin{aligned}
\mathbf{X}-\iota\bar{X} &= \mathbf{X}-\iota\frac{1}{n}\iota'\mathbf{X} \\
&= \left(\mathbf{I}_n - \iota\frac{1}{n}\iota'\right)\mathbf{X} \\
&= \left(\mathbf{I}_n - \iota\frac{1}{n}\iota'\right)\left(\mathbf{X}-\mathbb{E}(\mathbf{X})\right),
\end{aligned}
$$

da

$$
\begin{aligned}
\left(\mathbf{I}_n - \iota\frac{1}{n}\iota'\right)\mathbb{E}(\mathbf{X}) &= \left(\mathbf{I}_n - \iota\frac{1}{n}\iota'\right)\mu\iota \\
&= \mu\left(\mathbf{I}_n\iota - \frac{1}{n}\iota\iota'\iota\right) \\
&= \mu\left(\iota - \frac{n}{n}\iota\right) \\
&= \mu\left(\iota-\iota\right) \\
&= \mathbf{0}.
\end{aligned}
$$

$$\begin{aligned}
\mathbb{E}\left(\left(\bar{X}-\mu\right)\left(\mathbf{X}-\iota\bar{X}\right)'\right) &= \mathbb{E}\left(\left(\frac{1}{n}\iota'(\mathbf{X}-\mathbb{E}(\mathbf{X}))\right)\left((\mathbf{X}-\mathbb{E}(\mathbf{X}))'\left(\mathbf{I}_n-\iota\frac{1}{n}\iota'\right)\right)\right)\\
&= \frac{1}{n}\iota'\mathbb{E}\left(((\mathbf{X}-\mathbb{E}(\mathbf{X})))\left((\mathbf{X}-\mathbb{E}(\mathbf{X}))'\right)\right)\left(\mathbf{I}_n-\iota\frac{1}{n}\iota'\right)\\
&= \frac{1}{n}\iota' VC(\mathbf{X})\left(\mathbf{I}_n-\iota\frac{1}{n}\iota'\right)\\
&= \sigma^2\frac{1}{n}\iota'\left(\mathbf{I}_n-\iota\frac{1}{n}\iota'\right)\\
&= \sigma^2\frac{1}{n}\left(\iota'-\iota'\iota\frac{1}{n}\iota'\right)\\
&= \sigma^2\frac{1}{n}(\iota'-\iota')\\
&= \mathbf{0}.
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}\left(\left(\mathbf{X}-\iota\bar{X}\right)\left(\mathbf{X}-\iota\bar{X}\right)'\right) &= VC(\mathbf{X}-\iota\bar{X})\\
&= VC(\mathbf{X}-\iota\frac{1}{n}\iota'\mathbf{X})\\
&= VC\left(\left(\mathbf{I}_n-\iota\frac{1}{n}\iota'\right)\mathbf{X}\right)\\
&= \left(\mathbf{I}_n-\iota\frac{1}{n}\iota'\right)VC(\mathbf{X})\left(\mathbf{I}_n-\iota\frac{1}{n}\iota'\right)'\\
&= \left(\mathbf{I}_n-\iota\frac{1}{n}\iota'\right)\sigma^2\left(\mathbf{I}_n-\iota\frac{1}{n}\iota'\right)'\\
&= \sigma^2\left(\mathbf{I}_n-\iota\frac{1}{n}\iota'\right)\left(\mathbf{I}_n-\iota\frac{1}{n}\iota'\right)'\\
&= \sigma^2\left(\mathbf{I}_n-\iota\frac{1}{n}\iota'-\iota\frac{1}{n}\iota'+\iota\frac{1}{n}\iota'\iota\frac{1}{n}\iota'\right)\\
&= \sigma^2\left(\mathbf{I}_n-\iota\frac{1}{n}\iota'-\iota\frac{1}{n}\iota'+\iota\frac{1}{n}\iota'\right)\\
&= \sigma^2\left(\mathbf{I}_n-\iota\frac{1}{n}\iota'\right).
\end{aligned}$$

Also:

$$\mathrm{VC}\begin{pmatrix}\bar{X}\\ \mathbf{X}-\iota\bar{X}\end{pmatrix} = \begin{pmatrix}\frac{\sigma^2}{n} & \mathbf{0}'\\ \mathbf{0}' & \sigma^2\left(\mathbf{I}_n-\iota\frac{1}{n}\iota'\right)\end{pmatrix}.$$

Man sieht: $\bar{X}$ und $\mathbf{X}-\iota\bar{X}$ sind komponentenweise unkorreliert.

Bezeichnung: Für zwei Zufallsvektoren $\mathbf{X}\in\mathbb{R}^n$ und $\mathbf{Y}\in\mathbb{R}^m$ ist die $n\times m$ Matrix $\mathrm{Cov}(\mathbf{X},\mathbf{Y})$ definiert als $\mathrm{Cov}(\mathbf{X},\mathbf{Y})=\mathbb{E}((\mathbf{X}-\mathbb{E}(\mathbf{X}))(\mathbf{Y}-\mathbb{E}(\mathbf{Y}))')$.

Weiters ist die $m\times n$ Matrix $\mathrm{Cov}(\mathbf{Y},\mathbf{X})$ definiert als $\mathrm{Cov}(\mathbf{Y},\mathbf{X})=\mathbb{E}((\mathbf{Y}-\mathbb{E}(\mathbf{Y}))(\mathbf{X}-\mathbb{E}(\mathbf{X}))')$.

Beachte: $\mathrm{Cov}(\mathbf{X},\mathbf{Y})=(\mathrm{Cov}(\mathbf{Y},\mathbf{X}))'$

Bemerkung: Zwei Zufallsvektoren $\mathbf{X}, \mathbf{Y}$ sind unabhängig, wenn für beliebige Funktionen $f(x)$ und $g(y)$ die Beziehung $\mathbb{E}(f(x)g(y)) = \mathbb{E}(f(x))\mathbb{E}(g(y))$ gilt.

Proposition 3.4 Sind $\mathbf{X}, \mathbf{Y}$ unabhängige Zufallsvektoren, dann ist $\mathrm{Cov}(\mathbf{X}, \mathbf{Y}) = \mathbf{0}$.

Beweis. Siehe Übung. □

Proposition 3.5 Ist $\mathbf{X}$ ein Zufallsvektor mit $\mathrm{VC}(\mathbf{X}) = \Sigma$, dann gilt

$$\begin{aligned} \Sigma' &= \Sigma, && \text{(symetrisch)} \\ \Sigma &\geq 0. && \text{(positiv semidefinit)} \end{aligned}$$

Beweis. Sei $\mathbf{X}$ ein n-dimensionaler Zufallsvektor.

$$\begin{aligned} \Sigma' &= (\mathrm{VC}(\mathbf{X}))' \\ &= (\mathbb{E}((\mathbf{X} - \mathbb{E}(\mathbf{X})))(\mathbf{X} - \mathbb{E}(\mathbf{X})'))' \\ &= \mathbb{E}((\mathbf{X} - \mathbb{E}(\mathbf{X}))(\mathbf{X} - \mathbb{E}(\mathbf{X}))') \\ &= \mathrm{VC}(\mathbf{X}) \\ &= \Sigma. \checkmark \end{aligned}$$

$\Sigma \geq 0 \ldots$ d.h. für jeden Vektor $\boldsymbol{\alpha} \in \mathbb{R}^n$ ist $\boldsymbol{\alpha}'\Sigma\boldsymbol{\alpha} \geq 0$. Sei also $\boldsymbol{\alpha} \in \mathbb{R}^n \setminus \{0\}$. Betrachte die Zufallsvariable $\boldsymbol{\alpha}'\mathbf{X}$:

$$\begin{aligned} 0 \leq \mathrm{Var}(\boldsymbol{\alpha}'\mathbf{X}) &= \mathrm{VC}(\boldsymbol{\alpha}'\mathbf{X}) \\ &= \boldsymbol{\alpha}'\mathrm{VC}(\mathbf{X})\boldsymbol{\alpha} \\ &= \boldsymbol{\alpha}'\Sigma\boldsymbol{\alpha}. \checkmark \end{aligned}$$

□

Bemerkung: Ist $\Sigma = \mathrm{VC}(\mathbf{X}) > 0$, dann ist für jeden Vektor $\boldsymbol{\alpha} \in \mathbb{R}^n$, $\boldsymbol{\alpha} \neq \mathbf{0}$, die Varianz von $\boldsymbol{\alpha}'\mathbf{X}$ positiv (und umgekehrt): Für $\boldsymbol{\alpha} \in \mathbb{R}^n \setminus \{0\}$ ist $\mathrm{Var}(\boldsymbol{\alpha}'\mathbf{X}) = \boldsymbol{\alpha}'\Sigma\boldsymbol{\alpha}$.

Erinnerung: Sei $\mathbf{A}$ eine symmetrische Matrix $(n \times n)$.
Seien $\lambda_1, \ldots, \lambda_n$ die Eigenwerte, und seien $\mathbf{u}_1, \ldots, \mathbf{u}_n$ die entsprechenden Eigenvektoren.
Setze $\mathbf{U} = (\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n)_{(n \times n)}$,
sowie

$$\boldsymbol{\Omega} = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}_{(n \times n)} .$$

Dann gilt: $\mathbf{A} = \mathbf{U}\boldsymbol{\Omega}\mathbf{U}'$.

Mit R-Beispiel vom 08.03.2021 (oder auch ganz allgemein) sieht man:
Sind λ_i die Eigenwerte und $\mathbf{u}_i$ die Eigenvektoren von $\mathbf{A}_{(n\times n)}$, dann gilt

$$\begin{aligned}
\mathbf{A}\mathbf{u}_i \quad &= \lambda_i \mathbf{u}_i, \quad 1 \le i \le n. \\
&\equiv \mathbf{A}\underbrace{(\mathbf{u}_1,\dots,\mathbf{u}_n)}_{\mathbf{U}} && = (\lambda_1\mathbf{u}_1,\dots,\lambda_n\mathbf{u}_n) \\
&\equiv \mathbf{A}\mathbf{U} && = \mathbf{U}\boldsymbol{\Omega} \\
&\Rightarrow \mathbf{A}\ \underbrace{\mathbf{U}\mathbf{U}'}_{\mathbf{I}_n} && = \mathbf{U}\boldsymbol{\Omega}\mathbf{U}'
\end{aligned}$$

Also: $\mathbf{A} = \mathbf{U}\boldsymbol{\Omega}\mathbf{U}'$.

$$\mathbf{A}x = \mathbf{U}\boldsymbol{\Omega}\mathbf{U}'x$$

Rücktransformation

Stauchung/Streckung

Transformation in neues Koordinatensystem

$$\mathbf{A}^2 = \mathbf{A}\mathbf{A} = \mathbf{U}\boldsymbol{\Omega}\mathbf{U}'\mathbf{U}\boldsymbol{\Omega}\mathbf{U}' = \mathbf{U}\boldsymbol{\Omega}^2\mathbf{U}'.$$

$$\mathbf{A}^{\frac{1}{2}} := \mathbf{U}\underset{\substack{\uparrow \\ \operatorname{diag}\left(\sqrt{\lambda_1},\dots,\sqrt{\lambda_n}\right)}}{\boldsymbol{\Omega}^{\frac{1}{2}}}\mathbf{U}'.$$

$$\mathbf{A}^{\frac{1}{2}}\mathbf{A}^{\frac{1}{2}} = \mathbf{U}\boldsymbol{\Omega}^{\frac{1}{2}}\mathbf{U}'\mathbf{U}\boldsymbol{\Omega}^{\frac{1}{2}}\mathbf{U}' = \mathbf{U}\boldsymbol{\Omega}^{\frac{1}{2}}\boldsymbol{\Omega}^{\frac{1}{2}}\mathbf{U}' = \mathbf{A}.$$

Ist $\mathbf{A}$ invertierbar, dann sind alle $\lambda_i \neq 0$, sodass $\boldsymbol{\Omega}^{-1} = \operatorname{diag}\left(\frac{1}{\lambda_1},\dots,\frac{1}{\lambda_n}\right)$ wohldefiniert ist.
Dann ist aber $\mathbf{A}^{-1} = \mathbf{U}\boldsymbol{\Omega}^{-1}\mathbf{U}'$, denn

$$\begin{aligned}
\mathbf{A}\mathbf{U}\boldsymbol{\Omega}^{-1}\mathbf{U} = \mathbf{U}\boldsymbol{\Omega}\ \underbrace{\mathbf{U}'\mathbf{U}}_{\mathbf{I}_n}\ \boldsymbol{\Omega}^{-1}\mathbf{U}' \quad &= \mathbf{U}\mathbf{U}' = \mathbf{I}_n. \\
\mathbf{U}\boldsymbol{\Omega}^{-1}\mathbf{U}'\mathbf{A} = \dots \quad &= \mathbf{I}_n.
\end{aligned}$$

Weiters ist dann

$$\mathbf{A}^{-\frac{1}{2}} = \mathbf{U}\underset{\substack{\uparrow \\ \operatorname{diag}\left(\frac{1}{\sqrt{\lambda_1}},\dots,\frac{1}{\sqrt{\lambda_n}}\right)}}{\boldsymbol{\Omega}^{-\frac{1}{2}}}\mathbf{U}'$$

ebenfalls wohldefiniert und erfüllt die Beziehung

$$\mathbf{A}^{-\frac{1}{2}}\mathbf{A}\mathbf{A}^{-\frac{1}{2}} = \mathbf{I}_n \quad \text{(Details: Übung.)}$$

Bemerkung: Ist $\mathbf{X}$ ein n-dimensionaler Zufallsvektor mit Mittelwert 0 und VC-Matrix $\Sigma > 0$, dann gilt für den Zufallsvektor $\mathbf{Y} = \Sigma^{-\frac{1}{2}}\mathbf{X}$ (n-dim.):

$$\begin{aligned}
\mathbb{E}(\mathbf{Y}) &= \mathbf{0} \\
\mathrm{VC}(\mathbf{Y}) &= \mathbf{I}_n.
\end{aligned}$$

Nachrechnen.

$$\mathbb{E}(\mathbf{Y}) = \mathbb{E}\left(\Sigma^{-\frac{1}{2}}\mathbf{X}\right) = \Sigma^{-\frac{1}{2}}\mathbb{E}(\mathbf{X}) = \Sigma^{-\frac{1}{2}} \cdot \mathbf{0} = \mathbf{0} \in \mathbb{R}^n.\checkmark$$

$$\begin{aligned}
\mathrm{VC}(\mathbf{Y}) &= \mathrm{VC}\left(\Sigma^{-\frac{1}{2}}\mathbf{X}\right) \\
&= \Sigma^{-\frac{1}{2}}\,\mathrm{VC}(\mathbf{X}) \underbrace{\left(\Sigma^{-\frac{1}{2}}\right)'}_{\text{symmetrisch}} \\
&= \Sigma^{-\frac{1}{2}}\,\mathrm{VC}(\mathbf{X})\Sigma^{-\frac{1}{2}} \\
&= \Sigma^{-\frac{1}{2}} \underbrace{\Sigma}_{\Sigma^{\frac{1}{2}}\Sigma^{\frac{1}{2}}} \Sigma^{-\frac{1}{2}} \\
&= \underbrace{\Sigma^{-\frac{1}{2}}\Sigma^{\frac{1}{2}}}_{\mathbf{I}_n} \underbrace{\Sigma^{\frac{1}{2}}\Sigma^{-\frac{1}{2}}}_{I_n} \\
&= \mathbf{I}_n.\checkmark
\end{aligned}$$

□

Bemerkung: Ist $\mathbf{Y}$ ein Zufallsvektor der Dimension n mit $\mathbb{E}(\mathbf{Y}) = 0$ und $\mathrm{VC}(\mathbf{Y}) = \mathbf{I}_n$. Sei weiters $\Sigma_{(n\times n)}$, sodass $\Sigma' = \Sigma$ und $\Sigma \geq 0$.
Für $\mathbf{X} = \Sigma^{\frac{1}{2}}\mathbf{Y}$ gilt dann:

$$\mathbb{E}(\mathbf{X}) = \mathbf{0}$$

und

$$\mathrm{VC}(\mathbf{X}) = \Sigma.$$

Nachrechnen.

$$\mathbb{E}(\mathbf{X}) = \mathbb{E}(\Sigma^{\frac{1}{2}}\mathbf{Y}) = \Sigma^{\frac{1}{2}}\mathbb{E}(\mathbf{Y}) = \Sigma^{\frac{1}{2}}\mathbf{0} = \mathbf{0} \in \mathbb{R}^n.\checkmark$$

$$\mathrm{VC}(\mathbf{X}) = \mathrm{VC}(\Sigma^{\frac{1}{2}}\mathbf{Y}) = \Sigma^{\frac{1}{2}}\,\mathrm{VC}(\mathbf{Y})\Sigma^{\frac{1}{2}} = \Sigma^{\frac{1}{2}}\mathbf{I}_n\Sigma^{\frac{1}{2}} = \Sigma^{\frac{1}{2}}\Sigma^{\frac{1}{2}} = \Sigma.\checkmark$$

□

3.2 Die Multivariate Normalverteilung

Definition 3.3 Sei $\mathbf{A}$ eine $n \times k$ Matrix und $\mathbf{b}$ ein n-dim. Vektor.
Seien weiters $Z_1, \ldots, Z_k$ iid. $N(0,1)$-verteilte reellwertige Zufallsvariablen.
Setze $\mathbf{Z} = (Z_1, \ldots, Z_k)'$ und $\mathbf{X} = \mathbf{A} \cdot \mathbf{Z} + \mathbf{b}$ ein n- dim. Zufallsvektor.
Die Verteilung von $\mathbf{X}$ nennt man die (multivariate) Normalverteilung mit Mittelwert $\mathbf{b}$ und VC Matrix $\mathbf{AA}'$.
Kurz:

$$\mathbf{X} \sim N(\mathbf{b}, \mathbf{AA}').$$

Nomination:

univariat	multivariat
σ^2	Σ
μ	$\boldsymbol{\mu}$ (Vektor)

Proposition 3.6 Für $\mathbf{X} \sim N(\boldsymbol{\mu}, \Sigma)$ gilt:

$$\mathbb{E}(\mathbf{X}) = \boldsymbol{\mu},$$

$$\mathrm{VC}(\mathbf{X}) = \Sigma.$$

Beweis. Sei $\boldsymbol{\mu} \in \mathbb{R}^n$ und $\Sigma_{(n\times n)}$.
Wähle $\mathbf{Z} = (Z_1, \ldots, Z_n)'$, sodass $Z_1, \ldots, Z_n$ i.i.d. $N(0,1)$.
Weiters sei $\Sigma^{\frac{1}{2}}$ die "Wurzel" von Σ.
Betrachte $\Sigma^{\frac{1}{2}}\mathbf{Z} + \boldsymbol{\mu}$.
Laut Definition ist

$$\Sigma^{\frac{1}{2}}\mathbf{Z} + \boldsymbol{\mu} \sim N\left(\boldsymbol{\mu}, \Sigma^{\frac{1}{2}}\Sigma^{\frac{1}{2}'}\right) \equiv N(\boldsymbol{\mu}, \Sigma).$$

Damit gilt:

$$\begin{aligned}\mathbb{E}(\mathbf{X}) &= \mathbb{E}\left(\Sigma^{\frac{1}{2}}\mathbf{Z} + \boldsymbol{\mu}\right)\\ &= \Sigma^{\frac{1}{2}}\underbrace{\mathbb{E}(Z)}_{\mathbf{0}\in\mathbb{R}^n} + \boldsymbol{\mu}\\ &= \boldsymbol{\mu}.\end{aligned}$$

$$\begin{aligned}\mathrm{VC}(\mathbf{X}) &= \mathrm{VC}\left(\Sigma^{\frac{1}{2}}\mathbf{Z} + \boldsymbol{\mu}\right)\\ &= \Sigma^{\frac{1}{2}}\,\mathrm{VC}(\mathbf{Z})\left(\Sigma^{\frac{1}{2}}\right)'\\ &= \Sigma^{\frac{1}{2}}\mathbf{I}_n\Sigma^{\frac{1}{2}}\\ &= \Sigma^{\frac{1}{2}}\Sigma^{\frac{1}{2}}\\ &= \Sigma.\end{aligned}$$

□

Proposition 3.7 **(Reproduktionseigenschaft)** Ist $\mathbf{X}$ ein n-dim. Zufallsvektor mit $\mathbf{X} \sim N(\boldsymbol{\mu}, \Sigma)$, ist $\mathbf{A}$ eine $m \times n$ Matrix und $\mathbf{b}$ ein m-dim. Vektor, dann ist

$$\mathbf{AX} + \mathbf{b} \sim N(\mathbf{A}\boldsymbol{\mu} + \mathbf{b}, \mathbf{A}\Sigma\mathbf{A}').$$

Beweis. Sei $\mathbf{Z} = (Z_1, \ldots, Z_n)' \in \mathbb{R}^n$ mit $Z_1, \ldots, Z_n$ iid. $N(0,1)$.
Dann gilt:

$$\Sigma^{\frac{1}{2}}\mathbf{Z} + \boldsymbol{\mu} \sim N(\boldsymbol{\mu}, \Sigma).$$

Also:

$$\Sigma^{\frac{1}{2}}\mathbf{Z}+\boldsymbol{\mu}\sim\mathbf{X}.$$

$$\begin{aligned}\Rightarrow \mathbf{AX}+\mathbf{b} &\sim \mathbf{A}\left(\Sigma^{\frac{1}{2}}\mathbf{Z}+\boldsymbol{\mu}\right)+\mathbf{b}\\ &= \mathbf{A}\Sigma^{\frac{1}{2}}\mathbf{Z}+\mathbf{A}\boldsymbol{\mu}+\mathbf{b}\\ &= \left(\mathbf{A}\Sigma^{\frac{1}{2}}\right)\mathbf{Z}+(\mathbf{A}\boldsymbol{\mu}+\mathbf{b}).\end{aligned}$$

Laut Definition ist die Verteilung von $\left(\mathbf{A}\Sigma^{\frac{1}{2}}\right)\mathbf{Z}+(\mathbf{A}\boldsymbol{\mu}+\mathbf{b})$ gegeben durch

$$\begin{aligned}N\left(\mathbf{A}\boldsymbol{\mu}+\mathbf{b},\left(\mathbf{A}\Sigma^{\frac{1}{2}}\right)\left(\mathbf{A}\Sigma^{\frac{1}{2}}\right)'\right) &= N\left(\mathbf{A}\boldsymbol{\mu}+\mathbf{b},\mathbf{A}\Sigma^{\frac{1}{2}}\left(\Sigma^{\frac{1}{2}}\right)'\mathbf{A}'\right)\\ &= N\left(\mathbf{A}\boldsymbol{\mu}+\mathbf{b},\mathbf{A}\Sigma\mathbf{A}'\right)\\ &\equiv N\left(\mathbf{A}\boldsymbol{\mu}+\mathbf{b},\mathbf{A}\Sigma\mathbf{A}'\right).\end{aligned}$$

Also:

$$\begin{aligned}\mathbf{AX}+\mathbf{b} &\sim \mathbf{A}\Sigma^{\frac{1}{2}}\mathbf{Z}+\mathbf{A}\boldsymbol{\mu}+\mathbf{b}\\ &\sim N\left(\mathbf{A}\boldsymbol{\mu}+\mathbf{b},\mathbf{A}\Sigma\mathbf{A}'\right).\end{aligned}$$

□

Bemerkung:
Die $N(\boldsymbol{\mu},\Sigma)$-Verteilung kann degeneriert sein, nämlich genau dann, wenn Σ nicht vollen Rang hat ($N(\boldsymbol{\mu},\mathbf{0})\equiv\boldsymbol{\mu}$ ist ein degenerierter Fall).

Siehe die folgenden Beispiele.

Beispiel 3.5 ($\mathbf{Z}\sim N(\mathbf{0},\mathbf{1})$, $\mathbf{A}=\mathbf{0}$, $\mathbf{b}\in\mathbb{R}$.) Laut Definition ist

$$\mathbf{X}=\mathbf{A}_{(1\times1)}\mathbf{Z}_{(1\times1)}+\mathbf{b}_{(1\times1)}\sim N\left(\mathbf{b},\mathbf{AA}'\right)=N(\mathbf{b},\mathbf{0}).$$

$$\mathbf{X}=\mathbf{AZ}+\mathbf{b}=\mathbf{0}\cdot\mathbf{Z}+\mathbf{b}=\mathbf{b}.$$

Also: $N(\mathbf{b},\mathbf{0})\equiv\mathbf{b}$.

Beispiel 3.6 Seien $X_1,\ldots,X_n$ reellwertige Zufallsvariablen, i.i.d. mit $X_i\sim N\left(\mu,\sigma^2\right)$, $\sigma^2>0$.
Setze $\mathbf{X}=(X_1,\ldots,X_n)'$.
Beachte:

$$\mathbb{E}(\mathbf{X})=\begin{pmatrix}\mu\\ \vdots\\ \mu\end{pmatrix}=\mu\iota.$$

$$\mathrm{VC}(\mathbf{X}) = \begin{pmatrix} \sigma^2 & & 0 \\ & \ddots & \\ 0 & & \sigma^2 \end{pmatrix} = \sigma^2 \mathbf{I}_n.$$

Sei $\mathbf{Z} = (Z_1, \dots, Z_n)$ mit Z_i iid. $N(0,1)$.
Betrachte

$$\begin{aligned} \sigma\mathbf{Z} + \mu\iota &= \sigma\mathbf{I}_n\mathbf{Z} + \mu\iota \\ &\sim N\left(\mu\iota, \sigma\mathbf{I}_n(\sigma\mathbf{I}_n)'\right) \text{ (aus Def.)} \\ &\equiv N\left(\mu\iota, \sigma^2\mathbf{I}_n\right). \end{aligned}$$

Weiters gilt: $(\sigma\mathbf{Z} + \mu\iota)_i = \sigma Z_i + \mu \sim N\left(\mu, \sigma^2\right) \sim X_i$, und $(\sigma\mathbf{Z} + \mu\iota)_i$, $1 \le i \le n$, sind unabhängig, und X_i, $1 \le i \le n$, sind unabhängig.
Damit ist $\mathbf{X} \sim \sigma\mathbf{Z} + \mu\iota \sim N\left(\mu\iota, \sigma^2\mathbf{I}_n\right)$.
Also ist $\mathbf{X} \sim N\left(\mu\iota, \sigma^2\mathbf{I}_n\right)$.
Zerlege $\mathbf{X}$ in $\mathbf{X} = \iota\bar{X} + \left(\mathbf{X} - \iota\bar{X}\right)$.
Wissen:

$$\begin{aligned} \iota\overline{X} &= \iota\frac{1}{n}\iota'\mathbf{X} \\ &\sim N\left(\iota\frac{1}{n}\iota'\mu\iota, \left(\iota\frac{1}{n}\iota'\right)\sigma^2\mathbf{I}_n\left(\iota\frac{1}{n}\iota'\right)'\right) \text{ (aufgrund der Reprod. Eigenschaft)} \\ &= N\left(\mu\iota\frac{1}{n}\iota'\iota, \sigma^2\iota\frac{1}{n}\iota'\iota\frac{1}{n}\iota'\right) \\ &= N\left(\mu\iota, \frac{\sigma^2}{n}\iota\iota'\right), \end{aligned}$$

eine Normalverteilung die auf eine Gerade $[\iota]$ konzentriert ist.
Analog sieht man: $\mathbf{X} - \bar{X}\iota \dots$ eine Normalverteilung in $[\iota]^\perp$.
Wissen:

$$\mathbb{E}\left(\mathbf{X} - \bar{X}\iota \in \mathbb{R}^n\right)$$

$$\mathrm{VC}\left(\mathbf{X} - \bar{X}\iota\right) = \sigma^2\left(\mathbf{I}_n - \iota\frac{1}{n}\iota'\right).$$

Nun ist $\mathbf{X} - \bar{X}\iota = \mathbf{X} - \iota\frac{1}{n}\iota'\mathbf{X} = \left(\mathbf{I}_n - \iota\frac{1}{n}\iota'\right)\mathbf{X}$, sodass

$$\mathbf{X} - \iota\bar{X} \sim N\left(\mathbf{0}, \sigma^2\left(\mathbf{I}_n - \iota\frac{1}{n}\iota'\right)\right) \dots \text{ eine Normalverteilung in} [\iota]^\perp.$$

Proposition 3.8 Betrachte einen $m+n$-dimensionalen Zufallsvektor $\begin{pmatrix}\mathbf{X}\\ \mathbf{Y}\end{pmatrix}$, wobei $\mathbf{X} \in \mathbb{R}^m$ und $\mathbf{Y} \in \mathbb{R}^n$, sodass

$$\begin{pmatrix}\mathbf{X}\\ \mathbf{Y}\end{pmatrix} \sim N(\boldsymbol{\mu}, \Sigma).$$

Ist $\mathrm{Cov}(\mathbf{X},\mathbf{Y}) = \mathbf{0}$, dann sind $\mathbf{X}$ und $\mathbf{Y}$ unabhängig.

Erinnerung: Zufallsvektoren $\mathbf{X}, \mathbf{Y}$ sind unabhängig, wenn für jede reellwertige Funktion $f(x)$ und für jede reellwertige Funktion $g(y)$ die Beziehung $\mathbb{E}(f(x)g(y)) = \mathbb{E}(f(x))\mathbb{E}(g(y))$ gilt.

Beweis. Partitioniere $\boldsymbol{\mu} = \mathbb{E}\begin{pmatrix}\mathbf{X}\\ \mathbf{Y}\end{pmatrix} = \begin{pmatrix}\boldsymbol{\mu}_X\\ \boldsymbol{\mu}_Y\end{pmatrix}$ mit $m+n$ Zeilen und $m+n$ Spalten, sodass $\mathbb{E}(\mathbf{X}) = \boldsymbol{\mu}_X$, $\mathbb{E}(\mathbf{Y}) = \boldsymbol{\mu}_Y$, sowie

$$\Sigma_{((m+n)\times(m+n))} = \begin{pmatrix}\Sigma_X & \Sigma_{XY}\\ \Sigma_{YX} & \Sigma_Y\end{pmatrix}$$

, sodass $\mathrm{VC}(\mathbf{X}) = \Sigma_X$, $\mathrm{VC}(\mathbf{Y}) = \Sigma_Y$ und $\mathrm{Cov}(\mathbf{X},\mathbf{Y}) = \Sigma_{XY}$.
Laut Voraussetzung ist $\Sigma_{XY} = \mathbf{0}$ $(m\times n)$ und $\Sigma_{YX} = \mathbf{0}$ $(n\times m)$. Seien $Z_1,\ldots,Z_{(m+n)}$ i.i.d $N(0,1)$ und $\mathbf{Z} = \begin{pmatrix}Z_1\\ \vdots\\ Z_{(m+n)}\end{pmatrix}$.
Partitioniere $\mathbf{Z} = \begin{pmatrix}\mathbf{Z}_X\\ \mathbf{Z}_Y\end{pmatrix}$ mit $m+n$ Zeilen.

$$\Sigma^{\frac{1}{2}} = \begin{pmatrix}\Sigma_X & \mathbf{0}\\ \mathbf{0} & \Sigma_Y\end{pmatrix}^{\frac{1}{2}} \overset{\ast}{=} \begin{pmatrix}\Sigma_X^{\frac{1}{2}} & \mathbf{0}\\ \mathbf{0} & \Sigma_Y^{\frac{1}{2}}\end{pmatrix}$$

Nachrechnen von $\ast$:

$$\begin{pmatrix}\Sigma_X^{\frac{1}{2}} & \mathbf{0}\\ \mathbf{0} & \Sigma_Y^{\frac{1}{2}}\end{pmatrix}\begin{pmatrix}\Sigma_X^{\frac{1}{2}} & \mathbf{0}\\ \mathbf{0} & \Sigma_Y^{\frac{1}{2}}\end{pmatrix} = \begin{pmatrix}\Sigma_X & \mathbf{0}\\ \mathbf{0} & \Sigma_Y\end{pmatrix}. \checkmark$$

Betrachte den Zufallsvektor

$$\begin{aligned}\Sigma^{\frac{1}{2}}\mathbf{Z} + \boldsymbol{\mu} &\sim N\left(\boldsymbol{\mu}, \Sigma^{\frac{1}{2}}\Sigma^{\frac{1}{2}}\right)\\ &= N(\boldsymbol{\mu}, \Sigma)\\ &\sim \begin{pmatrix}\mathbf{X}\\ \mathbf{Y}\end{pmatrix}.\end{aligned}$$

Beachte:

$$\begin{aligned}\Sigma^{\frac{1}{2}}\mathbf{Z} + \boldsymbol{\mu} &= \begin{pmatrix}\Sigma_X^{\frac{1}{2}} & \mathbf{0}\\ \mathbf{0} & \Sigma_Y^{\frac{1}{2}}\end{pmatrix}\begin{pmatrix}\mathbf{Z}_X\\ \mathbf{Z}_Y\end{pmatrix} + \begin{pmatrix}\boldsymbol{\mu}_X\\ \boldsymbol{\mu}_Y\end{pmatrix}\\ &= \begin{pmatrix}\Sigma_X^{\frac{1}{2}}\mathbf{Z}_X + \boldsymbol{\mu}_X\\ \Sigma_Y^{\frac{1}{2}}\mathbf{Z}_Y + \boldsymbol{\mu}_Y\end{pmatrix} \sim \begin{pmatrix}\mathbf{X}\\ \mathbf{Y}\end{pmatrix}.\end{aligned}$$

Nun sind $\mathbf{Z}_X, \mathbf{Z}_Y$ unabhängig.
$\Rightarrow \Sigma_X^{\frac{1}{2}}\mathbf{Z}_X + \boldsymbol{\mu}_X$, $\Sigma_Y^{\frac{1}{2}}\mathbf{Z}_Y + \boldsymbol{\mu}_Y$ unabhängig.

$\Rightarrow \mathbf{X}, \mathbf{Y}$ unabhängig.
Zu zeigen: $\mathbf{X}, \mathbf{Y}$ unabhängig.
Sei: $f(x)$ eine reellwertige Funtion von X und $g(y)$ eine reellwertige Funktion von Y. Zu zeigen: $\mathbb{E}(f(x)g(y)) = \mathbb{E}(f(x))E(g(y))$.

$$\begin{aligned}\mathbb{E}(f(x)g(y)) &= \mathbb{E}\left(f\left(\Sigma_X^{\frac{1}{2}}\mathbf{Z}_X + \boldsymbol{\mu}_X\right)g\left(\Sigma_Y^{\frac{1}{2}}\mathbf{Z}_Y + \boldsymbol{\mu}_Y\right)\right)\\ &= \mathbb{E}\left(f\left(\Sigma_X^{\frac{1}{2}}\mathbf{Z}_X + \boldsymbol{\mu}_X\right)\right)\mathbb{E}\left(g\left(\Sigma_Y^{\frac{1}{2}}\mathbf{Z}_Y + \boldsymbol{\mu}_Y\right)\right)\\ &= \mathbb{E}(f(x))\mathbb{E}(g(y)).\end{aligned}$$

□

Beispiel 3.7 $X_1, \ldots, X_n$ i.i.d. $N(\mu, \sigma^2), \mu \in \mathbb{R}, 0 < \sigma^2 < \infty$. Setze $\mathbf{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} \sim N(\boldsymbol{\mu\iota}, \sigma^2 \mathbf{I}_n)$.

Wissen:

$$\begin{aligned}\begin{pmatrix} \bar{X} \\ \mathbf{X} - \iota\bar{X} \end{pmatrix} &\ldots \quad \text{eine lineare Funktion von } \mathbf{X}\\ &= \begin{pmatrix} \frac{1}{n}\boldsymbol{\iota}' \\ \mathbf{I}_n - \boldsymbol{\iota}\frac{1}{n}\boldsymbol{\iota}' \end{pmatrix}\mathbf{X} \sim \quad \text{Normalverteilung.}\end{aligned}$$

Siehe Beispiel **??** und 3.4.
$\Rightarrow \bar{X}, \mathbf{X} - \iota\bar{X}$ sind unabhängig!
$\Rightarrow \bar{X}, \hat{\sigma}_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2 = \|\mathbf{X} - \iota\bar{X}\|^2$ sind unabhängig!

Erinnerung: Für $Z_1, \ldots, Z_n$ i.i.d. $N(0,1)$, ist

$$\sum_{i=1}^n Z_i^2 \sim \chi_k^2.$$

Lemma 3.1 Ist $\mathbf{P}_{(n \times n)}$ die Matrix einer Orthogonalprojektion, dann gilt: $\mathbf{P}^2 = \mathbf{P}$ und $\mathbf{P} = \mathbf{P}'$.

Beweis. Für $\mathbf{X}, \mathbf{Y} \in \mathbb{R}^n$ ist

$$\mathbf{X} = \mathbf{PX} + (\mathbf{I}_n - \mathbf{P})\mathbf{X} \quad \text{(orthogonale Zerlegung).}$$
$$\mathbf{Y} = \mathbf{PY} + (\mathbf{I}_n - \mathbf{P})\mathbf{Y} \quad \text{(orthogonale Zerlegung).}$$

$$\begin{aligned}\mathbf{X}'\mathbf{PY} &= (\mathbf{X})'\mathbf{PY}\\ &= (\mathbf{PX} + (\mathbf{I}_n - \mathbf{P})\mathbf{X})'\mathbf{PY}\\ &= (\mathbf{X}'\mathbf{P}' + \mathbf{X}'(\mathbf{I}_n - \mathbf{P})')\mathbf{PY}\\ &= \mathbf{X}'\mathbf{P}'\mathbf{PY} + \underbrace{\mathbf{X}'(\mathbf{I}_n - \mathbf{P})'\mathbf{PY}}_{\substack{=((\mathbf{I}_n - \mathbf{P})\mathbf{X})'\mathbf{PY} \\ =\mathbf{0}}}\\ &= \mathbf{X}'\mathbf{P}'\mathbf{PY}.\end{aligned}$$

$$\begin{aligned}\mathbf{X}'\mathbf{P}'\mathbf{Y} &= (\mathbf{PX})'\mathbf{Y}\\ &= (\mathbf{PX})'(\mathbf{PY} + (\mathbf{I}_n - \mathbf{P})\mathbf{Y})\\ &= (\mathbf{X}'\mathbf{P}'(\mathbf{PY} + (\mathbf{I}_n - \mathbf{P})\mathbf{Y})\\ &= \mathbf{X}'\mathbf{P}'\mathbf{PY} + \underbrace{\mathbf{X}'\mathbf{P}'(\mathbf{I}_n - \mathbf{P})\mathbf{Y}}_{\substack{=(\mathbf{PX})'(\mathbf{I}_n-\mathbf{P})\mathbf{Y}\\ =\mathbf{0}}}\\ &= \mathbf{X}'\mathbf{P}'\mathbf{PY}.\end{aligned}$$

$\Rightarrow \mathbf{X}'\mathbf{PY} = \mathbf{X}'\mathbf{P}'\mathbf{Y}$.

Da dies für beliebige $\mathbf{X}, \mathbf{Y} \in \mathbb{R}^n$ gilt, folgt $\mathbf{P} = \mathbf{P}'$.

Weiters ist $\mathbf{X}'\mathbf{PY} = \mathbf{X}'\mathbf{P}'\mathbf{PY}$

$$\Rightarrow \mathbf{P} = \mathbf{P}'\mathbf{P} \underset{\mathbf{P}'=\mathbf{P}}{\overset{\uparrow}{=}} \mathbf{PP} = \mathbf{P}^2.$$

□

Es gilt auch die Umkehrung: Ist $\mathbf{P}_{(n\times n)}$ mit $\mathbf{P}^2 = \mathbf{P}$ und $\mathbf{P} = \mathbf{P}'$, dann ist $\mathbf{P}$ die Matrix einer Orthogonalprojektion (ohne Beweis).

Korollar 3.1 Ist $\mathbf{P}_{(n\times n)}$ die Matrix einer Orthogonalprojektion auf einen k-dimensionalen Teilraum ($0 \le k \le n$), dann sind genau k Eigenwerte von $\mathbf{P}$ gleich 1 und der Rest ist 0.

Beweis. Spektralzerlegung $\mathbf{P} = \mathbf{U\Omega U}'$ ($\mathbf{U}$ orthogonal : $\mathbf{U}'\mathbf{U} = \mathbf{UU}' = \mathbf{I}_n$, $\quad \mathbf{\Omega} = \operatorname{diag}(\lambda_1, \dots, \lambda_n)$).
$\mathbf{P}^2 = \mathbf{P}$, d.h.

$$\begin{aligned} & \mathbf{U\Omega U}'\mathbf{U\Omega U}' = \mathbf{U\Omega U}'\\ \Leftrightarrow\quad & \mathbf{U\Omega}^2\mathbf{U}' = \mathbf{U\Omega U}'\\ \Leftrightarrow\quad & \mathbf{U\Omega}^2 = \mathbf{U\Omega}\\ \Leftrightarrow\quad & \mathbf{\Omega}^2 = \mathbf{\Omega}.\end{aligned}$$

D.h. für $i = 1, \dots, n$ ist $\lambda_i^2 = \lambda_i$, also $\lambda_i = 0$ oder $\lambda_i = 1$.
Weiters ist

$$\begin{aligned}\operatorname{rang}(\mathbf{P}) &= k\\ &= \operatorname{rang}(\mathbf{U\Omega U}')\\ &= \operatorname{rang}(\mathbf{\Omega})\\ &= \#\{i : 1 \le i \le n, \lambda_i = 1\}.\end{aligned}$$

Also sind genau k Eigenwerte gleich 1 und $n-k$ Eigenwerte sind 0. □

Proposition 3.9 Ist $\mathbf{P}_{(n\times n)}$ die Matrix einer Orthogonalprojektion auf einen k-dimensionalen Teilraum ($1 \le k \le n$), und ist $\mathbf{Z} \sim N(0,1)$. Dann gilt

$$\mathbf{Z}'\mathbf{PZ} \sim \chi_k^2$$

Beweis. $\mathbf{Z}'\mathbf{P}\mathbf{Z} = \mathbf{Z}'\mathbf{U}\mathbf{\Omega}\mathbf{U}'\mathbf{Z}$.
Nun ist $\mathbf{U}'\mathbf{Z} \sim N(\underbrace{\mathbf{U}'0}_{=0}, \underbrace{\mathbf{U}'\mathbf{I}_n U}_{\substack{=\mathbf{U}'\mathbf{U} \\ =\mathbf{I}_n}}) \equiv N(0, \mathbf{I}_n)$.

Insbesondere ist $\mathbf{Z} \sim \mathbf{U}'\mathbf{Z}$. Für $\mathbf{\Omega} = \operatorname{diag}(\underbrace{1,\dots,1}_{k}, \underbrace{0,\dots,0}_{n-k})$ ist damit

$$\begin{aligned}
\mathbf{Z}'\mathbf{P}\mathbf{Z} &= \mathbf{Z}'\mathbf{U}\mathbf{\Omega}\mathbf{U}'\mathbf{Z} \\
&= (\mathbf{U}'\mathbf{Z})'\mathbf{\Omega}(\mathbf{U}'\mathbf{Z}) \\
&\sim \mathbf{Z}'\mathbf{\Omega}\mathbf{Z} \\
&= \sum_{i=1}^{k} Z_i^2 \\
&\sim \chi_k^2.
\end{aligned}$$

□

Beispiel 3.8 Seien $X_1, \dots, X_n$ i.i.d. $N(\mu, \sigma^2)$, sodass $\mathbf{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} \sim N(\mu^2 . \sigma^2 \mathbf{I}_n)$.

$\mathbf{X} - \iota\bar{X} = \underbrace{\left(\mathbf{I}_n - \iota\frac{1}{n}\iota'\right)}_{:=\mathbf{P}}\mathbf{X}\dots$ Orthogonalprojektion von $\mathbf{X}$ auf $[\iota]^\perp$, ein $(n-1)$.dimensionaler Teilraum.
Prüfe Eigenschaften:

$$\begin{aligned}
\mathbf{P}^2 &= \left(\mathbf{I}_n - \iota\frac{1}{n}\iota'\right)\left(\mathbf{I}_n - \iota\frac{1}{n}\iota'\right) \\
&= \mathbf{I}_n - \iota\frac{1}{n}\iota' - \iota\frac{1}{n}\iota' + \iota\frac{1}{n}\iota'\iota\frac{1}{n}\iota' \\
&= \mathbf{I}_n - \iota\frac{1}{n}\iota' - \iota\frac{1}{n}\iota' + \iota\frac{1}{n}\iota' \\
&= \mathbf{I}_n - \iota\frac{1}{n}\iota' \\
&= \mathbf{P}. \checkmark
\end{aligned}$$

$\mathbf{P}' = \left(\mathbf{I}_n - \iota\frac{1}{n}\iota'\right)' = \mathbf{I}_n - \iota\frac{1}{n}\iota' = \mathbf{P}. \checkmark$
Beachte: $\dim([\iota]^\perp) = n-1$.

? Verteilung von $\|\mathbf{X} - \iota\bar{X}\|^2$?

(Beachte: $\hat{\sigma}_n^2 = \frac{1}{n-1}\|\mathbf{X} - \iota\bar{X}\|^2$).
Wissen: $\mathbf{X} - \iota\bar{X} \sim N\left(0, \sigma^2\left(\mathbf{I}_n - \iota\frac{1}{n}\iota'\right)\right)$.

Für $\mathbf{Z} \sim N(0, \mathbf{I}_n)$ ist

$$\underbrace{\sigma\left(\mathbf{I}_n - \iota\frac{1}{n}\iota'\right)}_{P}\mathbf{Z} \sim N(0, \underbrace{\sigma\mathbf{P}\mathbf{I}_n\mathbf{P}'\sigma}_{\substack{=\sigma^2\mathbf{PP}' \\ =\sigma^2\mathbf{PP} \\ =\sigma^2\mathbf{P}}})$$
$$= N(0, \sigma^2\mathbf{P})$$
$$\sim \mathbf{X} - \iota\bar{X}.$$

$$\|\mathbf{X} - \iota\bar{X}\|^2 \sim \|\sigma\underbrace{\left(\mathbf{I}_n - \iota\frac{1}{n}\iota'\right)}_{P}\mathbf{Z}\|^2$$
$$= (\sigma\mathbf{PZ})'(\sigma\mathbf{PZ})$$
$$= \sigma^2\mathbf{Z}'\mathbf{P}'\mathbf{PZ}$$
$$= \sigma^2\mathbf{Z}'\mathbf{PZ}$$
$$\underset{\substack{\uparrow \\ \text{Proposition}}}{\sim} \sigma^2\chi^2_{n-1}.$$

Proposition 3.10 Für $X_1, \ldots, X_n$ i.i.d. $N(\mu, \sigma^2)$ ist

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

und

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

Beweis. Erinnerung: Für Zufallsvariablen A, B, die unabhängig voneinander sind und so, dass $A \sim N(0,1)$ und $B \sim \chi^2_k$ ist.

$$\frac{A}{\sqrt{\frac{B}{n}}} \sim t_k$$

$$\frac{\sqrt{n}}{\hat{\sigma}_n}(\bar{X} - \mu) = \frac{\frac{\sqrt{n}}{\sigma}(\bar{X} - \mu)}{\frac{\hat{\sigma}_n}{\sigma}}$$

$$\frac{\sqrt{n}}{\sigma}(\bar{X} - \mu) \sim N(0, 1).$$

$$\begin{aligned}
\frac{\hat{\sigma}_n}{\sigma} &= \sqrt{\frac{\hat{\sigma}_n^2}{\sigma^2}} \\
&= \sqrt{\frac{1}{\sigma^2}\frac{1}{n-1}\underbrace{\sum_{i=1}^{n}(X_i-\bar{X})^2}_{\sim\sigma^2\chi^2_{n-1}}} \\
&\sim \sqrt{\frac{1}{\sigma^2}\frac{1}{n-1}\sigma^2\chi^2_{n-1}} \\
&\sim \sqrt{\frac{\chi^2_{n-1}}{n-1}}.
\end{aligned}$$

Zähler ist eine Funktion von $\bar{X}$ und Nenner ist eine Funktion von $\mathbf{X}-\iota\bar{X}$
$\Rightarrow$ Zähler und Nenner sind unabhängig. Also ist

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

□

3.3 Der Multivariate Zentrale Grenzwertsatz

Satz 3.1 Sind $\mathbf{X}_i$, $i \geq 1$, k-dimensionale Zufallsvektoren, die i.i.d. sind mit $\mathbb{E}(\mathbf{X}_i) = \boldsymbol{\mu}_{(n\times 1)}$, $\mathrm{VC}(\mathbf{X}_i) = \Sigma_{(n\times n)}$. Damit gilt für $\bar{\mathbf{X}}_n = \frac{1}{n}\sum_{i=1}^n \mathbf{X}_i$, dass

$$\sqrt{n}(\bar{\mathbf{X}}_n - \boldsymbol{\mu}) \xrightarrow{w} N(0,\Sigma)$$

3.4 Normalverteilungen - stetig und singulär (degeneriert)

Betrachte $\mathbf{X} \sim N(\boldsymbol{\mu},\Sigma)$, $\boldsymbol{\mu}\in\mathbb{R}^n$, $\Sigma_{(n\times n)}$. Es gilt $\mathrm{rang}(\Sigma)\in\{0,1,2,\ldots,n\}$.

Proposition 3.11 Ist $\mathrm{rang}(\Sigma) = n$, dann besitzt $\mathbf{X}$ eine Dichte, die gegeben ist durch

$$\Phi_{\mu,\Sigma}(\mathbf{x}) = (2\pi)^{-\frac{n}{2}}(\det\Sigma)^{-\frac{1}{2}}e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})}$$

für $\mathbf{x} = (x_1,\ldots,x_n)' \in \mathbb{R}^n$.

Beweis. Betrachte Spektralzerlegung (Eigenwertzerlegung) von $\Sigma = \mathbf{U\Omega U}'$, wobei $\boldsymbol{\Omega} = \mathrm{diag}(\lambda_1,\ldots,\lambda_n)$, wobei alle $\lambda_i > 0$ sind.
$\Sigma^{\frac{1}{2}} = \mathbf{U}\boldsymbol{\Omega}^{\frac{1}{2}}\mathbf{U}'$ mit $\boldsymbol{\Omega}^{\frac{1}{2}} = \mathrm{diag}\left(\sqrt{\lambda_1},\ldots,\sqrt{\lambda_n}\right)$,
$\Sigma^{-\frac{1}{2}} = \mathbf{U}\boldsymbol{\Omega}^{\frac{1}{2}}\mathbf{U}'$ mit $\boldsymbol{\Omega}^{-\frac{1}{2}} = \mathrm{diag}\left(\frac{1}{\sqrt{\lambda_1}},\ldots,\frac{1}{\sqrt{\lambda_n}}\right)$,
$\Sigma^{-1} = \mathbf{U}\boldsymbol{\Omega}^{\frac{1}{2}}\mathbf{U}'$ mit $\boldsymbol{\Omega}^{-1} = \mathrm{diag}\left(\frac{1}{\lambda_1},\ldots,\frac{1}{\lambda_n}\right)$.

Seien $Z_1,\dots,Z_n$ i.i.d. $N(0,1)$ und $\mathbf{Z} = (Z_1,\dots,Z_n)'$.
Dann ist

$$\begin{aligned}\Sigma^{\frac{1}{2}}\mathbf{Z} + \boldsymbol{\mu} &\sim N(\boldsymbol{\mu},\Sigma)\\ &\sim \mathbf{X}.\end{aligned}$$

Für $\mathbf{B} \in \mathbb{R}^n$ ist

$$\begin{aligned}\mathbb{P}(x \in \mathbf{B}) &= \mathbb{P}\left(\Sigma^{\frac{1}{2}}\mathbf{Z} + \boldsymbol{\mu} \in \mathbf{B}\right)\\ &= \int\limits_{\substack{\mathbf{Z}=(Z_1,\dots,Z_n)':\\ \Sigma^{\frac{1}{2}}\mathbf{Z}+\boldsymbol{\mu}\in\mathbf{B}}} \cdots \int \underbrace{\Phi(Z_1)\cdot\ldots\cdot\Phi(Z_n)}_{\substack{=\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{Z_1^2}{2}\right)\cdots\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{Z_n^2}{2}\right)\\ =(2\pi)^{-\frac{n}{2}}\exp\left(-\frac{Z_1^2}{2}-\frac{Z_2^2}{2}-\ldots-\frac{Z_n^2}{2}\right)\\ =(2\pi)^{-\frac{n}{2}}\exp\left(-\frac{1}{2}\sum_{i=1}^n Z_i^2\right)\\ =(2\pi)^{-\frac{n}{2}}\exp\left(-\frac{1}{2}\mathbf{Z}'\mathbf{Z}\right)}} \cdot dz_1\cdots dz_n\\ &= \int\limits_{\substack{\mathbf{Z}=(Z_1,\dots,Z_n)':\\ \Sigma^{\frac{1}{2}}Z+\boldsymbol{\mu}\in\mathbf{B}}} \cdots \int (2\pi)^{-\frac{n}{2}}\exp\left(-\frac{1}{2}Z'Z\right)\cdot dz_1\cdots dz_n = \circledast.\end{aligned}$$

$\mathbf{X} := \Sigma^{\frac{1}{2}}\mathbf{Z} + \boldsymbol{\mu}$.
$\mathbf{Z} = \Sigma^{-\frac{1}{2}}(\mathbf{X} - \boldsymbol{\mu})$.
$dz_1,\cdots,dz_n = |\det\Sigma|^{-\frac{1}{2}}$.

$$\begin{aligned}\circledast &= \int\limits_{\substack{\mathbf{X}=(X_1,\dots,X_n)':\\ x\in\mathbf{B}}} \cdots \int (2\pi)^{-\frac{n}{2}}\exp\left(-\frac{1}{2}(\mathbf{X}-\boldsymbol{\mu})'\Sigma^{-\frac{1}{2}}\Sigma^{-\frac{1}{2}}(\mathbf{X}-\boldsymbol{\mu})\right)|\det\Sigma|^{-\frac{1}{2}}\cdot dx_1\cdots dx_n\\ &= \int\limits_{\substack{\mathbf{X}=(X_1,\dots,X_n)':\\ x\in\mathbf{B}}} \cdots \int \underbrace{(2\pi)^{-\frac{n}{2}}|\det\Sigma|^{-\frac{1}{2}}\exp\left(-\frac{1}{2}(\mathbf{X}-\boldsymbol{\mu})'\Sigma^{-1}(\mathbf{X}-\boldsymbol{\mu})\right)}_{\substack{\text{... dies ist also die Dichte von}\mathbf{X}\\ =\Phi_{\boldsymbol{\mu},\Sigma}(x)}}\cdot dx_1\cdots dx_n\end{aligned}$$

□

Proposition 3.12 Ist $k = \text{rang}(\Sigma)$ so, dass $0 < k < n$, dann lässt sich $\mathbf{X}$ schreiben als

$$\mathbf{X} = \mathbf{B}\mathbf{W} + \boldsymbol{\mu},$$

wobei $\mathbf{B}$ eine $m \times k$ Matrix mit orthogonalen ($\mathbf{B}'\mathbf{B} = \mathbf{I}_k$) Spalten ist und wobei $\mathbf{W}$ ein k-dimensionaler Zufallsvektor ist mit

$$\mathbf{W} \sim N(0,\mathbf{D}),$$

mit $\mathbf{D} = \text{diag}(d_1,\dots,d_k) > 0$.

Beweis. Wieder $\Sigma = \mathbf{U}\boldsymbol{\Omega}\mathbf{U}'$. Laut Voraussetzung ($k = \text{rang}(\Sigma) < n$) kann man $\boldsymbol{\Omega}$ schreiben als $\boldsymbol{\Omega} = \text{diag}(\underbrace{\lambda_1,\dots,\lambda_k}_{k\text{ Komponente}}, \underbrace{0,\dots,0}_{n\text{ Komponente}})$.

Mit anderen Worten:

$$\mathbf{\Omega} = \begin{pmatrix} \lambda_1 & & & & & \mathbf{0} \\ & \ddots & & & & \\ & & \lambda_k & & & \\ & & & 0 & & \\ \mathbf{0} & & & & \ddots & \\ & & & & & 0 \end{pmatrix} = \begin{pmatrix} \mathbf{\Omega}_1 & \mathbf{0}_{k,n-k} \\ \mathbf{0}_{n-k,k} & \mathbf{0}_{n-k,n-k} \end{pmatrix},$$

wobei $\mathbf{\Omega}_1 = \operatorname{diag}(\lambda_1, \ldots, \lambda_k)$.

Partitioniere $\mathbf{U} = (\mathbf{U}_1, \mathbf{U}_2)$.

$$\begin{aligned} \mathbf{X} - \boldsymbol{\mu} &= \mathbf{U}\mathbf{U}'(\mathbf{X} - \boldsymbol{\mu}) \\ &= (\mathbf{U}_1\mathbf{U}_1' + \mathbf{U}_2\mathbf{U}_2')(\mathbf{X} - \boldsymbol{\mu}) \\ &= \mathbf{U}_1\mathbf{U}_1'(\mathbf{X} - \boldsymbol{\mu}) + \mathbf{U}_2\mathbf{U}_2'(\mathbf{X} - \boldsymbol{\mu}). \end{aligned}$$

$$\mathbf{U}_1'(\underbrace{\mathbf{X} - \boldsymbol{\mu}}_{\sim N(0,\mathbf{\Sigma})}) \sim N(0, \mathbf{U}_1'\mathbf{\Sigma}\mathbf{U}_1).$$

$$\begin{aligned} \mathbf{U}_1'\mathbf{\Sigma}\mathbf{U}_1 &= \mathbf{U}_1'\mathbf{U}\mathbf{\Omega}\mathbf{U}'\mathbf{U}_1 \\ &= \mathbf{U}_1'\begin{pmatrix}\mathbf{U}_1 & \mathbf{U}_2\end{pmatrix}\mathbf{\Omega}\begin{pmatrix}\mathbf{U}_1' \\ \mathbf{U}_2'\end{pmatrix}\mathbf{U}_1 \\ &= \begin{pmatrix}\mathbf{U}_1'\mathbf{U}_1 & \mathbf{U}_1'\mathbf{U}_2\end{pmatrix}\mathbf{\Omega}\begin{pmatrix}\mathbf{U}_1'\mathbf{U}_1 \\ \mathbf{U}_2'\mathbf{U}_1\end{pmatrix} \\ &= \begin{pmatrix}\mathbf{I}_k & \mathbf{0}_{k,n-k}\end{pmatrix}\begin{pmatrix} \mathbf{\Omega}_1 & \mathbf{0}_{k,n-k} \\ \mathbf{0}_{n-k,k} & \mathbf{0}_{n-k,n-k} \end{pmatrix}\begin{pmatrix}\mathbf{I}_k \\ \mathbf{0}_{k,n-k}\end{pmatrix} \\ &= \begin{pmatrix}\mathbf{I}_k & \mathbf{0}_{k,n-k}\end{pmatrix}\begin{pmatrix}\mathbf{\Omega}_1 \\ \mathbf{0}\end{pmatrix} \\ &= \mathbf{\Omega}_1. \end{aligned}$$

Also: $\mathbf{U}_1'(\mathbf{X} - \boldsymbol{\mu}) \sim N(0, \mathbf{\Omega}_1)$. Beachte $\mathbf{\Omega}_1 > 0$.

$$\begin{aligned} \mathbf{U}_2'\mathbf{\Sigma}\mathbf{U}_2 &= \mathbf{U}_2'\mathbf{U}\mathbf{\Omega}\mathbf{U}'\mathbf{U}_2 \\ &= \mathbf{U}_2'\begin{pmatrix}\mathbf{U}_1 & \mathbf{U}_2\end{pmatrix}\mathbf{\Omega}\begin{pmatrix}\mathbf{U}_1' \\ \mathbf{U}_2'\end{pmatrix}\mathbf{U}_2 \\ &= \begin{pmatrix}\mathbf{U}_2'\mathbf{U}_2 & \mathbf{U}_2'\mathbf{U}_2\end{pmatrix}\mathbf{\Omega}\begin{pmatrix}\mathbf{U}_1'\mathbf{U}_2 \\ \mathbf{U}_2'\mathbf{U}_2\end{pmatrix} \\ &= \begin{pmatrix}0 & \mathbf{I}_{n-k}\end{pmatrix}\begin{pmatrix} \mathbf{\Omega}_1 & \mathbf{0}_{k,n-k} \\ \mathbf{0}_{n-k,k} & \mathbf{0}_{n-k,n-k} \end{pmatrix}\begin{pmatrix}0 \\ \mathbf{I}_{n-k}\end{pmatrix} \\ &= \begin{pmatrix}0 & \mathbf{I}_{n-k}\end{pmatrix}\begin{pmatrix}0 \\ 0\end{pmatrix} \\ &= 0. \end{aligned}$$

Also:

$$\begin{aligned} \mathbf{U}_2'(\mathbf{X} - \boldsymbol{\mu}) &\sim N(0, 0) \\ &\sim \mathbf{0} \in \mathbb{R}^{n-k}. \end{aligned}$$

$$
\begin{aligned}
\mathbf{X}-\boldsymbol{\mu} &= \mathbf{U}_1\left(\mathbf{U}_1'(\mathbf{X}-\boldsymbol{\mu})\right)+\mathbf{U}_2\left(\mathbf{U}_2'(\mathbf{X}-\boldsymbol{\mu})\right)\\
\Leftrightarrow \mathbf{X} &= \mathbf{U}_1\underbrace{(\mathbf{U}_1'(\mathbf{X}-\boldsymbol{\mu}))}_{\sim N(0,\boldsymbol{\Omega}_1)}+\underbrace{\mathbf{U}_2\left(\mathbf{U}_2'(\mathbf{X}-\boldsymbol{\mu})\right)}_{0}+\boldsymbol{\mu}.
\end{aligned}
$$

Setze $\mathbf{B} := \mathbf{U}_1$, $\mathbf{W} = \mathbf{U}_1'(\mathbf{X}-\boldsymbol{\mu})$, und bin fertig. □

Bemerkung: Ist $\operatorname{rang}(\Sigma) = 0$, dann ist $\mathbf{X} \sim 0\mathbf{Z} + \boldsymbol{\mu}$ für $Z_1, \ldots, Z_n$ i.i.d. $N(0,1)$, denn hier ist $\Sigma = \mathbf{0}_{(n\times n)}$. Also $\mathbf{X} \sim 0\mathbf{Z} + \boldsymbol{\mu} \equiv \boldsymbol{\mu}$.

Bemerkung: Mit den beiden letzten Propositionen ist auch sichergestellt, dass die $N(\boldsymbol{\mu}, \Sigma)$-Verteilung, wie in Abschnitt I.2 angegeben, auch wohldefiniert ist.