Statistik II

Vorlesungsskript

Prof. Dr. Evgeny Spodarev

Ulm
2009

Inhaltsverzeichnis

Vorwort

Dieses Skript entstand aus dem Zyklus der Vorlesungen über Statistik, die ich in den Jahren 2006-2009 an der Universität Ulm gehalten habe. Dabei handelt es sich um die aufbauende Vorlesung Statistik II, die auf der Vorlesung *Statistik I* (im neuen Bachelor-Studienplan: *Stochastik I*) basiert.

Ich möchte gerne meinen Kollegen aus dem Institut für Stochastik, Herrn Prof. Volker Schmidt und Herrn Dipl.-Math. Malte Spiess, für ihre Unterstützung und anregenden Diskussionen während der Entstehung des Skriptes danken. Herr Marco Baur hat eine hervorragende Arbeit beim Tippen des Skriptes und bei der Erstellung zahlreicher Abbildungen, die den Text begleiten, geleistet. Dafür gilt ihm mein herzlicher Dank.

Ulm, den 14.04.2009
Evgeny Spodarev

1 Konfidenzintervalle

1.1 Einführung

Konfidenz- oder Vertrauensintervalle wurden bereits in Statistik I exemplarisch behandelt (vgl. Folgerung 3.3.2 und Bemerkung 3.3.4 des Vorlesungsskriptes Statistik I). In dieser vertiefenden Vorlesung werden wir eine formale Definition eines Konfidenzintervalles angeben um Vertrauensintervalle in größerer Tiefe studieren zu können. Dabei werden sowohl *Ein-* als auch *Zweistichprobenprobleme* behandelt.

Rufen wir uns die Annahmen eines parametrischen Modells in Erinnerung: es sei eine Stichprobe $(X_1, \ldots, X_n)$ von unabhängigen, identisch verteilten Zufallsvariablen mit $X_i \sim F_\theta$ gegeben, wobei F_θ eine Verteilungsfunktion aus einer parametrischen Familie von Verteilungen $\{F_\theta : \theta \in \Theta\}$, $\Theta \subset \mathbb{R}^m$ ist, dem m-dimensionalen Parameterraum, $m \geq 1$.

Die Punktschätzer von θ liefern jeweils einen Wert für den Parametervektor. Es wäre allerdings auch vorteilhaft, die Genauigkeit solcher Schätzansätze zu nennen, das heißt, einen Bereich anzugeben, in dem θ mit hoher Wahrscheinlichkeit $1 - \alpha$ liegt. Dabei heißt α *Irrtumswahrscheinlichkeit*; übliche Werte für α sind $\alpha = 0,01; 0,05; 0,1$. Die Wahrscheinlichkeit $1 - \alpha$, daß θ im vorgegebenen *Konfidenzintervall* liegt, heißt dann *Überdeckungswahrscheinlichkeit* oder *Konfidenzniveau* und soll dann entsprechend hoch ausfallen, z.B. $0,99; 0,95; 0,9$.

Definition 1.1.1. Es sei $1 - \alpha$ ein Konfidenzniveau und $\underline{\theta} : \mathbb{R}^n \to \overline{\mathbb{R}} = \mathbb{R} \cup \{\pm\infty\}$, $\overline{\theta} : \mathbb{R}^n \to \overline{\mathbb{R}}$ zwei Stichprobenfunktionen mit der Eigenschaft

$$\underline{\theta}(X_1, \ldots X_n) \leq \overline{\theta}(X_1, \ldots, X_n) \quad \forall (X_1, \ldots X_n) \in \mathbb{R}^n.$$

Falls

1. $P_\theta\left(\theta \in \left[\underline{\theta}(X_1, \ldots, X_n), \overline{\theta}(X_1, \ldots X_n)\right]\right) \geq 1 - \alpha \quad \forall \theta \in \Theta$

2. $\inf\limits_{\theta \in \Theta} P_\theta\left(\theta \in \left[\underline{\theta}(X_1, \ldots, X_n), \overline{\theta}(X_1, \ldots, X_n)\right]\right) = 1 - \alpha$

3. $\lim\limits_{n \to \infty} P_\theta\left(\theta \in \left[\underline{\theta}(X_1, \ldots, X_n), \overline{\theta}(X_1, \ldots, X_n)\right]\right) = 1 - \alpha \quad \forall \theta \in \Theta$

dann heißt $I = \left[\underline{\theta}(X_1, \ldots, X_n), \overline{\theta}(X_1, \ldots X_n)\right]$ ein

1. *Konfidenzintervall*

2. *minimales Konfidenzintervall*

3. *asymptotisches Konfidenzintervall*

zum Konfidenzniveau $1-\alpha$. Dabei heißt $l_\theta(X_1,\dots X_n) = \overline{\theta}(X_1,\dots X_n) - \underline{\theta}(X_1,\dots X_n)$ die *Länge* des Konfidenzintervalls. Es ist erwünscht, möglichst kleine Konfidenzintervalle (mit minimaler Länge) bei großem Konfidenzniveau für θ zu konstruieren.

Wie bereits bei den Beispielen im Statistik I-Skript ersichtlich ist, folgt die Konstruktion eines Konfidenzintervalls einem bestimmten Muster, das wir jetzt genauer studieren werden:

1. Finde eine Statistik $T(X_1,\dots,X_n,\theta)$, die
   - vom Parameter θ abhängt und
   - eine bekannte (Prüf-) Verteilung F besitzt (möglicherweise asymptotisch für $n\to\infty$).
2. Bestimme von der Verteilung F die Quantile $F^{-1}(\alpha_1)$ und $F^{-1}(1-\alpha_2)$ für Niveaus α_1 und $1-\alpha_2$, sodaß $\alpha_1+\alpha_2=\alpha$.
3. Löse (falls möglich) die Ungleichung $F^{-1}(\alpha_1) \le T(X_1,\dots,X_n,\theta) \le F^{-1}(1-\alpha_2)$ bzgl. θ auf. Das entsprechende Ergebnis $I = \left[T^{-1}(F^{-1}(\alpha)), T^{-1}(F^{-1}(1-\alpha_2))\right]$ (im Falle einer monoton in θ steigenden Statistik T) ist ein Konfidenzintervall für θ zum Niveau $1-\alpha$, denn es gilt

$$\begin{aligned}
\mathbb{P}_\theta(\theta\in I) &= \mathbb{P}_\theta\left(T_\theta^{-1}(F^{-1}(\alpha_1)) \le \theta \le T^{-1}(F^{-1}(1-\alpha_2))\right)\\
&= \mathbb{P}_\theta\left(F^{-1}(\alpha_1) \le T_\theta(X_1,\dots,X_n,\theta) \le F^{-1}(1-\alpha_2)\right)\\
&= F(F^{-1}(1-\alpha_2)) - F(F^{-1}(\alpha_1))\\
&= 1-\alpha_2-\alpha_1\\
&= 1-\alpha \text{ für alle } \theta\in\Theta.
\end{aligned}$$

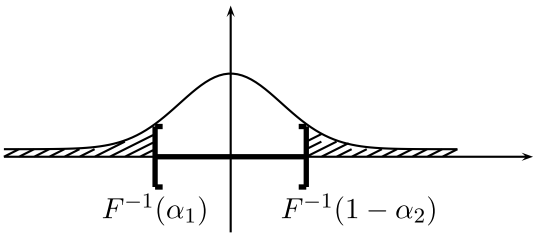
Für asymptotische Konfidenzintervalle soll überall noch $\lim\limits_{n\to\infty}$ geschrieben werden: $\lim\limits_{n\to\infty} \mathbb{P}_\theta(\theta\in I) = \dots = 1-\alpha$. Hierbei ist T_θ^{-1} die Inverse von $T(X_1,\dots,X_n,\theta)$ bezüglich θ. Grafisch kann dies wie folgt veranschaulicht werden:

Definition 1.1.2. 1. Falls $\alpha_1=\alpha_2=\alpha/2$, dann heißt das Konfidenzintervall $I = \left[T^{-1}(F^{-1}(\frac{\alpha}{2})), T^{-1}(F^{-1}(1-\frac{\alpha}{2}))\right]$ symmetrisch.

2. Falls $\alpha_1 = 0$ (bzw. $\underline{\theta}(X_1,\dots,X_n) = -\infty$), dann heißt das Konfidenzintervall $\left(-\infty, \overline{\theta}(X_1,\dots,X_n)\right]$ einseitig. Das selbe gilt für $\alpha_2=0$ (bzw. $\overline{\theta}(X_1,\dots,X_n) = +\infty$) und das Vertrauensintervall $[\underline{\theta}(X_1,\dots,X_n), +\infty)$.

In der Zukunft werden wir oft, ohne Beschränkung der Allgemeinheit, symmetrische Konfidenzintervalle konstruieren, obwohl man auch ein allgemeineres, nicht-symmetrisches Intervall leicht angeben kann.

Abbildung 1.1: asymptotisches Konfidenzintervall



Bemerkung 1.1.1. Man sieht leicht, daß der Algorithmus zur Konstruktion eines Vertrauensbereiches sich sehr dem eines statistischen Tests ähnelt. Im letzten Fall heißt $T(X_1, \ldots, X_n)$ *Teststatistik*. Im Allgemeinen kann man für jedes Konfidenzintervall einen entsprechenden statistischen Test angeben, aber nicht umgekehrt.

In Kapitel 2 werden wir einige Beispiele dieser Übertragung „Konfidenzintervall $\mapsto$ Test“sehen.

1.2 Ein-Stichproben-Probleme

In diesem Abschnitt werden wir einige Beispiele von Vertrauensbereichen für Parameter einiger bekannter Verteilungen nach dem oben genannten Schema konstruieren. Dabei werden wir immer mit einer Stichprobe $(X_1, \ldots, X_n)$ wie in Abschnitt 1.1 arbeiten.

1.2.1 Normalverteilung

Es seien $X_1, \ldots, X_n$ unabhängig, identisch verteilt, mit $X_i \sim N(\mu, \sigma^2)$.

Konfidenzintervalle für den Erwartungswert μ

- **bei bekannter Varianz** σ^2 Wenn wir annehmen, daß σ^2 bekannt ist, so ermöglicht uns der Satz 3.3.1, 4. (Vorlesungsskript Statistik I), ein exaktes Konfidenzintervall für μ zum Niveau $1-\alpha$ zu berechnen. Denn es gilt $\overline{X}_n \sim N\left(\mu, \sigma^2/n\right)$ und somit

$$T(X_1, \ldots, X_n, \mu) = \sqrt{n}\frac{\overline{X}_n - \mu}{\sigma} \sim N(0,1)$$

Es seien z_{α_1} und $z_{1-\alpha_2}$ Quantile der $N(0,1)$-Verteilung, $\alpha_1 + \alpha_2 = \alpha$ und $1-\alpha$ das vorgegebene Konfidenzniveau.

Dann gilt

$$\begin{aligned}1-\alpha &= \mathbb{P}\left(z_{\alpha_1} \le T(X_1,\dots,X_n,\mu) \le z_{1-\alpha_2}\right)\\ &= \mathbb{P}\left(z_{\alpha_1} \le \sqrt{n}\frac{\overline{X}_n-\mu}{\sigma} \le z_{1-\alpha_2}\right)\\ &\overset{(-z_{\alpha_1}=z_{1-\alpha_1})}{=} \mathbb{P}\left(\overline{X}_n - \frac{z_{1-\alpha_2}\sigma}{\sqrt{n}} \le \mu \le \overline{X}_n + \frac{z_{1-\alpha_1}\sigma}{\sqrt{n}}\right).\end{aligned}$$

Somit ist $\left[\underline{\theta}(X_1,\dots,X_n), \overline{\theta}(X_1,\dots,X_n)\right]$ mit $\underline{\theta}(X_1,\dots,X_n) = \overline{X}_n - z_{1-\alpha_2}\frac{\sigma}{\sqrt{n}}$ und $\overline{\theta}(X_1,\dots,X_n) = \overline{X}_n + z_{1-\alpha_1}\frac{\sigma}{\sqrt{n}}$ ein exaktes Konfidenzintervall für μ zum Niveau $1-\alpha$.

Es hat die Länge $l_\mu(X_1,\dots,X_n) = \frac{\sigma}{\sqrt{n}}\left(z_{1-\alpha_2}+z_{1-\alpha_1}\right)$. Es gilt $l_\mu(X_1,\dots,X_n) \to 0$, für $n\to\infty$ was bedeutet, daß bei wachsendem Informationsumfang ($n\to\infty$) die Präzision der Schätzung immer besser wird.

Im Symmetriefall ($\alpha_1=\alpha_2=\alpha/2$) müssen wir schreiben $\underline{\theta}(X_1,\dots,X_n) = \overline{X}_n - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}$, $\overline{\theta}(X_1,\dots,X_n) = \overline{X}_n + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}$ und $l_\mu(X_1,\dots,X_n) = \frac{2\sigma}{\sqrt{n}}z_{1-\alpha/2}$.

Daraus folgt, daß man bei vorgegebener Länge $\varepsilon>0$ die Anzahl der Beobachtungen n bestimmen kann, die dann notwendig sind um die vorgegebene Präzision zu erreichen:

$$\frac{2\sigma}{\sqrt{n}}z_{1-\alpha/2} \le \varepsilon \Longleftrightarrow n \ge \left(\frac{2\sigma z_{1-\alpha/2}}{\varepsilon}\right)^2 \tag{1.2.1}$$

Für $\alpha_1=0$ bzw. $\alpha_2=0$ kann man einseitige Intervalle $\left(-\infty, \overline{X}_n + z_{1-\alpha}\frac{\sigma}{\sqrt{n}}\right]$ und $\left[\overline{X}_n - z_{1-\alpha}\frac{\sigma}{\sqrt{n}}, +\infty\right)$ genauso angeben.

- **bei unbekannter Varianz** σ^2: siehe Bemerkung 3.3.4, Statistik I.

  Dort wurde das Konfidenzintervall $\left[\overline{X}_n - \frac{t_{n-1,1-\alpha/2}}{\sqrt{n}}S_n, \overline{X}_n + \frac{t_{n-1,1-\alpha/2}}{\sqrt{n}}S_n\right]$ für μ zum Konfidenzniveau $1-\alpha$ konstruiert, wobei $t_{n-1,1-\alpha/2}$ das $(1-\frac{\alpha}{2})$-Quantil der t_{n-1}- Verteilung ist.

  Wie man sieht, ist sie Länge des Konfidenzintervalls zufällig: $l_\mu(X_1,\dots X_n) = \frac{2S_n}{\sqrt{n}}t_{n-1,1-\alpha/2}$, somit macht es Sinn, mit erwarteter Länge

  $$\mathbb{E}\, l_\mu(X_1,\dots X_n) = \frac{2}{\sqrt{n}}\mathbb{E}\, S_n t_{n-1,1-\alpha/2}$$

  zu arbeiten, um zum Beispiel. die Frage nach der notwendigen Anzahl n von Beobachtungen bei vorgegebener Genauigkeit $\varepsilon>0$ (vergleiche Gleichung (1.2.1)) zu beantworten.

Konfidenzintervalle für die Varianz σ^2

- **bei bekanntem Erwartungswert μ:**

  Betrachten wir den Schätzer $\tilde{S}_n^2 = \frac{1}{n}\sum\limits_{i=1}^{n}(X_i-\mu)^2$ für σ^2. Aus Satz 3.3.5, 2. des Vorlesungsskriptes Statistik I folgt $\frac{n\tilde{S}_n^2}{\sigma^2} \sim \chi_n^2$. Wir setzen $T(X_1,\dots,X_n,\sigma^2) = \frac{n\tilde{S}_n^2}{\sigma^2}$ und bekommen

  $$\mathbb{P}\left(\chi^2_{n,\alpha_2} \le \frac{n\tilde{S_n}^2}{\sigma^2} \le \chi^2_{n,1-\alpha_1}\right) = \mathbb{P}\left(\frac{n\tilde{S}_n^2}{\chi^2_{n,1-\alpha_1}} \le \sigma^2 \le \frac{n\tilde{S}_n^2}{\chi^2_{n,\alpha_2}}\right) = 1-\alpha.$$

  Somit ist $\left[\frac{n\tilde{S}_n^2}{\chi^2_{n,1-\alpha_1}}, \frac{n\tilde{S}_n^2}{\chi^2_{n,\alpha_2}}\right]$ ein Konfidenzintervall für σ^2 zum Niveau $1-\alpha, \alpha=\alpha_1+\alpha_2$ mit der mittleren Länge $\mathbb{E}\, l_{\sigma^2} = n\sigma^2\left(\frac{1}{\chi^2_{n,\alpha_2}} - \frac{1}{\chi^2_{n,1-\alpha_1}}\right)$. Da die χ^2-Verteilung nicht symmetrisch ist, ist auch das Konfidenzintervall nicht symmetrisch.

- **bei unbekanntem Erwartungswert μ:**

  Ähnlich wie oben beschrieben folgt das Konfidenzintervall $\left[\frac{(n-1)S_n^2}{\chi^2_{n-1,1-\alpha_1}}, \frac{(n-1)S_n^2}{\chi^2_{n-1,\alpha_2}}\right]$ zum Niveau $1-\alpha, \alpha=\alpha_1+\alpha_2$ aus Satz 3.3.5, 1. des Vorlesungsskriptes Statistik I, weil $\frac{(n-1)S_n^2}{\sigma^2} \sim \chi^2_{n-1}$ für die Stichprobenvarianz $S_n^2 = \frac{1}{n-1}\sum\limits_{i=1}^{n}\left(X_i - \overline{X}_n\right)^2$. Die erwartete Länge ist $\mathbb{E}\, l_{\sigma^2} = (n-1)\sigma^2\left(\frac{1}{\chi^2_{n-1,\alpha_2}} - \frac{1}{\chi^2_{n-1,1-\alpha_1}}\right)$.

1.2.2 Konfidenzintervalle aus stochastischen Ungleichungen

Eine alternative Methode zur Gewinnung von Konfidenzintervallen besteht in der Anwendung stochastischer Ungleichungen. So kann man zum Beispiel bei einer Stichprobe $(X_1,\dots,X_n)$ von unabhängigen und identisch verteilten Zufallsvariablen mit $\mathbb{E}\, X_i = \mu$, $\operatorname{Var} X_i = \sigma^2 \in (0,\infty)$ die Ungleichung von Tschebyschew benutzen, um ein einfaches, aber grobes Konfidenzintervall für μ zu konstruieren:

$$\begin{aligned}
\mathbb{P}\left(|\overline{X}_n - \mu| > \varepsilon\right) \le \frac{\operatorname{Var}\overline{X}_n}{\varepsilon^2} &= \frac{\sigma^2}{n\varepsilon^2} = \alpha \\
&\Rightarrow \text{ für } \varepsilon = \frac{\sigma}{\sqrt{n\alpha}} \text{ gilt: } 1-\alpha = \mathbb{P}\left(|\overline{X}_n - \mu| \le \varepsilon\right) \\
&= \mathbb{P}\left(-\frac{\sigma}{\sqrt{n\alpha}} \le -\overline{X}_n + \mu \le \frac{\sigma}{\sqrt{n\alpha}}\right) \\
&= \mathbb{P}\left(\overline{X}_n - \frac{\sigma}{\sqrt{n\alpha}} \le \mu \le \overline{X}_n + \frac{\sigma}{\sqrt{n\alpha}}\right).
\end{aligned}$$

Das Konfidenzintervall $\left[\overline{X}_n - \frac{\sigma}{\sqrt{n\alpha}}, \overline{X}_n + \frac{\sigma}{\sqrt{n\alpha}}\right]$ für μ bei bekannter Varianz σ^2 ist verteilungsunabhängig, da keinerlei Annahmen über die Verteilung von X_i gemacht wurden.

Präzisere Konfidenzintervalle können bei der Verwendung folgender *Ungleichung von Hoeffding* konstruiert werden:

Satz 1.2.1 (Ungleichung von Hoeffding)**.** Es seien $Y_1, \ldots, Y_n$ unabhängige Zufallsvariablen mit $\mathbb{E}\, Y_i = 0, a_i \leq Y_i \leq b_i$ fast sicher, $i = 1, \ldots, n$. Für alle $\varepsilon > 0$ gilt

$$\mathbb{P}\left(\sum_{i=1}^n Y_i \geq \varepsilon\right) \leq \exp\left(-\frac{2\varepsilon^2}{\sum\limits_{i=1}^n (b_i - a_i)^2}\right)$$

(ohne Beweis).

Diese Ungleichung ist schärfer als die Tschebyschew-Ungleichung. Falls man spezielle Annahmen über die Verteilung von Y_i macht, kann man mit ihrer Hilfe auf gute Konfidenzintervalle unter Verwendung des Satzes 1.1 kommen.

Nehmen wir z.B. an, daß $X_1, \ldots, X_n$ unabhängige, identisch verteilte Zufallsvariablen sind, $X_i \sim \text{Bernoulli}(p)$, $p \in (0,1)$. Wir wollen ein Konfidenzintervall für p bestimmen.

Folgerung 1.2.1. Es seien $X_1, \ldots, X_n$ unabhängige Bernoulli(p)-verteilte Zufallsvariablen. Dann gilt $\mathbb{P}\left(|\overline{X}_n - p| > \varepsilon\right) \leq 2e^{-2n\varepsilon^2}, \quad \varepsilon > 0$.

Beweis. Es gilt

$$\overline{X}_n - p = \frac{1}{n} \sum_{i=1}^n \underbrace{(X_i - p)}_{Y_i},\ Y_i \in [-p, 1-p],$$

das heißt $a_i = -p$, $b_i = 1 - p$, $b_i - a_i = 1$, $i = 1, \ldots, n$, $\mathbb{E}\, Y_i = p - p = 0$. Dann gilt:

$$\begin{aligned}
\mathbb{P}_p\left(|\overline{X}_n - p| > \varepsilon\right) &\leq \mathbb{P}_p\left(\left|\sum_{i=1}^n Y_i\right| \geq \varepsilon n\right) \\
&= \mathbb{P}_p\left(\sum_{i=1}^n Y_i \geq \varepsilon n\right) + \mathbb{P}_p\left(\sum_{i=1}^n (-Y_i) \geq \varepsilon n\right) \\
&\overset{\text{(Satz 1.2.1)}}{\leq} 2e^{-\frac{2\varepsilon^2 n^2}{n}} = 2e^{-2\varepsilon^2 n},
\end{aligned}$$

wobei man den Satz 1.2.1 sowohl für die Folge $\{Y_i\}$ als auch $\{-Y_i\}$ anwendet. Damit ist die Behauptung bewiesen. □

Bemerkung 1.2.1. Die Form der Ungleichung von Hoeffding ähnelt sehr der von Dvoretzky-Kiefer-Wolfowitz, Satz 3.3.10 aus dem Vorlesungsskript Statistik I.

Nun fixieren wir $\alpha > 0$ und wählen $\varepsilon_n = \sqrt{\frac{1}{2n} \log \frac{2}{\alpha}}$. Durch Anwendung von Folgerung 1.2.1 mit diesem ε_n erhalten wir $\mathbb{P}_p\left(|\overline{X}_n - p| > \varepsilon_n\right) \leq \alpha$, somit $\mathbb{P}_p\left(|\overline{X}_n - p| \leq \varepsilon_n\right) \geq 1 - \alpha$ und darum ist $\left[\overline{X}_n - \sqrt{\frac{1}{2n} \log \frac{2}{\alpha}}, \overline{X}_n + \sqrt{\frac{1}{2n} \log \frac{2}{\alpha}}\right]$ ein Konfidenzintervall für p zum Niveau $1 - \alpha$.

1.2.3 Asymptotische Konfidenzintervalle

Die Philosophie der Konstruktion von asymptotischen Konfidenzintervallen ist relativ einfach: Wir erläutern sie am Beispiel eines asymptotisch normalverteilten Schätzers $\hat{\theta}$ für einen Parameter θ.

Sei $(X_1, \dots, X_n)$ eine Stichprobe von unabhängigen und identisch verteilten Zufallsvariablen, $X_i \sim F_\theta, \theta \in \Theta \subseteq \mathbb{R}$. Sei $\hat{\theta}_n = \hat{\theta}(X_1, \dots, X_n)$ ein Schätzer für θ, der asymptotisch normalverteilt ist. Dann gilt für erwartungstreue $\hat{\theta}_n$

$$\frac{\hat{\theta}_n - \theta}{\hat{\sigma}_n} \xrightarrow{d} Y \sim N(0,1),$$

wobei $\hat{\sigma}_n$ ein konsistenter Schätzer der Varianz von $\hat{\theta}_n$ ist.

$$\begin{aligned}
&\lim_{n\to\infty} \mathbb{P}_\theta \left(z_{\alpha/2} \le \frac{\hat{\theta}_n - \theta}{\hat{\sigma}_n} \le z_{1-\alpha/2} \right) \\
&= \lim_{n\to\infty} \mathbb{P}_\theta \left(\theta \in \left[\hat{\theta}_n - z_{1-\alpha/2}\hat{\sigma}_n,\, \hat{\theta}_n + z_{1-\alpha/2}\hat{\sigma}_n \right] \right) = 1 - \alpha.
\end{aligned}$$

Somit ist $\left[\hat{\theta}_n - z_{1-\alpha/2}\hat{\sigma}_n,\, \hat{\theta}_n + z_{1-\alpha/2}\hat{\sigma}_n \right]$ ein asymptotisches Konfidenzintervall für θ zum Niveau $1 - \alpha$.

Diese Vorgehensweise werden wir jetzt anhand von zwei Beispielen klar machen:

- **Bernoulli-Verteilung:**

  Seien $X_i \sim$ Bernoulli(p)-verteilt, $i = 1, \dots, n$. Dann gilt $\theta = p$, $\hat{\theta}_n = \hat{p}_n = \overline{X}_n$. $\mathbb{E}_p\, \hat{p}_n = p, \mathrm{Var}_p\, \hat{p}_n = \frac{p(1-p)}{n}$. Wir wählen $\hat{\sigma}^2 = \frac{1}{n}\,\hat{p}(1-\hat{p}_n) = \frac{\overline{X}_n}{n}(1-\overline{X}_n)$ als Plug-In-Schätzer für σ^2. Dann gilt nach dem zentralen Grenzwertsatz (Satz 7.2.1, WR) und dem Satz von Slutzky (Satz 6.4.2, 3. WR):

  $$\sqrt{n}\frac{\overline{X}_n - p}{\sqrt{\overline{X}_n(1-\overline{X}_n)}} \xrightarrow[n\to\infty]{d} Y \sim N(0,1),$$

  das heißt $p \in \left[\overline{X}_n - z_{1-\alpha/2}\sqrt{\frac{\overline{X}_n(1-\overline{X}_n)}{n}},\, \overline{X}_n + z_{1-\alpha/2}\sqrt{\frac{\overline{X}_n(1-\overline{X}_n)}{n}} \right]$ stellt ein asymptotisches Konfidenzintervall für p zum Niveau $1-\alpha$ dar. Da aber $p \in [0,1]$ sein soll, betrachtet man

  $$\underline{p}(X_1, \dots, X_n) = \max\left\{ 0,\, \overline{X}_n - z_{1-\alpha/2}\sqrt{\frac{\overline{X}_n(1-\overline{X}_n)}{n}} \right\}$$

  und

  $$\overline{p}(X_1, \dots, X_n) = \min\left\{ 1,\, \overline{X}_n + z_{1-\alpha/2}\sqrt{\frac{\overline{X}_n(1-\overline{X}_n)}{n}} \right\}.$$

Bemerkung 1.2.2. Ein anderes asymptotisches Konfidenzintervall für den Parameter p der Bernoulli-Verteilung bekommt man, wenn man die Aussage des zentralen Grenzwertsatzes $\lim_{n\to\infty} \mathbb{P}_p\left(-z_{1-\alpha/2} \le \sqrt{n}\frac{\overline{X}_n - p}{\sqrt{p(1-p)}} \le z_{1-\alpha/2}\right) = 1-\alpha$ nimmt und die quadratische Ungleichung dann bezüglich p auflöst.

Übung 1.2.1. Lösen Sie die Ungleichung auf!

- **Poissonverteilung**:

Es seien $X_i \sim \text{Poisson}(\lambda)$, $i = 1, \ldots, n$, dann gilt $\theta = \lambda$, $\hat{\theta}_n = \hat{\lambda} = \overline{X}_n$. Da $\mathbb{E}_\lambda X_i = \operatorname{Var}_\lambda X_i = \lambda$, kann man den zentralen Grenzwertsatz (Satz 7.2.1, WR) anwenden

$$\sqrt{n}\frac{\overline{X}_n - \lambda}{\sqrt{\lambda}} \xrightarrow[n\to\infty]{d} Y \sim N(0,1),$$

Da $\overline{X}_n$ stark konsistent für λ ist, gilt nach dem Satz von Slutsky (Satz 6.4.2, 4, WR)

$$\sqrt{n}\frac{\overline{X}_n - \lambda}{\sqrt{\overline{X}_n}} \xrightarrow[n\to\infty]{d} Y \sim N(0,1).$$

Daraus folgt ein asymptotisches Konfidenzintervall

$$\left[\overline{X}_n - z_{1-\alpha/2}\sqrt{\frac{\overline{X}_n}{n}},\ \overline{X}_n + z_{1-\alpha/2}\sqrt{\frac{\overline{X}_n}{n}}\right]$$

für den Parameter λ zum Konfidenzniveau $1-\alpha$.

Bemerkung 1.2.3.
1. Ähnlich wie in Bemerkung 1.2.2 angegeben, kann man durch Auflösen der quadratischen Ungleichung in

$$\lim_{n\to\infty} \mathbb{P}_\lambda\left(\sqrt{n}\frac{\overline{X}_n - \lambda}{\sqrt{\lambda}} \in [-z_{1-\alpha/2}, z_{1-\alpha/2}]\right) = 1-\alpha$$

bezüglich λ ein alternatives asymptotisches Konfidenzintervall für λ angeben.
Übung 1.2.2. Bitte führen Sie diese Berechnungen durch.

2. Da $\lambda > 0$ ist, kann man die untere Schranke diesbezüglich korrigieren:

$$\underline{\lambda}(X_1, \ldots, X_n) = \max\left\{0,\ \overline{X}_n - z_{1-\alpha/2}\sqrt{\frac{\overline{X}_n}{n}}\right\}$$

1.3 Zwei-Stichproben-Probleme

In diesem Abschnitt werden Charakteristiken bzw. Parameter von zwei unterschiedlichen Stichproben miteinander verglichen, indem man Konfidenzintervalle für einfache Funktionen dieser Parameter konstruiert.

Betrachten wir zwei Zufallsstichproben $Y_1 = (X_{11}, \ldots, X_{1n_1})$, $Y_2 = (X_{21}, \ldots, X_{2n_2})$ von Zufallsvariablen $X_{i1}, \ldots X_{in_i}$, $i = 1, 2$, die innerhalb der Stichprobe Y_i jeweils unabhängig und identisch verteilt sind, $X_{ij} \stackrel{d}{=} X_i$, $j = 1, \ldots n_i$, $i = 1, 2$ und die Prototyp-Zufallsvariable $X_i \sim F_{\theta_i}$, $\theta_i \in \Theta \subset \mathbb{R}^m$. Es wird im Allgemeinen nicht gefordert, daß Y_1 und Y_2 unabhängig sind. Falls sie voneinander abhängen, spricht man von *verbundenen Stichproben* Y_1 und Y_2. Betrachten wir eine Funktion $g : \mathbb{R}^{2m} \to \mathbb{R}$ von den Parametervektoren θ_1 und θ_2. In diesem Skript werden dabei meistens die Fälle $m = 1, 2$, $g(\theta_1, \theta_2) = \theta_{1j} - \theta_{2j}$, $g(\theta_1, \theta_2) = \frac{\theta_{1j}}{\theta_{2j}}$ untersucht, wobei $\theta_i = (\theta_{i1}, \ldots, \theta_{im})$, $i = 1, 2$.

Unsere Zielstellung wird sein, ein (möglicherweise asymptotisches) Konfidenzintervall für $g(\theta_1, \theta_2)$ mit Hilfe der Stichprobe (Y_1, Y_2) zu gewinnen.

Dabei wird die selbe Philosophie wie in Abschnitt 1.1 beschrieben verfolgt. Es wird eine Statistik $T(Y_1, Y_2, g(\theta_1, \theta_2))$ gesucht, die eine (möglicherweise asymptotische) Prüfverteilung F besitzt und von $g(\theta_1, \theta_2)$ explizit abhängt.

Durch das Auflösen der Ungleichung $F^{-1}_{\alpha_1} \leq T(Y_1, Y_2, g(\theta_1, \theta_2)) \leq F^{-1}_{1-\alpha_2}$ bzgl. $g(\theta_1, \theta_2)$ bekommt man dann ein (möglicherweise asymptotisches) Konfidenzintervall zum Niveau $1 - \alpha$, $\quad \alpha = \alpha_1 + \alpha_2$.

1.3.1 Normalverteilte Stichproben

Hier wird angenommen, daß $X_i \sim N(\mu_i, \sigma_i^2)$, $i = 1, 2$.

Konfidenzintervall für die Differenz $\mu_1 - \mu_2$ bei bekannten Varianzen σ_1^2 und σ_2^2 und unabhängigen Stichproben

Seien Y_1 und Y_2 voneinander unabhängig und σ_1^2, σ_2^2 bekannt. Wir betrachten die Parameterfunktion $g(\mu_1, \mu_2) = \mu_1 - \mu_2$. Es seien $\overline{X}_{in_i} = \frac{1}{n_i} \sum\limits_{j=1}^{n_i} X_{ij}$, $i = 1, 2$ die Stichprobenmittel der Stichproben Y_1 und Y_2. Es gilt $\overline{X}_{in_i} \sim N(\mu_i, \frac{\sigma_i^2}{n_i})$, $i = 1, 2$. Nach Satz 3.3.3, 4) aus dem Vorlesungsskript Statistik I sind $\overline{X}_{1n_1}$ und $\overline{X}_{2n_2}$ unabhängig. Dann ist wegen der Faltungsstabilität der Normalverteilung $X_{1n_1} - X_{2n_2} \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$. Nach dem Normieren erhält man die Statistik $T(Y_1, Y_2, \mu_1 - \mu_2) = \frac{\overline{X}_{1n_1} - \overline{X}_{2n_2}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0, 1)$.

Daraus bekommt man das Konfidenzintervall

$$\left[\overline{X}_{1n_1} - \overline{X}_{2n_2} - z_{1-\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}, \overline{X}_{1n_1} - \overline{X}_{2n_2} + z_{1-\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right]$$

für $\mu_1 - \mu_2$ zum Niveau $1 - \alpha$.

Konfidenzintervall für den Quotienten $\frac{\sigma_1^2}{\sigma_2^2}$ bei unbekannten Erwartungswerten μ_1 und μ_2 und unabhängigen Stichproben

Seien Y_1 und Y_2 voneinander unabhängig. Sei $g(\sigma_1, \sigma_2) = \frac{\sigma_1^2}{\sigma_2^2}$. Wir konstruieren die Statistik $T(Y_1, Y_2, \frac{\sigma_1^2}{\sigma_2^2})$ folgendermaßen: Seien $S_{in_i}^2 = \frac{1}{n_i-1} \sum\limits_{j=1}^{n_i} \left(X_{ij} - \overline{X}_{in_i} \right)$, $i = 1,2$ die Stichprobenvarianzen der Stichproben Y_1 und Y_2. Dann gilt $\frac{(n_i-1)S_{in_i}^2}{\sigma_i^2} \sim \chi_{n_i-1}^2$, $i = 1,2$ nach Satz 3.3.5 aus dem Vorlesungsskript Statistik I.

Da die $S_{in_i}^2$ voneinander unabhängig sind, gilt

$$T\left(Y_1, Y_2, \frac{\sigma_1^2}{\sigma_2^2}\right) = \frac{\frac{(n_2-1)S_{2n_2}^2}{(n_2-1)\sigma_2^2}}{\frac{(n_1-1)S_{1n_1}^2}{(n_1-1)\sigma_1^2}} = \frac{S_{2n_2}^2}{S_{1n_1}^2} \cdot \frac{\sigma_2^2}{\sigma_1^2} \sim F_{n_2-1,\, n_1-1}$$

nach der Definition der F - Verteilung. Daraus ergibt sich das Konfidenzintervall

$$\left[\frac{S_{1n_1}^2}{S_{2n_2}^2} F_{n_2-1,\, n_1-1,\, \alpha_1},\ \frac{S_{1n_1}^2}{S_{2n_2}^2} F_{n_2-1,\, n_1-1,\, 1-\alpha_2} \right]$$

für $\frac{\sigma_1^2}{\sigma_2^2}$ zum Niveau $1 - \alpha$.

Konfidenzintervall für die Differenz $\mu_1 - \mu_2$ der Erwartungswerte bei verbundenen Stichproben

Dieses Mal seien Y_1 und Y_2 verbunden, $X_1 - X_2 \sim N(\mu_1 - \mu_2,\, \sigma^2)$ für ein unbekanntes $\sigma^2 > 0$, $n_1 = n_2 = n$. Da $X_{ij}, j = 1, \ldots, n$ unabhängig und identisch verteilt sind, gilt $Z_j = X_{1j} - X_{2j} \sim N(\mu_1 - \mu_2, \sigma^2)$, $j = 1, \ldots, n$.

Unser Ziel ist es, ein Konfidenzintervall für $\mu_1 - \mu_2$ zu bekommen. Wenn wir die Stichprobe $(Z_1, \ldots, Z_n)$ betrachten, und Ergebnisse des Abschnittes 1.2.1, 2. anwenden, so erhalten wir sofort folgendes Konfidenzintervall:

$$\left[\overline{Z}_n - t_{n-1,1-\frac{\alpha}{2}} \frac{S_n}{\sqrt{n}},\ \overline{Z}_n + t_{n-1,1-\frac{\alpha}{2}} \frac{S_n}{\sqrt{n}} \right]$$

für $\mu_1 - \mu_2$ zum Niveau $1 - \frac{\alpha}{2}$, wobei $\overline{Z}_n = \frac{1}{n} \sum\limits_{j=1}^{n} Z_j = \frac{1}{n} \sum\limits_{j=1}^{n} (X_{1j} - X_{2j}) = \overline{X}_{1n} - \overline{X}_{2n}$, $S_n^2 = \frac{1}{n-1} \sum\limits_{j=1}^{n} \left(Z_j - \overline{Z}_n \right)^2 = \frac{1}{n-1} \sum\limits_{j=1}^{n} \left(X_{1j} - X_{2j} - \overline{X}_{1n} + \overline{X}_{2n} \right)^2$

1.3.2 Poissonverteilte Stichproben

Wir nehmen jetzt an, daß die Stichproben Y_1 und Y_2 unabhängig sind, und $X_i \sim \text{Poisson}(\lambda_i)$, $i = 1,2$. Konstruieren wir asymptotische Konfidenzintervalle für $g(\lambda_1, \lambda_2) = \lambda_1 - \lambda_2$ und $g(\lambda_1, \lambda_2) = \frac{n_2\lambda_2}{n_1\lambda_1 + n_2\lambda_2} = \frac{\lambda_2}{\rho\lambda_1 + \lambda_2}$, $\rho = \frac{n_1}{n_2} = \text{const}$, wobei $n_1, n_2 \to \infty$.

Asymptotisches Konfidenzintervall für $\lambda_1 - \lambda_2$

Um zu einer Statistik $T(Y_1, Y_2, \lambda_1 - \lambda_2)$ zu kommen, die asymptotisch (für $n_1, n_2 \to \infty$) $N(0,1)$-verteilt ist, verwenden wir den zentralen Grenzwertsatz von Ljapunow (vergleiche Satz 7.2.6, WR).

Lemma 1.3.1. Es gilt

$$\frac{\overline{X}_{1n_1} - \overline{X}_{2n_2} - \lambda_1 + \lambda_2}{\sqrt{\frac{\lambda_1}{n_1} + \frac{\lambda_2}{n_2}}} \xrightarrow[\substack{n_1\to\infty \\ n_2\to\infty}]{d} Y \sim N(0,1)$$

Beweis. Führen wir die Zufallsvariable

$$Z_{nk} = \begin{cases} \frac{X_{1k}-\lambda_1}{n_1\sqrt{\frac{\lambda_1}{n_1}+\frac{\lambda_2}{n_2}}}, & k = 1, \dots, n_2, \\ -\frac{X_{2k-n_1}-\lambda_2}{n_2\sqrt{\frac{\lambda_1}{n_1}+\frac{\lambda_2}{n_2}}}, & k = n_2+1, \dots, n_1+n_2 \end{cases}$$

ein, wobei $n = n_1 + n_2$. Es gilt: $\mathbb{E}\, Z_{nk} = 0$ für alle $k = 1, \dots, n$, und

$$0 < \sigma_{nk}^2 = \operatorname{Var} Z_{nk} = \begin{cases} \frac{\operatorname{Var} X_{1k}}{n_1^2\left(\frac{\lambda_1}{n_1}+\frac{\lambda_2}{n_2}\right)} = \frac{\lambda_1}{n_1^2\left(\frac{\lambda_1}{n_1}+\frac{\lambda_2}{n_2}\right)}, & k = 1, \dots, n_1, \\ \frac{\lambda_2}{n_2^2\left(\frac{\lambda_1}{n_1}+\frac{\lambda_2}{n_2}\right)}, & k = n_1+1, \dots, n, \end{cases}$$

somit

$$\sum_{k=1}^{n} \sigma_{nk}^2 = \left(\frac{\lambda_1}{n_1^2} n_1 + \frac{\lambda_2}{n_2^2} n_2\right) \frac{1}{\frac{\lambda_1}{n_1} + \frac{\lambda_2}{n_2}} = 1.$$

Außerdem gilt für $\delta > 0$ und $n_1, n_2 \to \infty$:

$$\lim_{n\to\infty} \sum_{k=1}^{n} \mathbb{E}\left(|Z_{nk}|\right)^{2+\delta} = \lim_{n_1,n_2\to\infty} \left(\frac{\mathbb{E}\left(|X_{11} - \lambda_1|^{2+\delta}\right)}{n_1^{1+\delta}\left(\frac{\lambda_1}{n_1} + \frac{\lambda_2}{n_2}\right)^{(2+\delta)/2}} + \frac{\mathbb{E}\left(|X_{21-\lambda_2}|\right)^{2+\delta}}{n_2^{1+\delta}\left(\frac{\lambda_1}{n_1} + \frac{\lambda_2}{n_2}\right)^{(2+\delta)/2}} i \right) = 0$$

Somit ist die Ljapunow-Bedingung erfüllt und nach Satz 7.2.6 (WR) gilt

$$\sum_{k=1}^{n} Z_{nk} \xrightarrow[\substack{n_1\to\infty \\ n_2\to\infty}]{d} Y \sim N(0,1).$$

Es gilt aber auch $\sum_{n=1}^{n} Z_{nk} = \frac{\overline{X}_{1n_1} - \overline{X}_{2n_2} - \lambda_1 + \lambda_2}{\sqrt{\frac{\lambda_1}{n_1} + \frac{\lambda_2}{n_2}}}$, somit ist das Lemma bewiesen. □

Da $X_{in_i} \xrightarrow{f.s.} \lambda_i$, $i = 1, 2$ nach dem starken Gesetz der großen Zahlen, gilt mit Hilfe des Satzes von Slutsky

$$T(Y_1, Y_2, \lambda_1 - \lambda_2) = \frac{\overline{X}_{1n_1} - \overline{X}_{2n_2} - \lambda_1 + \lambda_2}{\sqrt{\overline{X}_{1n_1}/n_1 + \overline{X}_{n_2}/n_2}} \xrightarrow[n_1, n_2 \to \infty]{d} Y \sim N(0,1)$$

Daraus läßt sich sofort das asymptotische Konfidenzintervall für $\lambda_1 - \lambda_2$ zum Niveau $1 - \alpha$ ableiten:

$$\left[\overline{X}_{1n_1} - \overline{X}_{2n_2} - z_{1-\alpha/2}\sqrt{\frac{\overline{X}_{1n_1}}{n_1} + \frac{\overline{X}_{2n_2}}{n_2}},\, \overline{X}_{1n_1} - \overline{X}_{2n_2} + z_{1-\alpha/2}\sqrt{\frac{\overline{X}_{1n_1}}{n_1} + \frac{\overline{X}_{2n_2}}{n_2}}\right]$$

Asymptotisches Konfidenzintervall für $\frac{n_2\lambda_2}{n_1\lambda_1+n_2\lambda_2}$

Es sei $n_1/n_2 = \beta = \text{const}$ und $g(\lambda_1, \lambda_2) = \frac{n_2\lambda_2}{n_1\lambda_1+n_2\lambda_2} = \frac{\lambda_2}{\beta\lambda_1+\lambda_2} \overset{\text{Def.}}{=} p$. Es wird ein asymptotisches Konfidenzintervall für p gesucht. Wir führen die Statistik

$$T(Y_1, Y_2, p) = \frac{S_{2n_2} - p(S_{1n_1} + S_{2n_2})}{\sqrt{\hat{p}(1-\hat{p})(S_{1n_1} + S_{2n_2})}}$$

ein, wobei $S_{in_i} = \sum\limits_{j=1}^{n_i} X_{ij}$, $i = 1, 2$ und

$$\hat{p} = \frac{S_{2n_2}}{S_{1n_1} + S_{2n_2}} = \frac{n_2\overline{X}_{2n_2}}{n_1\overline{X}_{1n_1} + n_2\overline{X}_{2n_2}} \xrightarrow[n_1, n_2 \to \infty]{f.s.} p$$

ein konsistenter Schätzer für p (wegen des starken Gesetzes der großen Zahlen) ist. Falls wir zeigen können, daß $T(Y_1, Y_2, p) \xrightarrow[n_1, n_2 \to \infty]{d} Y \sim N(0,1)$, so wird daraus folgendes Konfidenzintervall ableitbar: Aus

$$\lim_{\substack{n_1\to\infty \\ n_2\to\infty}} \mathbb{P}\left(-z_{1-\alpha/2} \leq \frac{\frac{S_{2n_2}}{S_{1n_1}+S_{2n_2}} - p}{\sqrt{S_{1n_1}\cdot S_{2n_2}}} \cdot (S_{1n_1} + S_{2n_2})^{3/2} \leq z_{1-\alpha/2}\right) = 1 - \alpha$$

folgt, daß

$$\left[\underline{\theta}(Y_1, Y_2),\, \overline{\theta}(Y_1, Y_2)\right]$$

mit

$$\underline{\theta}(\lambda_1, \lambda_2) = \frac{S_{2n_2}}{S_{1n_1} + S_{2n_2}} - z_{1-\alpha/2} \cdot \sqrt{\frac{S_{1_{n_1}} \cdot S_{2n_2}}{\left(S_{1_{n_1}} + S_{2n_2}\right)^3}}$$

$$\overline{\theta}(\lambda_1, \lambda_2) = \frac{S_{2n_2}}{S_{1n_1} + S_{2n_2}} + z_{1-\alpha/2} \cdot \sqrt{\frac{S_{1_{n_1}} \cdot S_{2n_2}}{\left(S_{1_{n_1}} + S_{2n_2}\right)^3}}$$

ein asymptotisches Konfidenzintervall für p zum Niveau $1-\alpha$ ist.

Da $0 < p < 1$ sein soll, können die Schranken des Intervalls diesbezüglich korrigiert werden:

$$\begin{aligned}\underline{\theta}^*(Y_1, Y_2) &= \max\{0, \underline{\theta}(Y_1, Y_2)\},\\ \overline{\theta}^*(Y_1, Y_2) &= \min\{1, \overline{\theta}(Y_1, Y_2)\}.\end{aligned}$$

Nun soll die asymptotische Normalverteiltheit von $T(Y_1, Y_2, p)$ gezeigt werden. Sie folgt aus dem Satz von Slutsky und folgendem Lemma:

Lemma 1.3.2. Es gilt:

$$\frac{S_{2n_2} - p(S_{1n_1} + S_{2n_2})}{\sqrt{p(1-p)(S_{1n_1} + S_{2n_2})}} \xrightarrow[n_1\to\infty]{d} Y \sim N(0,1)$$

Beweis. Um die Aussage des Lemmas zu zeigen, verwenden wir einen zentralen Grenzwertsatz für Summen von Zufallsvariablen in zufälliger Anzahl (vgl. Satz 7.2.2 (WR)). Führen wir die Folge $N_n = S_{1n_1} + S_{2n_2}$ von nichtnegativen Zufallsvariablen ein. Die Summe ist monoton wachsend. Gleichzeitig setzen wir $a_{n_2} = n_1\lambda_1 + n_2\lambda_2$. Offensichtlich gilt

$$\begin{aligned}\frac{N_n}{a_{n_2}} &= \frac{S_{1n_1}}{n_1\lambda_1 + n_2\lambda_2} + \frac{S_{2n_2}}{n_1\lambda_1 + n_2\lambda_2}\\ &= \frac{\overline{X}_{1n_1}}{\lambda_1 + \beta^{-1}\lambda_2} + \frac{\overline{X}_{2n_2}}{\beta\lambda_1 + \lambda_2}\\ &\xrightarrow[n_1,n_2\to\infty]{f.s.} \frac{\lambda_1}{\lambda_1 + \beta^{-1}\lambda_2} + \frac{\lambda_2}{\beta\lambda_1 + \lambda_2}\\ &= \frac{\beta\lambda_1}{\beta\lambda_1 + \lambda_2} + \frac{\lambda_2}{\beta\lambda_1 + \lambda_2} = 1\end{aligned}$$

Außerdem gilt:

$$\begin{aligned}\mathbb{P}\left(S_{2n_2} = k \mid N_n = m\right) &= \frac{\mathbb{P}\left(S_{2n_2} = k,\, S_{1n_1} + S_{2n_2} = m\right)}{\mathbb{P}\left(S_{1n_1} + S_{2n_2} = m\right)}\\ &= \frac{\mathbb{P}\left(S_{2n_2} = k,\, S_{1n_1} = m-k\right)}{\mathbb{P}\left(S_{1n_1} + S_{2_{n-2}} = m\right)}\\ &= \frac{e^{-n_2\lambda_2}\frac{(\lambda_2 n_2)^k}{k!}\cdot e^{-n_1\lambda_1}\frac{(n_1\lambda_1)^{m-k}}{(m-k)!}}{e^{-n_1\lambda_1 - n_2\lambda_2}\frac{(n_1\lambda_1 + n_2\lambda_2)^m}{m!}}\\ &= \frac{m!}{(m-k)!k!}\left(\frac{n_2\lambda_2}{n_1\lambda_1 + n_2\lambda_2}\right)^m\left(\frac{n_1\lambda_1}{n_1\lambda_1 + n_2\lambda_2}\right)^{m-k}\\ &= \binom{m}{k}p^k(1-p)^{m-k}\end{aligned}$$

was bedeutet, daß $S_{2n_2} \mid N_n = m \sim \mathrm{Bin}(m,p)$. Dann gilt $\frac{S_{2n_2}-mp}{\sqrt{mp(1-p)}} \mid N_n = m \stackrel{d}{=} \frac{S_m-mp}{\sqrt{mp(1-p)}}$, wobei $S_m = \sum\limits_{i=1}^{m} Z_i$ eine Summe von unabhängigen, identisch verteilten Zufallsvariablen $Z_i \sim \mathrm{Bernoulli}(p)$ ist. Nach Satz 7.2.2 (WR) gilt dann

$$\frac{S_{Nn_N} - N_n p}{\sqrt{N_n p(1-p)}} \xrightarrow{d} Y \sim N(0,1) \Longleftrightarrow \frac{S_{2n_2} - N_n p}{\sqrt{N_n p(1-p)}} \xrightarrow{d} Y \sim N(0,1).$$

□

2 Tests statistischer Hypothesen

In der Vorlesung Statistik I haben wir schon Beispiele von statistischen Tests kennengelernt, wie etwa den Kolmogorow-Smirnow-Test (vergleiche Bemerkung 3.3.38, 3), Skript Statistik I). Jetzt sollen statistische Signifikanztests formal eingeführt und ihre Eigenschaften untersucht werden.

2.1 Allgemeine Philosophie des Testens

Es sei eine Zufallsstichprobe $(X_1, \ldots, X_n)$ von unabhängigen, identisch verteilten Zufallsvariablen X_i gegeben, mit Verteilungsfunktion $F \in \Lambda$, wobei Λ eine Klasse von Verteilungsfunktionen ist. Es sei $(x_1, \ldots, x_n)$ eine konkrete Stichprobe, die als Realisierung von $(X_1, \ldots, X_n)$ interpretiert wird. In der Theorie des statistischen Testens werden Hypothesen über die Beschaffenheit der (unbekannten) Verteilungsfunktion F gestellt und geprüft. Dabei unterscheidet man

$$\begin{array}{ccc} & \boxed{\text{Statistische Tests}} & \\ \swarrow & & \searrow \\ \boxed{\text{parametrische Tests}} & & \boxed{\text{nichtparametrische Tests}} \\ \text{falls } \Lambda = \{F_\theta, \theta \in \Theta\}, & & \\ \text{wobei } \Theta \subseteq \mathbb{R}^m \text{ ist.} & & \text{sonst.} \end{array}$$

Bei parametrischen Tests prüft man, ob der Parameter θ bestimmte Werte annimmt (zum Beispiel $\theta = 0$). Bekannte Beispiele von nichtparametrischen Tests sind Anpassungstests, bei denen man prüft, ob die Verteilungsfunktion F gleich einer vorgegebenen Funktion F_0 ist.

Formalisieren wir zunächst den Begriff *Hypothese*. Die Menge Λ von zulässigen Verteilungsfunktionen F wird in zwei disjunkte Teilmengen Λ_0 und Λ_1 zerlegt, $\Lambda_0 \cup \Lambda_1 = \Lambda$. Die Aussage

„Man testet die *Haupthypothese* $H_0 : F \in \Lambda_0$ gegen die *Alternative* $H_1 : F \in \Lambda_1$,“

bedeutet, daß man an Hand der konkreten Stichprobe $(x_1, \ldots, x_n)$ versucht, eine Entscheidung zu fällen, ob die Verteilungsfunktion der Zufallsvariable X_i zu Λ_0 oder zu Λ_1 gehört. Dies passiert auf Grund einer statistischen *Entscheidungsregel*

$$\varphi : \mathbb{R}^n \to [0, 1],$$

die eine Statistik mit folgender Interpretation ist:

Der Stichprobenraum $\mathbb{R}^n$ wird in drei disjunkte Bereiche K_0, K_{01} und K_1 unterteilt, sodaß $\mathbb{R}^n = K_0 \cup K_{01} \cup K_1$, wobei

$$\begin{aligned}
K_0 &= \varphi^{-1}(\{0\}) &&= \{x \in \mathbb{R}^n : \varphi(x) = 0\}\,, \\
K_1 &= \varphi^{-1}(\{1\}) &&= \{x \in \mathbb{R}^n : \varphi(x) = 1\}\,, \\
K_{01} &= \varphi^{-1}((0,1)) &&= \{x \in \mathbb{R}^n : 0 < \varphi(x) < 1\}\,.
\end{aligned}$$

Dementsprechend wird $H_0 : F \in \Lambda_0$

- verworfen, falls $\varphi(x) = 1$, also $x \in K_1$,
- nicht verworfen, falls $\varphi(x) = 0$, also $x \in K_0$;
- falls $\varphi(x) \in (0,1)$, also $x \in K_{01}$, wird $\varphi(x)$ als Bernoulli-Wahrscheinlichkeit interpretiert, und es wird eine Zufallsvariable $Y \sim \text{Bernoulli}(\varphi(x))$ generiert, für die gilt:

$$Y = \begin{cases} 1 & \Longrightarrow H_0 \text{ wird verworfen} \\ 0 & \Longrightarrow H_0 \text{ wird nicht verworfen} \end{cases}$$

Falls $K_{01} \neq \emptyset$, wird eine solche Entscheidungsregel *randomisiert* genannt. Bei $K_{01} = \emptyset$, also $\mathbb{R}^n = K_0 \cup K_1$ spricht man dagegen von *nicht-randomisierten* Tests. Dabei heißt K_0 bzw. K_1 *Annahmebereich* bzw. *Ablehnungsbereich* (*kritischer Bereich*) von H_0. K_{01} heißt *Randomisierungsbereich.*

Bemerkung 2.1.1. 1. Man sagt absichtlich „H_0 *wird nicht verworfen*“, statt „H_0 *wird akzeptiert*“, weil die schließende Statistik generell keine positiven, sondern nur negative Entscheidungen treffen kann. Dies ist generell ein philosophisches Problem der Falsifizierbarkeit von Hypothesen oder wissenschaftlichen Theorien, von denen aber keiner behaupten kann, daß sie der Wahrheit entsprechen (vergleiche die *wissenschaftliche Erkenntnistheorie vom Karl Popper (1902-1994)*).

2. Die randomisierten Tests sind hauptsächlich von theoretischem Interesse (vergleiche Abschnitt 2.3). In der Praxis werden meistens nichtrandomisierte Regeln verwendet, bei denen man aus der Stichprobe $(x_1, \ldots, x_n)$ allein die Entscheidung über H_0 treffen kann. Hier gilt $\varphi(x) = \mathbb{I}_{K_1}, x = (x_1, \ldots, x_n) \in \mathbb{R}^n$.

In diesem und in folgendem Abschnitt betrachten wir ausschließlich nichtrandomisierte Tests, um in Abschnitt 2.3 zu der allgemeinen Situation zurückzukehren.

Definition 2.1.1. Man sagt, daß die nicht-randomisierte Testregel $\varphi : \mathbb{R}^n \to \{0,1\}$ einen *(nichtrandomisierten) statistischen Test zum Signifikanzniveau* α angibt, falls für $F \in \Lambda_0$ gilt.

$$\mathbb{P}_F\left(\varphi(X_1, \ldots, X_n) = 1\right) = P\left(H_0 \text{ verwerfen } \mid H_0 \text{ richtig }\right) \leq \alpha$$

Definition 2.1.2. 1. Wenn man H_0 verwirft, obwohl H_0 richtig ist, begeht man den sogenannten *Fehler 1. Art*. Die Wahrscheinlichkeit

$$\alpha_n(F) = \mathbb{P}_F\left(\varphi(x_1, \dots, x_n) = 1\right), \quad F \in \Lambda_0$$

heißt die *Wahrscheinlichkeit des Fehlers 1. Art* und soll unter dem Niveau α bleiben.

2. Den *Fehler 2. Art* begeht man, wenn man die falsche Hypothese H_0 nicht verwirft. Dabei ist

$$\beta_n(F) = \mathbb{P}_F\left(\varphi(x_1, \dots, x_n) = 0\right), \quad F \in \Lambda_1$$

die *Wahrscheinlichkeit des Fehlers 2. Art.*

Eine Zusammenfassung aller Möglichkeiten wird in folgender Tabelle festgehalten:

	H_0 richtig	H_0 falsch
H_0 verwerfen	Fehler 1. Art, Wahrscheinlichkeit $\alpha_n(F) \leq \alpha$	richtige Entscheidung
H_0 nicht verwerfen	richtige Entscheidung	Fehler 2. Art mit Wahrscheinlichkeit $\beta_n(F)$

Dabei sollen α_n und β_n möglichst klein sein, was gegenläufige Tendenzen darstellt, weil beim Kleinwerden von α die Wahrscheinlichkeit des Fehlers 2. Art notwendigerweise wächst.

Definition 2.1.3. 1. Die Funktion

$$G_n(F) = \mathbb{P}_F\left(\varphi(X_1, \dots, X_n) = 1\right), \quad F \in \Lambda$$

heißt *Gütefunktion* eines Tests φ.

2. Die Einschränkung von G_n auf Λ_1 heißt *Stärke*, *Schärfe* oder *Macht* (englisch *power*) des Tests φ.

   Es gilt

$$\begin{cases} G_n(F) = \alpha_n(F) \leq \alpha, \; F \in \Lambda_0 \\ G_n(F) = 1 - \beta_n(F), \; F \in \Lambda_1 \end{cases}$$

Beispiel 2.1.1. Parametrische Tests. Wie sieht ein parametrischer Test aus? Der Parameterraum Θ wird als $\Theta_0 \cup \Theta_1$ dargestellt, wobei $\Theta_0 \cap \Theta_1 = \emptyset$. Es gilt $\Lambda_0 = \{F_\theta : \theta \in \Theta_0\}$, $\Lambda_1 = \{F_\theta : \theta \in \Theta_1\}$. P_F wird zu P_θ, α_n, G_n und β_n werden statt auf Λ auf Θ definiert.

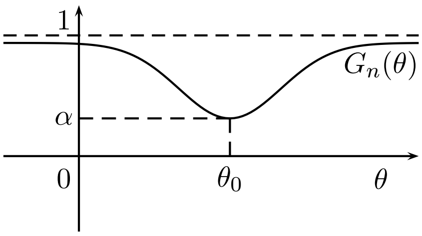
Welche Hypothesen H_0 und H_1 kommen oft bei parametrischen Tests vor? Zur Einfachheit betrachten wir den Spezialfall $\Theta = \mathbb{R}$.

1. $H_0 : \theta = \theta_0$ vs. $H_1 : \theta \neq \theta_0$
2. $H_0 : \theta \geq \theta_0$ vs. $H_1 : \theta < \theta_0$
3. $H_0 : \theta \leq \theta_0$ vs. $H_1 : \theta > \theta_0$
4. $H_0 : \theta \in [a, b]$ vs. $H_1 : \theta \notin [a, b]$

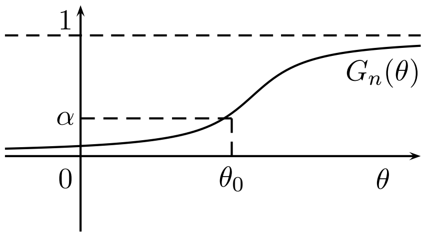
Im Fall (1) heißt der parametrische Test *zweiseitig*, in den Fällen (2) und (3) *einseitig* (*rechts*- bzw. *linksseitig*). In Fall (4) spricht man von der *Intervallhypothese* H_0.

Bei einem zweiseitigen bzw. einseitigen Test kann die Gütefunktion wie in Abbildung 2.1 (a) bzw. 2.1 (b) aussehen,

Abbildung 2.1: Gütefunktion



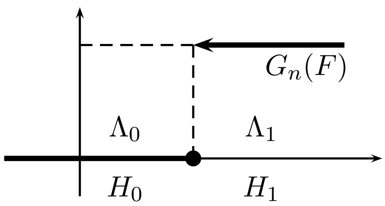
(a) eines zweiseitigen Tests



(b) eines einseitigen Tests

Bei einem allgemeinen (nicht notwendigerweise parametrischen) Modell kann man die ideale Gütefunktion wie in Abbildung 2.2 schematisch darstellen.

Abbildung 2.2: Schematische Darstellung der idealen Gütefunktion



- Man sieht aus Definition 2.1.2, dem Fehler 1. und 2. Art und der Ablehnungsregel, daß die Hypothesen H_0 und H_1 nicht symmetrisch behandelt werden, denn nur die Wahrscheinlichkeit des Fehlers 1. Art wird kontrolliert. Dies ist der Grund dafür, daß Statistiker die eigentlich interessierende Hypothese nicht als H_0, sondern als H_1 formulieren, damit, wenn man sich für H_1 entscheidet, man mit Sicherheit sagen kann, daß die Wahrscheinlichkeit der Fehlentscheidung unter dem Niveau α liegt.

- Wie wird ein statistischer, nicht randomisierter Test praktisch konstruiert? Die Konstruktion der Ablehnungsregel φ ähnelt sich sehr der von Konfidenzintervallen:
  1. Finde eine Teststatistik $T : \mathbb{R}^n \to \mathbb{R}$, die unter H_0 eine (möglicherweise asymptotisch für $n \to \infty$) bestimmte Prüfverteilung hat.
  2. Definiere $B_0 = [t_{\alpha_1}, t_{1-\alpha_2}]$, wobei t_{α_1} und $t_{1-\alpha_2}$ Quantile der Prüfverteilung von T sind, $\alpha_1 + \alpha_2 = \alpha \in [0, 1]$.
  3. Falls $T(X_1, \ldots, X_n) \in \mathbb{R} \setminus B_0 = B_1$, setze $\varphi(X_1, \ldots, X_n) = 1$. H_0 wird verworfen. Ansonsten setze $\varphi(X_1, \ldots, X_n) = 0$.
- Falls die Verteilung von T nur asymptotisch bestimmt werden kann, so heißt φ *asymptotischer Test.*
- Sehr oft aber ist auch die asymptotische Verteilung von T nicht bekannt. Dann verwendet man sogenannte *Monte-Carlo Tests*, in denen dann Quantile t_α näherungsweise aus sehr vielen Monte-Carlo-Simulationen von T (unter H_0) bestimmt werden: Falls t^i, $i = 1, \ldots, m$ die Werte von T in m unabhängigen Simulationsvorgängen sind, das heißt $t^i = T(x_1^i, \ldots, x_n^i)$, x_j^i sind unabhängige Realisierungen von $X_i \sim F \in \Lambda_0$, dann bildet man ihre Ordnungsstatistiken $t^{(1)}, \ldots, t^{(m)}$ und setzt $t_\alpha \approx t^{(\lfloor \alpha \cdot m \rfloor)}, \alpha \in [0, 1]$, wobei $t^0 = -\infty$.

Bemerkung 2.1.2. Man sieht deutlich, daß aus einem beliebigen Konfidenzintervall $I_\theta = \left[I_1^\theta(X_1, \ldots, X_n), I_2^\theta(X_1, \ldots, X_n)\right]$ zum Niveau $1 - \alpha$ für einen Parameter $\theta \in \mathbb{R}$ ein Test für θ konstruierbar ist. Die Hypothese $H_0 : \theta = \theta_0$ vs. $H_1 : \theta \neq \theta_0$ wird mit folgender Entscheidungsregel getestet:

$$\varphi(X_1, \ldots, X_n) = 1, \text{ falls } \theta_0 \in \left[I_1^{\theta_0}(X_1, \ldots, X_n),\, I_2^{\theta_0}(X_1, \ldots, X_n)\right].$$

Das Signifikanzniveau des Tests ist α.

Beispiel 2.1.2. *Normalverteilung, Test des Erwartungswertes bei bekannter Varianz.* Es seien $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$ mit bekannter Varianz σ^2. Ein Konfidenzintervall für μ ist

$$I^\mu = [I_1^\mu(X_1, \ldots, X_n),\, I_2^\mu(X_1, \ldots, X_n)] = \left[\overline{X}_n - \frac{z_{1-\alpha/2} \cdot \sigma}{\sqrt{n}}, \overline{X}_n + \frac{z_{1-\alpha/2} \cdot \sigma}{\sqrt{n}}\right]$$

(vergleiche Abschnitt 1.2.1, 1.)). H_0 wird verworfen, falls $|\mu_0 - \overline{X}_n| > \frac{z_{1-\alpha/2} \cdot \sigma}{\sqrt{n}}$. In der Testsprache bedeutet daß,

$$\varphi(x_1, \ldots, x_n) = \mathbb{I}\left((x_1, \ldots x_n) \in K_1\right),$$

wobei

$$K_1 = \left\{(x_1, \ldots, x_n) \in \mathbb{R}^n : |\mu_0 - \overline{x}_n| > \frac{z_{1-\alpha/2}}{\sqrt{n}}\right\}$$

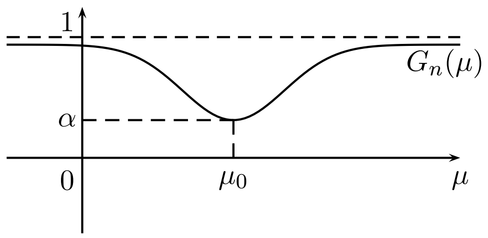
der Ablehnungsbereich ist. Für die Teststatistik $T(X_1, \ldots, X_n)$ gilt:

$$\begin{aligned} T(X_1, \ldots, X_n) &= \frac{\overline{X}_n - \mu_0}{\sigma}\sqrt{n} \sim N(0,1) \mid \text{ unter } H_0, \\ \alpha_n(\mu) &= \alpha. \end{aligned}$$

Berechnen wir nun die Gütefunktion (vergleiche Abbildung 2.3).

$$\begin{aligned} G_n(\mu) &= \mathbb{P}_\mu\left(|\mu_0 - \overline{X}_n| > \frac{z_{1-\alpha/2}}{\sqrt{n}}\right) = 1 - \mathbb{P}_\mu\left(\left|\overline{X}_n = \mu_0\right| \leq \frac{z_{1-\alpha/2}}{\sqrt{n}}\right) \\ &= 1 - \mathbb{P}_\mu\left(\left|\sqrt{n}\frac{\overline{X}_n - \mu}{\sigma} + \frac{\mu - \mu_0}{\sigma}\sqrt{n}\right| \leq z_{1-\alpha/2}\right) \\ &= 1 - \mathbb{P}_\mu\left(-z_{1-\alpha/2} - \frac{\mu - \mu_0}{\sigma}\sqrt{n} \leq \sqrt{n}\frac{\overline{X}_n - \mu}{\sigma} \leq z_{1-\alpha/2} - \frac{\mu - \mu_0}{\sigma}\sqrt{n}\right) \\ &= 1 - \Phi\left(z_{1-\alpha/2} - \frac{\mu - \mu_0}{\sigma}\sqrt{n}\right) + \Phi\left(-z_{1-\alpha/2} - \frac{\mu - \mu_0}{\sigma}\sqrt{n}\right) \\ &= \Phi\left(-z_{1-\alpha/2} + \frac{\mu - \mu_0}{\sigma}\sqrt{n}\right) + \Phi\left(-z_{1-\alpha/2} - \frac{\mu - \mu_0}{\sigma}\sqrt{n}\right) \end{aligned}$$

Abbildung 2.3: Gütefunktion für den zweiseitigen Test des Erwartungswertes einer Normalverteilung bei bekannter Varianz



Die „Ja-Nein "- Entscheidung des Testens wird oft als zu grob empfunden. Deswegen versucht man, ein feineres Maß der Verträglichkeit der Daten mit den Hypothesen H_0 und H_1 zu bestimmen. Dies ist der sogenannte p-Wert, der von den meisten Statistik-Softwarepaketen angegeben wird.

Definition 2.1.4. Es sei $(x_1, \ldots, x_n)$ die konkrete Stichprobe von Daten, die als Realisierung von $(X_1, \ldots, X_n)$ interpretiert wird und $T(X_1, \ldots, X_n)$ die Teststatistik, mit deren Hilfe die Entscheidungsregel φ konstruiert wurde. *Der p-Wert des statistischen Tests φ* ist das kleinste Signifikanzniveau, zu dem der Wert $t = T(x_1, \ldots, x_n)$ zur Verwerfung der Hypothese H_0 führt.

Im Beispiel eines einseitigen Tests mit dem Ablehnungsbereich $B_1 = (t, \infty)$ sagt man grob, daß

$$p = „\mathbb{P}\left(T(X_1, \ldots, X_n) > t \mid H_0\right)“,$$

wobei die Anführungszeichen bedeuten, daß dies keine klassische, sondern eine bedingte Wahrscheinlichkeit ist, die später präzise angegeben wird.

Bei der Verwendung des p-Wertes verändert sich die Ablehnungsregel: die Hypothese H_0 wird zum Signifikanzniveau α abgelehnt, falls $\alpha \geq p$. Früher hat man die Signifikanz der Testentscheidung (Ablehnung von H_0) an Hand folgender Tabelle festgesetzt:

p-Wert	Interpretation
$p \leq 0,001$	sehr stark signifikant
$0,001 < p \leq 0,01$	stark signifikant
$0,01 < p \leq 0,05$	schwach signifikant
$0,05 < p$	nicht signifikant

Da aber heute der p-Wert an sich verwendet werden kann, kann der Anwender der Tests bei vorgegebenem p-Wert selbst entscheiden, zu welchem Niveau er seine Tests durchführen will.

Bemerkung 2.1.3.

1. Das Signifikanzniveau darf nicht in Abhängigkeit von p festgelegt werden. Dies würde die allgemeine Testphilosophie zerstören!

2. Der p-Wert ist keine Wahrscheinlichkeit, sondern eine Zufallsvariable, denn er hängt von $(X_1, \ldots, X_n)$ ab. Der Ausdruck $p = \mathbb{P}(T(X_1, \ldots, X_n) > t \mid H_0)$, der in Definition 2.1.4 für den p-Wert eines einseitigen Tests mit Teststatistik T gegeben wurde, soll demnach als *Überschreitungswahrscheinlichkeit* interpretiert werden, daß bei Wiederholung des Zufallsexperiments unter H_0 der Wert $t = T(x_1, \ldots, x_n)$ oder extremere Werte in Richtung der Hypothese betrachtet werden:

$$p = \mathbb{P}\left(T(X_1', \ldots, X_n') > T(x_1, \ldots, x_n) \mid H_0\right),$$

wobei $(X_1', \ldots, X_n') \stackrel{d}{=} (X_1, \ldots, X_n)$. Falls wir von einer konkreten Realisierung $(x_1, \ldots, x_n)$ zur Zufallsstichprobe $(X_1, \ldots, X_n)$ übergehen, erhalten wir

$$p = p(X_1, \ldots, X_n) = \mathbb{P}\left(T(X_1', \ldots, X_n') > T(X_1, \ldots, X_n) \mid H_0\right)$$

3. Für andere Hypothesen H_0 wird der p-Wert auch eine andere Form haben. Zum Beispiel für

   a) einen symmetrischen zweiseitigen Test ist

   $$B_0 = \left[-t_{1-\alpha/2}, t_{1-\alpha/2}\right]$$

   der Akzeptanzbereich für H_0.

   $$\Rightarrow p = P\left(|T(X_1, \ldots, X_n)| > t \mid H_0\right), \, t = T(X_1, \ldots X_n)$$

   b) einen linksseitigen Test mit $B_0 = [t_\alpha, \infty]$ gilt

   $$p = P(T(X_1, \ldots, X_n) < t), \, t = T(X_1, \ldots, X_n)$$

c) Das Verhalten des p-Wertes kann folgendermaßen untersucht werden:
Lemma 2.1.1. Falls die Verteilungsfunktion F von X_i stetig monoton steigend ist (die Verteilung von T ist absolut stetig mit zum Beispiel steigender Dichte), dann ist $p \sim U[0,1]$.

Beweis. Wir zeigen es am speziellen Beispiel des rechtsseitigen Tests.

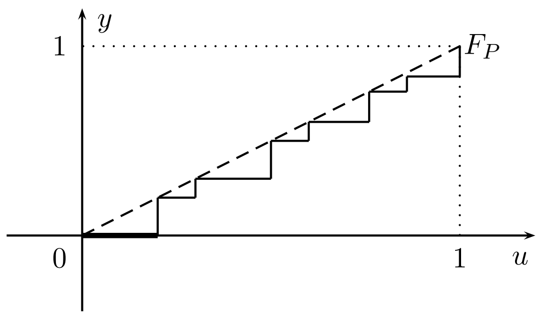
$$\begin{aligned}\mathbb{P}(p \le \alpha \mid H_0) &= \mathbb{P}\left(\overline{F}_T(T(X_1,\dots,X_n)) \le \alpha \mid H_0\right)\\ &= \mathbb{P}(F_T(T(X_1,\dots,X_n)) \ge 1-\alpha \mid H_0)\\ &= \mathbb{P}(U \ge 1-\alpha) = 1-(1-\alpha) = \alpha, \quad \alpha \in [0,1],\end{aligned}$$

da $F_T(T(X_1,\dots,X_n)) \stackrel{d}{=} U \sim U[0,1]$ und F_T absolut stetig ist. □

Übung 2.1.1. Zeigen Sie, daß für eine beliebige Zufallsvariable X mit absolut stetiger Verteilung, und streng monoton steigender Verteilungsfunktion F_X gilt:

$$F_X(X) \sim U[0,1]$$

Abbildung 2.4: Verteilung von p für diskrete T



Falls die Verteilung von T diskret ist, mit dem Wertebereich $\{t_1,\dots,t_n\}$, $t_i < t_j$ für $i < j$, so ist auch die Verteilung von p diskret, somit gilt nicht $p \sim U[0,1]$. In diesem Fall ist $F_T(x)$ eine Treppenfunktion, die die Gerade $y = u$ in den Punkten $u = \sum\limits_{i=1}^{k} \mathbb{P}(T(X_1,\dots,X_n) = t_i)$, $k = 1\dots n$ berührt (vgl. Abbildung 2.4).

Definition 2.1.5. 1. Falls die Macht $G_n(\cdot)$ eines Tests φ zum Niveau α die Ungleichung

$$G_n(F) \ge \alpha, \quad F \in \Lambda_1$$

erfüllt, dann heißt der Test *unverfälscht*.

2. Es seien φ und φ^* zwei Tests zum Niveau α mit Gütefunktionen $G_n(\cdot)$ und $G_n^*(\cdot)$. Man sagt, daß der Test φ *besser* als φ^* ist, falls er eine größere Macht besitzt:

$$G_n(F) \geq G_n^*(F) \quad \forall F \in \Lambda_1$$

3. Der Test φ heißt konsistent, falls $G_n(F) \underset{n\to\infty}{\longrightarrow} 1$ für alle $F \in \Lambda_1$.

Bemerkung 2.1.4. 1. Die einseitigen Tests haben oft eine größere Macht als ihre zweiseitigen Version.

Beispiel 2.1.3. Betrachten wir zum Beispiel den Gauß-Test des Erwartungswertes der Normalverteilung bei bekannter Varianz. Beim zweiseitigen Test

$$H_0 : \mu = \mu_0 \text{ vs. } H_1 : \mu \neq \mu_0.$$

erhalten wir die Gütefunktion

$$G_n(\mu) = \Phi\left(-z_{1-\alpha/2} + \sqrt{n}\frac{\mu - \mu_0}{\sigma}\right) + \Phi\left(-z_{1-\alpha/2} - \sqrt{n}\frac{\mu - \mu_0}{\sigma}\right).$$

Beim einseitigen Test φ^* der Hypothesen

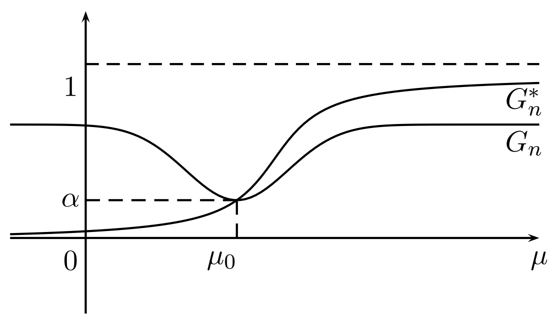
$$H_0^* : \mu \leq \mu_0 \text{ vs. } H_1^* : \mu > \mu_0$$

ist seine Gütefunktion gleich

$$G_n^*(\mu) = \Phi\left(-z_{1-\alpha} + \sqrt{n}\frac{\mu - \mu_0}{\sigma}\right)$$

Beide Tests sind offensichtlich konsistent, denn $G_n(\mu) \underset{n\to\infty}{\to} 1$, $G_n^*(\mu) \underset{n\to\infty}{\to} 1$. Dabei ist φ^* besser als φ. Beide Tests sind unverfälscht (vergleiche Abbildung 2.5).

Abbildung 2.5: Gütefunktionen eines ein- bzw. zweiseitigen Tests der Erwartungswertes einer Normalverteilung



2. Beim Testen einer Intervallhypothese $H_0 : \theta \in [a, b]$ vs. $H_1 : \theta \notin [a, b]$ zum Niveau α kann man wie folgt vorgehen: Teste

a) $H_0^a : \theta \geq a$ vs. $H_1 : \theta < a$ zum Niveau $\alpha/2$.
b) $H_0^b : \theta \leq b$ vs. $H_1 : \theta > b$ zum Niveau $\alpha/2$.

H_0 wird nicht abgelehnt, falls H_0^a und H_0^b nicht abgelehnt werden. Die Wahrscheinlichkeit des Fehlers 1. Art ist hier α. Die Macht dieses Tests ist im Allgemeinen schlecht.

Bemerkung 2.1.5. Je mehr Parameter für der Aufbau der Teststatistik T geschätzt werden müssen, desto kleiner wird in der Regel die Macht.

2.2 Nichtrandomisierte Tests

2.2.1 Parametrische Signifikanztests

In diesem Abschnitt geben wir Beispiele einiger Tests, die meistens aus den entsprechenden Konfidenzintervallen für die Parameter von Verteilungen entstehen. Deshalb werden wir sie nur kurz behandeln.

1. **Tests für die Parameter der Normalverteilung** $N(\mu, \sigma^2)$

   a) **Test von** μ **bei unbekannter Varianz**
      - Hypothesen: $H_0 : \mu = \mu_0$ vs. $H_1 : \mu \neq \mu_0$.
      - Teststatistik:

$$T(X_1, \ldots, X_n) = \frac{\overline{X}_n - \mu_0}{S_n} \sim t_{n-1} \quad | \; H_0$$

      - Entscheidungsregel:

$$\varphi(X_1, \ldots, X_n) = 1, \text{ falls } |T(X_1, \ldots, X_n)| > t_{n-1,1-\alpha/2}.$$

   b) **Test von** σ^2 **bei bekanntem** μ
      - Hypothesen: $H_0 : \sigma^2 = \sigma_0^2$ vs. $H_1 : \sigma^2 \neq \sigma_0^2$.
      - Teststatistik:

$$T(X_1, \ldots, X_n) = \frac{n\tilde{S}_n^2}{\sigma_0^2} \sim \chi_n^2 \quad | \; H_0$$

      mit $\tilde{S}_n^2 = \frac{1}{n} \sum\limits_{i=1}^{n} (X_i - \mu)^2$.
      - Entscheidungsregel:

$$\varphi(X_1, \ldots, X_n) = 1, \text{ falls } T(X_1, \ldots, X_n) \notin \left[\chi^2_{n,\alpha/2}, \chi^2_{n,1-\alpha/2}\right].$$

- Gütefunktion:

$$
\begin{aligned}
G_n(\sigma^2) &= 1 - \mathbb{P}_{\sigma^2}\left(\chi^2_{n,\alpha/2} \le \frac{n\tilde{S}_n^2}{\sigma_0^2} \le \chi^2_{n,1-\alpha/2}\right) \\
&= 1 - \mathbb{P}_{\sigma^2}\left(\frac{\chi^2_{n,\alpha/2}\sigma_0^2}{\sigma^2} \le \frac{n\tilde{S}_n^2}{\sigma_2} \le \frac{\chi^2_{n,1-\alpha/2}\sigma_0^2}{\sigma^2}\right) \\
&= 1 - F_{\chi^2_n}\left(\chi^2_{n,1-\alpha/2}\frac{\sigma_0^2}{\sigma^2}\right) + F_{\chi^2_n}\left(\chi^2_{n,\alpha/2}\frac{\sigma_0^2}{\sigma_2}\right)
\end{aligned}
$$

c) **Test von σ^2 bei unbekanntem μ**
   - Hypothesen: $H_0 : \sigma^2 = \sigma_0^2$ vs. $H_1 : \sigma^2 \neq \sigma_0^2$.
   - Teststatistik:

$$
T(X_1, \ldots, X_n) = \frac{(n-1)S_n^2}{\sigma_0^2} \sim \chi^2_{n-1} \quad | \; H_0,
$$

   wobei $S_n^2 = \frac{1}{n-1}\sum\limits_{i=1}^{n}\left(X_i - \overline{X}_n\right)^2$.
   - Entscheidungsregel:

$$
\varphi(X_1, \ldots, X_n) = 1, \text{ falls } T(X_1, \ldots, X_n) \notin \left[\chi^2_{n-1,\alpha/2}, \chi^2_{n-1,1-\alpha/2}\right].
$$

   Übung 2.2.1. (i) Finden Sie $G_n(\cdot)$ für die einseitige Version der obigen Tests.
   (ii) Zeigen Sie, daß diese einseitigen Tests unverfälscht sind, die zweiseitigen aber nicht.

2. **Asymptotische Tests**

   Bei asymptotischen Tests ist die Verteilung der Teststatistik nur näherungsweise (für große n) bekannt. Ebenso asymptotisch wird das Konfidenzniveau α erreicht. Ihre Konstruktion basiert meistens auf Verwendung der Grenzwertsätze.

   Die allgemeine Vorgehensweise wird im sogenannten *Wald-Test* (genannt nach dem Statistiker Abraham Wald (1902-1980)) fixiert:

   - Sei $(X_1, \ldots, X_n)$ eine Zufallsstichprobe, X_i seien unabhängig und identisch verteilt für $i = 1, \ldots, n$, mit $X_i \sim F_\theta$, $\theta \in \Theta \subseteq \mathbb{R}$.
   - Wir testen $H_0 : \theta = \theta_0$ vs. $H_1 : \theta \neq \theta_0$. Es sei $\hat{\theta}_n = \hat{\theta}(X_1, \ldots, X_n)$ ein erwartungstreuer, asymptotisch normalverteilter Schätzer für θ.

$$
\frac{\hat{\theta}_n - \theta_0}{\hat{\sigma}_n} \xrightarrow[n\to\infty]{d} Y \sim N(0,1) \quad | \; H_0,
$$

   wobei $\hat{\sigma}_n^2$ ein konsistenter Schätzer für die Varianz von $\hat{\theta}_n$ sei.
   Die Teststatistik ist

$$
T(X_1, \ldots, X_n) = \frac{\hat{\theta}_n(X_1, \ldots, X_n) - \theta_0}{\hat{\sigma}_n}.
$$

- Die Entscheidungsregel lautet: H_0 wird abgelehnt, wenn $|T(X_1, \ldots, X_n)| > z_{1-\alpha/2}$, wobei $z_{1-\alpha/2} = \Phi^{-1}(1-\alpha/2)$. Diese Entscheidungsregel soll nur bei großen n verwendet werden. Die Wahrscheinlichkeit des Fehlers 1. Art ist asymptotisch gleich α, denn $\mathbb{P}(|T(X_1, \ldots, X_n)| > z_{1-\alpha/2} \mid H_0) \underset{n\to\infty}{\to} \alpha$ wegen der asymptotischen Normalverteilung von T.

  Die Gütefunktion des Tests ist asymptotisch gleich

$$\lim_{n\to\infty} G_n(\theta) = 1 - \Phi\left(z_{1-\alpha/2} + \frac{\theta_0 - \theta}{\hat{\sigma}_n}\right) + \Phi\left(-z_{1-\alpha/2} + \frac{\theta_0 - \theta}{\hat{\sigma}_n}\right)$$

  Spezialfälle des Wald-Tests sind asymptotische Tests der Erwartungswerte bei einer Poisson- oder Bernoulliverteilten Stichprobe.

Beispiel 2.2.1. a) **Bernoulliverteilung**

Es seien $X_i \sim \text{Bernoulli}(p)$, $p \in [0, 1]$ unabhängige, identisch verteilte Zufallsvariablen.

- Hypothesen: $H_0 : p = p_0$ vs. $H_1 : p \neq p_0$.
- Teststatistik:

$$T(X_1, \ldots, X_n) = \begin{cases} \sqrt{n}\frac{\overline{X}_n - p_0}{\sqrt{\overline{X}_n(1-\overline{X}_n)}}, & \text{falls } \overline{X}_n \neq 0, 1, \\ 0, & \text{sonst.} \end{cases}$$

  Unter H_0 gilt: $T(X_1, \ldots, X_n) \xrightarrow[n\to\infty]{d} Y \sim N(0, 1)$.

b) **Poissonverteilung**

Es seien $X_i \sim \text{Poission}(\lambda)$, $\lambda > 0$ unabhängige, identisch verteilte Zufallsvariablen.

- Hypothesen: $H_0 : \lambda = \lambda_0$ vs. $H_1 : \lambda \neq \lambda_0$
- Teststatistik:

$$T(X_1, \ldots, X_n) = \begin{cases} \sqrt{n}\frac{\overline{X}_n - \lambda_0}{\sqrt{\overline{X}_n}}, & \text{falls } \overline{X}_n > 0, \\ 0, & \text{sonst.} \end{cases}$$

Unter H_0 gilt: $T(X_1, \ldots, X_n) \xrightarrow[n\to\infty]{d} Y \sim N(0, 1)$

3. **Zwei-Stichproben-Probleme**

Gegeben seien zwei Zufallsstichproben $Y_1 = (X_{11}, \ldots, X_{1n_1})$, $Y_2 = (Y_{21}, \ldots, Y_{2n_2})$, $n = \max\{n_1, n_2\}$. X_{ij} seien unabhängig für $j = 1, \ldots, n_i$, $X_{ij} \sim F_{\theta_i}$, $i = 1, 2$.

a) **Test der Gleichheit zweier Erwartungswerte bei normalverteilten Stichproben**

- **bei bekannten Varianzen**

  Es seien $X_{ij} \sim N(\mu_i, \sigma_i^2)$, $i = 1, 2$, $j = 1, \ldots, n$. Dabei seien σ_1^2, σ_2^2 bekannt, X_{ij} seien unabhängig voneinander für alle i, j.

Die Hypothesen sind $H_0 : \mu_1 = \mu_2$ vs. $H_1 : \mu_1 \neq \mu_2$. Wir betrachten die Testgröße:

$$T(Y_1, Y_2) = \frac{\overline{X}_{1n_1} - \overline{X}_{2n_2}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Unter H_0 gilt: $T(Y_1, Y_2) \sim N(0, 1)$. Als Entscheidungsregel gilt: H_0 wird abgelehnt, falls $|T(Y_1, Y_2)| > z_{1-\alpha/2}$.

- **bei unbekannten (jedoch gleichen) Varianzen**
  Es seien $X_{ij} \sim N(\mu_i, \sigma_i^2)$, $i = 1, 2$, $j = 1, \ldots, n$. Dabei seien σ_1^2, σ_2^2 unbekannt, $\sigma_1^2 = \sigma_2^2$ und X_{ij} seien unabhängig voneinander für alle i, j.
  Die Hypothesen sind: $H_0 : \mu_1 = \mu_2$ vs. $H_1 : \mu_1 \neq \mu_2$. Wir betrachten die Teststatistik

$$T(Y_1, Y_2) = \frac{\overline{X}_{1n_1} - \overline{X}_{2n_2}}{S_{n_1 n_2}} \sqrt{\frac{n_1 n_2}{n_1 + n_2}},$$

wobei

$$S^2_{n_1 n_2} = \frac{1}{n_1 + n_2 - 2} \cdot \left(\sum_{j=1}^{n_1} \left(X_{ij} - \overline{X}_{1n_1} \right)^2 + \sum_{j=1}^{n_2} \left(X_{2j} - \overline{X}_{2n_2} \right)^2 \right).$$

Man kann zeigen, daß unter H_0 gilt: $T(Y_1, Y_2) \sim t_{n_1 - n_2 - 2}$. Die Entscheidungsregel lautet: H_0 ablehnen, falls $|T(Y_1, Y_2)| > t_{n_1 + n_2 - 2, 1-\alpha/2}$.

b) **Test der Gleichheit von Erwartungswerten bei verbundenen Stichproben**
Es seien $Y_1 = (X_{11}, \ldots, X_{1n})$ und $Y_2 = (X_{21}, \ldots, X_{2n})$, $n_1 = n_2 = n$,

$$Z_j = X_{1j} - X_{2j} \sim N(\mu_1 - \mu_2, \sigma^2),\ j = 1, \ldots, n$$

unabhängig und identisch verteilt mit $\mu_i = \mathbb{E}\, X_{ij}$, $i = 1, 2$. Die Hypothesen sind: $H_0 : \mu_1 = \mu_2$ vs. $H_1 : \mu_1 \neq \mu_2$ bei unbekannter Varianz σ^2. Als Teststatistik verwenden wir

$$T(Z_1, Z_2) = \sqrt{n} \frac{\overline{Z}_n}{S_n},$$

wobei

$$S_n^2 = \frac{1}{n-1} \sum_{j=1}^{n} \left(Z_j - \overline{Z}_n \right)^2.$$

Unter H_0 gilt dann: $T(Z_1, \ldots, Z_n) \sim t_{n-1}$. Die Entscheidungsregel lautet: H_0 wird abgelehnt, falls $|T(z_1, \ldots, z_n)| > t_{n-1, 1-\alpha/2}$.

c) **Test der Gleichheit von Varianzen bei unabhängigen Gaußschen Stichproben**
Es seien $Y_1 = (X_{11}, \ldots, X_{1n_1})$ und $Y_2 = (X_{21}, \ldots, X_{2n_2})$ unabhängig und identisch verteilt mit $X_{ij} \sim N(\mu_i, \sigma_i^2)$, wobei μ_i und σ_i^2 beide unbekannt sind. Die Hypothesen sind: $H_0 : \sigma_1^2 = \sigma_2^2$ vs. $H_1 : \sigma_1^2 \neq \sigma_2^2$. Als Teststatistik verwenden wir

$$T(Y_1, Y_2) = \frac{S_{2n_2}^2}{S_{1n_1}^2},$$

wobei

$$S_{in_i}^2 = \frac{1}{n_i - 1} \sum_{j=1}^{n} \left(X_{ij} - \overline{X}_{in_i}\right)^2, \; i = 1, 2.$$

Unter H_0 gilt: $T(Y_1, Y_2) \sim F_{n_2-1,n_1-1}$. Die Entscheidungsregel lautet: H_0 wird abgelehnt, falls $T(Y_1, Y_2) \notin \left[F_{n_2-1,n_1-1,\alpha/2},\, F_{n_2-1,n_1-1,1-\alpha/2}\right]$.

d) **Asymptotische Zwei-Stichproben-Tests**
(i) **bei Bernoulli-verteilten Stichproben**
Es gilt $X_{ij} \sim \text{Bernoulli}(p_i)$, $j = 1, \ldots, n_i$, $i = 1, 2$. Die Hypothesen sind $H_0 : p_1 = p_2$ vs. $H_1 : p_1 \neq p_2$. Als Teststatistik verwenden wir

$$T(Y_1, Y_2) = \frac{\overline{X}_{1n_1} - \overline{X}_{2n_2}}{\sqrt{\frac{\overline{X}_{1n_1}(1-\overline{X}_{1n_1})}{n_1} + \frac{\overline{X}_{2n_2}(1-\overline{X}_{2n_2})}{n_2}}}$$

Unter H_0 gilt: $T(Y_1, Y_2) \underset{n_1,n_2\to\infty}{\xrightarrow{d}} Y \sim N(0, 1)$. Die Entscheidungsregel lautet: H_0 wird verworfen, falls $|T(Y_1, Y_2)| > z_{1-\alpha/2}$. Dies ist ein Test zum asymptotischen Signifikanzniveau α.

(ii) **bei Poisson-verteilten Stichproben**
Es seien X_{ij} unabhängig, $X_{ij} \sim \text{Poisson}(\lambda_i)$, $i = 1, 2$. Die Hypothesen sind: $H_0 : \lambda_1 = \lambda_2$ vs. $H_1 : \lambda_1 \neq \lambda_2$. Als Teststatistik verwenden wir:

$$T(Y_1, Y_2) = \frac{\overline{X}_{1n_1} - \overline{X}_{2n_2}}{\sqrt{\frac{\overline{X}_{1n_1}}{n_1} + \frac{\overline{X}_{2n_2}}{n_2}}}$$

Die Entscheidungsregel lautet: H_0 ablehnen, falls $|T(Y_1, Y_2)| > z_{1-\alpha/2}$. Dies ist ein Test zum asymptotischen Niveau α.

Bemerkung 2.2.1. Asymptotische Tests dürfen nur für große Stichprobenumfänge verwendet werden. Bei ihrer Verwendung für kleine Stichproben kann das asymptotische Signifikanzniveau nicht garantiert werden.

2.3 Randomisierte Tests

In diesem Abschnitt werden wir klassische Ergebnisse von Neyman-Pearson über die besten Tests präsentieren. Dabei werden randomisierte Tests eine wichtige Rolle spielen.

2.3.1 Grundlagen

Gegeben sei eine Zufallsstichprobe $(X_1, \ldots, X_n)$ von unabhängigen und identisch verteilten Zufallsvariablen X_i mit konkreter Ausprägung $(x_1, \ldots, x_n)$. Sei unser Stichprobenraum $(B, \mathcal{B})$ entweder $(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n})$ oder $(\mathbb{N}_0^n, \mathcal{B}_{\mathbb{N}_0^n})$, je nachdem, ob die Stichprobenvariablen X_i, $i = 1, \ldots, n$ absolut stetig oder diskret verteilt sind.

Hier wird zur Einfachheit im Falle einer diskret verteilten Zufallsvariable X_i ihr diskreter Wertebereich mit $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$ gleichgesetzt. Der Stichprobenraum sei mit einem Maß μ versehen, wobei

$$\mu = \begin{cases} \text{Lebesgue-Maß auf } \mathbb{R}^n, & \text{falls } B = \mathbb{R}^n, \\ \text{Zählmaß auf } \mathbb{N}_0^n, & \text{falls } B = \mathbb{N}_0^n. \end{cases}$$

Dementsprechend gilt

$$\int_{\mathcal{B}} g(x)\mu(dx) = \begin{cases} \int_{\mathbb{R}^n} g(x)dx, & \text{falls } B = \mathbb{R}^n, \\ \sum_{x \in \mathbb{N}_0^n} g(x), & \text{falls } B = \mathbb{N}_0^n. \end{cases}$$

Es sei zusätzlich $X_i \sim F_\theta$, $\theta \in \Theta \subseteq \mathbb{R}^m$, $i = 1, \ldots, n$ (parametrisches Modell). Für $\Theta = \Theta_0 \cup \Theta_1$, $\Theta_0 \cap \Theta_1 = \emptyset$ formulieren wir die Hypothesen $H_0 : \theta \in \Theta_0$ vs. $H_1 : \theta \in \Theta_1$, die mit Hilfe eines randomisierten Tests

$$\varphi(x) = \begin{cases} 1, & x \in K_1, \\ \gamma \in (0,1), & x \in K_{01} \\ 0, & x \in K_0 \end{cases} \qquad x = (x_1, \ldots, x_n),$$

getestet werden.

Im Falle $x \in K_{01}$ wird mit Hilfe einer Zufallsvariable $Y \sim \text{Bernoulli}(\varphi(x))$ entschieden, ob H_0 verworfen wird $(Y = 1)$ oder nicht $(Y = 0)$.

Definition 2.3.1. Die *Gütefunktion* eines randomisierten Tests φ sei

$$G_n(\theta) = G_n(\varphi, \theta) = \mathbb{E}_\theta\, \varphi(X_1, \ldots, X_n),\, \theta \in \Theta.$$

1. Der Test φ hat das *Signifikanzniveau* $\alpha \in [0, 1]$, falls $G_n(\varphi, \theta) \leq \alpha$, $\forall\, \theta \in \Theta_0$ ist. Die Zahl

$$\sup_{\theta \in \Theta_0} G_n(\varphi, \theta)$$

wird *Umfang* des Tests φ genannt. Offensichtlich ist der Umfang eines Niveau-α-Tests kleiner gleich α.

2. Sei $\Psi(\alpha)$ die Menge aller Tests zum Niveau α. Der Test $\varphi_1 \in \Psi(\alpha)$ ist *(gleichmäßig) besser* als Test $\varphi_2 \in \Psi(\alpha)$, falls $G_n(\varphi_1, \theta) \geq G_n(\varphi_2, \theta)$, $\theta \in \Theta_1$, also falls φ_1 eine größere Macht besitzt.

3. Ein Test $\varphi^* \in \Psi(\alpha)$ ist *(gleichmäßig) bester Test* in $\Psi(\alpha)$, falls
$$G_n(\varphi^*, \theta) \geq G_n(\varphi, \theta), \text{ für alle Tests } \varphi \in \Psi(\alpha),\ \theta \in \Theta_1.$$

Bemerkung 2.3.1. 1. Definition 2.3.1 1) ist eine offensichtliche Verallgemeinerung der Definition 2.1.3 der Gütefunktion eines nicht-randomisierten Tests φ. Nämlich, für $\varphi(x) = \mathbb{I}(x \in K_1)$ gilt:
$$\begin{aligned} G_n(\varphi, \theta) &= \mathbb{E}_\theta\, \varphi(X_1, \ldots, X_n) \\ &= \mathbb{P}_\theta\left((X_1, \ldots, X_n) \in K_1\right) \\ &= \mathbb{P}_\theta\left(H_0 \text{ ablehnen}\right),\ \theta \in \Theta. \end{aligned}$$

2. Ein bester Test φ^* in $\Psi(\alpha)$ existiert nicht immer, sondern nur unter gewissen Voraussetzungen an $\mathbb{P}_\theta, \Theta_0, \Theta_1$ und $\Psi(\alpha)$.

2.3.2 Neyman-Pearson-Tests bei einfachen Hypothesen

In diesem Abschnitt betrachten wir einfache Hypothesen
$$H_0 : \theta = \theta_0 \quad \text{vs.} \quad H_1 : \theta = \theta_1 \tag{2.3.1}$$
wobei $\theta_0, \theta_1 \in \Theta$, $\theta_1 \neq \theta_0$.

Dementsprechend sind $\Theta_0 = \{\theta_0\}$, $\Theta_1 = \{\theta_1\}$. Wir setzen voraus, daß F_{θ_i} eine Dichte $g_i(x)$ bezüglich μ besitzt, $i = 0, 1$. Führen wir einige abkürzende Bezeichnungen $\mathbb{P}_0 = \mathbb{P}_{\theta_0}$, $\mathbb{P}_1 = \mathbb{P}_{\theta_1}$, $\mathbb{E}_0 = \mathbb{E}_{\theta_0}$, $\mathbb{E}_1 = \mathbb{E}_{\theta_1}$ ein. Sei $f_i(x) = \prod_{j=1}^n g_i(x_j)$, $x = (x_1, \ldots, x_n)$, $i = 0, 1$ die Dichte der Stichprobe unter H_0 bzw. H_1.

Definition 2.3.2. Ein *Neyman-Pearson-Test* (*NP-Test*) der einfachen Hypothesen in (2.3.1) ist gegeben durch die Regel
$$\varphi(x) = \varphi_K(x) = \begin{cases} 1, & \text{falls } f_1(x) > K f_0(x), \\ \gamma, & \text{falls } f_1(x) = K f_0(x), \\ 0, & \text{falls } f_1(x) < K f_0(x) \end{cases} \tag{2.3.2}$$
für Konstanten $K \geq 0$ und $\gamma \in [0, 1]$.

Bemerkung 2.3.2. 1. Manchmal werden $K = K(x)$ und $\gamma = \gamma(x)$ als Funktionen von x und nicht als Konstanten betrachtet.

2. Der *Ablehnungsbereich* des Neyman-Pearson-Tests φ_K ist
$$K_1 = \{x \in B : f_1(x) > K f_0(x)\}.$$

3. Der *Umfang* des Neyman-Pearson-Tests φ_K ist

$$\begin{aligned}\mathbb{E}_0\,\varphi_K(X_1,\ldots,X_n) = \mathbb{P}_0\left(f_1(X_1,\ldots,X_n) > Kf_0(X_1,\ldots X_n)\right) + \\ + \gamma\mathbb{P}_0\left(f_1(X_1,\ldots,X_n)\right) = Kf_0(X_1,\ldots,X_n)\end{aligned}$$

4. Die Definition 2.3.2 kann man äquivalent folgendermaßen geben: Wir definieren eine Teststatistik

$$T(x) = \begin{cases} \frac{f_1(x)}{f_0(x)}, & x \in B : f_0(x) > 0, \\ \infty, & x \in B : f_0(x) = 0. \end{cases}$$

Dann wird der neue Test

$$\tilde{\varphi}_K(x) = \begin{cases} 1, & \text{falls } T(x) > K, \\ \gamma, & \text{falls } T(x) = K, \\ 0, & \text{falls } T(x) < K \end{cases}$$

eingeführt, der für P_0- und P_1- fast alle $x \in B$ äquivalent zu φ_k ist. In der Tat gilt $\varphi_K(x) = \tilde{\varphi}_K(x)\,\forall x \in B \setminus C$, wobei $C = \{x \in B : f_0(x) = f_1(x) = 0\}$ das P_0- bzw. P_1-Maß Null besitzt.

In der neuen Formulierung ist der Umfang von φ bzw. $\tilde{\varphi}_K$ gleich

$$\mathbb{E}_0\,\tilde{\varphi}_K = \mathbb{P}_0(T(X_1,\ldots,X_n) > K) + \gamma \cdot \mathbb{P}_0\left(T(X_1,\ldots,X_n) = K\right).$$

Satz 2.3.1. Optimalitätssatz

Es sei φ_K ein Neyman-Pearson-Test für ein $K \geq 0$ und $\gamma \in [0,1]$. Dann ist φ_K der beste Test zum Niveau $\alpha = \mathbb{E}_0\,\varphi_K$ seines Umfangs.

Beweis. Sei $\varphi \in \Psi(\alpha)$, also $\mathbb{E}_0\left(\varphi(X_1,\ldots,X_n)\right) \leq \alpha$. Um zu zeigen, daß φ_K besser als φ ist, genügt es bei einfachen Hypothesen H_0 und H_1 zu zeigen, daß $\mathbb{E}_1\,\varphi_K(X_1,\ldots,X_n) \geq \mathbb{E}_1\,\varphi(X_1,\ldots,X_n)$. Wir führen dazu die folgenden Mengen ein:

$$\begin{aligned}\mu^+ &= \{x \in B : \varphi_K(x) > \varphi(x)\} \\ \mu^- &= \{x \in B : \varphi_K(x) < \varphi(x)\} \\ \mu^= &= \{x \in B : \varphi_K(x) = \varphi(x)\}\end{aligned}$$

Es gilt offensichtlich $x \in \mu^+ \Rightarrow \varphi_K(x) > 0 \Rightarrow f_1(x) \geq Kf_0(x)$,

$$x \in \mu^- \Rightarrow \varphi_K(x) < 1 \Rightarrow f_1(x) \leq Kf_0(x).$$

Als Folgerung erhalten wir

$$\begin{aligned}
\mathbb{E}_1\left(\varphi_K(X_1,\ldots,X_n)-\varphi(X_1,\ldots,X_n)\right) &= \int_B (\varphi_K(x)-\varphi(x))f_1(x)\mu(dx)\\
&= \left(\int_{\mu^+}+\int_{\mu^-}+\int_{\mu^=}\right)(\varphi_K(x)-\varphi(x))f_1(x)\mu(dx)\\
&\geq \int_{\mu^+}(\varphi_K(x)-\varphi(x))Kf_0(x)\mu(dx)\\
&\quad+\int_{\mu^-}(\varphi_K(x)-\varphi(x))Kf_0(x)\mu(dx)\\
&= \int_{B=\mu^+\cup\mu^-\cup\mu^=}(\varphi_K(x)-\varphi(x))Kf_0(x)\mu(dx)\\
&= K\left[\mathbb{E}_0\,\varphi_K(X_1,\ldots,X_n)-\mathbb{E}_0\,\varphi(X_1,\ldots,X_n)\right]\\
&\geq K(\alpha-\alpha)=0,
\end{aligned}$$

weil beide Tests das Niveau α haben. Damit ist die Behauptung bewiesen. □

Bemerkung 2.3.3. 1. Da im Beweis γ nicht vorkommt, wird derselbe Beweis im Falle von $\gamma(x)\neq$ const gelten.

2. Aus dem Beweis folgt die Gültigkeit der Ungleichung

$$\int_B (\varphi_K(x)-\varphi(x))\left(f_1(x)-Kf_0(x)\right)\mu(dx)\geq 0$$

im Falle des konstanten K, bzw.

$$\mathbb{E}_1\left(\varphi_K(X_1,\ldots,X_n)-\varphi(X_1,\ldots,X_n)\right)\geq\int_B (\varphi_K(x)-\varphi(x))\,K(x)f_0(x)\mu(dx)$$

im allgemeinen Fall.

Satz 2.3.2. (Fundamentallemma von Neyman-Pearson)

1. Zu einem beliebigen $\alpha\in(0,1)$ gibt es einen Neyman-Pearson-Test φ_K mit Umfang α, der dann nach Satz 2.3.1 der beste Niveau-α-Test ist.
2. Ist φ ebenfalls bester Test zum Niveau α, so gilt $\varphi(x)=\varphi_K(x)$ für μ-fast alle $x\in K_0\cup K_1=\{x\in B: f_1(x)\neq Kf_0(x)\}$ und φ_K aus Teil 1).

Beweis. 1. Für $\varphi_K(x)$ gilt

$$\varphi_K(x)=\begin{cases}1, & \text{falls } x\in K_1=\{x: f_1(x)>K\cdot f_0(x)\},\\ \gamma, & \text{falls } x\in K_{01}=\{x: f_1(x)=K\cdot f_0(x)\},\\ 0, & \text{falls } x\in K_0=\{x: f_1(x)<K\cdot f_0(x)\}.\end{cases}$$

Der Umfang von φ_K ist

$$\mathbb{P}_0\left(T(X_1,\ldots,X_n)>K\right)+\gamma\mathbb{P}_0\left(T(X_1,\ldots,X_n)=K\right)=\alpha, \qquad (2.3.3)$$

wobei

$$T(x_1,\dots,x_n) = \begin{cases} \frac{f_1(x_1,\dots,x_n)}{f_0(x_1,\dots,x_n)}, & \text{falls } f_0(x) > 0, \\ \infty, & \text{sonst.} \end{cases}$$

Nun suchen wir ein $K > 0$ und ein $\gamma \in [0,1]$, sodaß Gleichung (2.3.3) stimmt. Es sei $\tilde{F}_0(x) = \mathbb{P}(T(X_1,\dots,X_n) \le x)$, $x \in \mathbb{R}$ die Verteilungsfunktion von T. Da $T \ge 0$ ist, gilt $\tilde{F}_0(x) = 0$, falls $x < 0$. Außerdem ist $\mathbb{P}(T(X_1,\dots,X_n) < \infty) = 1$, das heißt $\tilde{F}^{-1}(\alpha) \in [0,\infty)$, $\forall\, \alpha \in (0,1)$. Die Gleichung (2.3.3) kann dann folgendermaßen umgeschrieben werden:

$$1 - \tilde{F}_0(K) + \gamma\left(\tilde{F}_0(K) - \tilde{F}_0(K-)\right) = \alpha \tag{2.3.4}$$

wobei $\tilde{F}_0(K-) = \lim\limits_{x \to K-0} \tilde{F}_0(x)$.

Sei $K = \tilde{F}_0^{-1}(1-\alpha)$, dann gilt:

a) Falls K ein Stetigkeitspunkt von $\tilde{F}_0$ ist, ist Gleichung (2.3.4) erfüllt für alle $\gamma \in [0,1]$, zum Beispiel $\gamma = 0$.

b) Falls K kein Stetigkeitspunkt von $\tilde{F}_0$ ist, dann ist $\tilde{F}_0(K) - \tilde{F}_0(K-) > 0$, woraus folgt

$$\gamma = \frac{\alpha - 1 + \tilde{F}_0(K)}{\tilde{F}_0(K) - \tilde{F}_0(K-)}$$

$\Rightarrow$ es gibt einen Neyman-Pearson-Test zum Niveau α.

2. Wir definieren $M^{\neq} = \{x \in B : \varphi(x) \neq \varphi_K(x)\}$. Es muss gezeigt werden, daß

$$\mu\left((K_0 \cup K_1) \cap M^{\neq}\right) = 0.$$

Dazu betrachten wir

$$\begin{aligned} &\mathbb{E}_1\, \varphi(X_1,\dots,X_n) - \mathbb{E}_1\, \varphi_K(X_1,\dots,X_n) = 0 \quad (\varphi \text{ und } \varphi_K \text{ sind beste Tests}) \\ &\mathbb{E}_0\, \varphi(X_1,\dots,X_n) - \mathbb{E}_0\, \varphi_K(X_1,\dots,X_n) \le 0 \quad (\varphi \text{ und } \varphi_K \text{ sind } \alpha\text{-Tests} \\ &\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{mit Umfang von } \varphi_K = \alpha) \end{aligned}$$

$$\Rightarrow \int_B (\varphi - \varphi_K)) \cdot (f_1 - K \cdot f_0)\, \mu(dx) \ge 0.$$

In Bemerkung 2.3.3 wurde bewiesen, daß

$$\begin{aligned} &\int_B (\varphi - \varphi_K)(f_1 - K_0) d\mu \le 0 \\ &\Rightarrow \int_B (\varphi - \varphi_K)(f_1 - K f_0) d\mu = 0 = \int\limits_{M^{\neq} \cap (K_0 \cup K_1)} (\varphi - \varphi_K)(f_1 - K f_0) d\mu. \end{aligned}$$

Es gilt $\mu(M^{\neq} \cap (K_0 \cup K_1)) = 0$, falls der Integrand $(\varphi_K - \varphi)(f_1 - Kf_0) > 0$ auf $M^{\neq}$ ist. Wir zeigen, daß

$$(\varphi_K - \varphi)(f_1 - Kf_0) > 0 \text{ für } x \in M^{\neq} \tag{2.3.5}$$

ist. Es gilt

$$\begin{aligned} f_1 - Kf_0 > 0 &\Rightarrow \varphi_K - \varphi > 0, \\ f_1 - Kf_0 < 0 &\Rightarrow \varphi_K - \varphi < 0, \end{aligned}$$

weil

$$\begin{aligned} f_1(x) > Kf_0(x) &\Rightarrow \varphi_K(x) = 1 \\ &\qquad \text{und mit } \varphi(x) < 1 \Rightarrow \varphi_K(x) - \varphi(x) > 0 \text{ auf } M^{\neq}. \\ f_1(x) < Kf_0(x) &\Rightarrow \varphi_K(x) = 0 \\ &\qquad \text{und mit } \varphi(x) > 0 \Rightarrow \varphi_K(x) - \varphi(x) < 0 \text{ auf } M^{\neq}. \end{aligned}$$

Daraus folgt die Gültigkeit der Ungleichung (2.3.5) und somit

$$\mu\left((K_0 \cup K_1) \cap M^{\neq}\right) = 0.$$

□

Bemerkung 2.3.4. Falls φ und φ_K beste α-Tests sind, dann sind sie P_0- bzw. P_1- fast sicher gleich.

Beispiel 2.3.1 (Neyman-Pearson-Test für den Parameter der Poissonverteilung)**.** Es sei $(X_1, \ldots, X_n)$ eine Zufallsstichprobe mit $X_i \sim \text{Poisson}(\lambda)$, $\lambda > 0$, wobei X_i unabhängig und identisch verteilt sind für $i = 1, \ldots, n$. Wir testen die Hypothesen $H_0 : \lambda = \lambda_0$ vs. $H_1 : \lambda = \lambda_1$. Dabei ist

$$\begin{aligned} g_i(x) &= e^{-\lambda_i} \frac{\lambda_i^x}{x!}, \ x \in \mathbb{N}_0, \ i = 0, 1, \\ f_i(x) &= f_i(x_1, \ldots, x_n) = \prod_{j=1}^n g_i(x_j) = \prod_{j=1}^n e^{-\lambda_i} \frac{\lambda_i^{x_j}}{x_j!} = e^{-n\lambda_1} \cdot \frac{\lambda_1^{\sum_{j=1}^n x_j}}{(x_1! \cdot \ldots \cdot x_n!)} \end{aligned}$$

für $i = 0, 1$. Die Neyman-Pearson-Teststatistik ist

$$T(x_1, \ldots, x_n) = \begin{cases} \frac{f_1(x)}{f_0(x)} = e^{-n(\lambda_1 - \lambda_0)} \cdot (\lambda_1/\lambda_0)^{\sum_{j=1}^n x_j}, & \text{falls } x_1, \ldots, x_n \in \mathbb{N}_0, \\ \infty, & \text{sonst.} \end{cases}.$$

Die Neyman-Pearson-Entscheidungsregel lautet

$$\varphi_K(x_1, \ldots, x_n) = \begin{cases} 1, & \text{falls } T(x_1, \ldots, x_n) > K, \\ \gamma, & \text{falls } T(x_1, \ldots, x_n) = K, \\ 0, & \text{falls } T(x_1, \ldots, x_n) < K. \end{cases}$$

Wir wählen $K > 0$, $\gamma \in [0,1]$, sodaß φ_K den Umfang α hat. Dazu lösen wir

$$\alpha = \mathbb{P}_0(T(X_1,\dots,X_n) > K) + \gamma P_0(T(X_1,\dots,X_n) = K)$$

bezüglich γ und K auf.

$$\begin{aligned}
\mathbb{P}_0(T(X_1,\dots,X_n) > K) &= \mathbb{P}_0(\log T(X_1,\dots,X_n) > \log K) \\
&= \mathbb{P}_0\left(-n(\lambda_1-\lambda_0) + \sum_{j=1}^n X_j \cdot \log\left(\frac{\lambda_1}{\lambda_0}\right) > \log K\right) \quad = \mathbb{P}_0\left(\sum_{j=1}^n X_j > A\right) \\
\text{wobei } A &:= \left\lfloor \frac{\log K + n\cdot(\lambda_1-\lambda_0)}{\log\frac{\lambda_1}{\lambda_0}} \right\rfloor,
\end{aligned}$$

falls zum Beispiel $\lambda_1 > \lambda_0$. Im Falle $\lambda_1 < \lambda_0$ ändert sich das $>$ auf $<$ in der Wahrscheinlichkeit.

Wegen der Faltungsstabilität der Poissonverteilung ist unter H_0

$$\sum_{j=1}^n X_i \sim \text{Poisson}(n\lambda_0),$$

also wählen wir K als minimale, nichtnegative Zahl, für die gilt: $\mathbb{P}_0\left(\sum_{j=1}^n X_j > A\right) \leq \alpha$, und setzen

$$\gamma = \frac{\alpha - \mathbb{P}_0(\sum_{j=1}^n X_j > A)}{\mathbb{P}_0(\sum_{j=1}^n X_j = A)},$$

wobei

$$\begin{aligned}
\mathbb{P}_0\left(\sum_{j=1}^n X_j > A\right) &= 1 - \sum_{j=0}^{A} e^{-\lambda_0 n}\frac{(\lambda_0 n)^j}{j!} \\
\mathbb{P}_0\left(\sum_{j=1}^n X_j = A\right) &= e^{-\lambda_0 n}\frac{(\lambda_0 n)^A}{A!}
\end{aligned}$$

Somit haben wir die Parameter K und γ gefunden und damit einen Neyman-Pearson-Test φ_K konstruiert.

2.3.3 Einseitige Neyman-Pearson-Tests

Bisher betrachteten wir Neyman-Pearson-Tests für einfache Hypothesen der Form $H_i : \theta = \theta_i$, $i = 0, 1$. In diesem Abschnitt wollen wir einseitige Neyman-Pearson-Tests einführen, für Hypothesen der Form $H_0 : \theta \leq \theta_0$ vs. $H_1 : \theta > \theta_0$.

Zunächst konstruieren wir einen Test für diese Hypothesen: Sei $(X_1,\dots,X_n)$ eine Zufallsstichprobe, X_i seien unabhängig und identisch verteilt mit

$$X_i \sim F_\theta \in \Lambda = \{F_\theta : \theta \in \Theta\},$$

wobei $\Theta \subset \mathbb{R}$ offen ist und Λ eindeutig parametrisiert, das heißt

$$\theta \neq \theta' \Rightarrow F_\theta \neq F_{\theta'}.$$

Ferner besitze F_θ eine Dichte g_θ bezüglich des Lebesgue-Maßes (bzw. Zählmaßes) auf $\mathbb{R}$ (bzw. $\mathbb{N}_0$). Dann ist

$$f_\theta(x) = \prod_{j=1}^{n} g_\theta(x_i), \quad x = (x_1, \ldots, x_n)$$

eine Dichte von $(X_1, \ldots, X_n)$ bezüglich μ auf B.

Definition 2.3.3. Eine Verteilung auf B mit Dichte f_θ gehört zur Klasse von *Verteilungen mit monotonen Dichtekoeffizienten* in T, falls es für alle $\theta < \theta'$ eine Funktion $h : \mathbb{R} \times \Theta^2 \to \mathbb{R} \cup \infty$, die monoton wachsend in $t \in \mathbb{R}$ ist und eine Statistik $T : B \to \mathbb{R}$ gibt, mit der Eigenschaft

$$\frac{f_{\theta'}(x)}{f_\theta(x)} = h(T(x), \theta, \theta'),$$

wobei

$$h(T(x), \theta, \theta') = \infty \quad \text{für alle } x \in B: \ f_\theta(x) = 0, \ f_{\theta'}(x) > 0.$$

Der Fall $f_\theta(x) = f_{\theta'}(x) = 0$ tritt mit $\mathbb{P}_0$- bzw. $\mathbb{P}_1$-Wahrscheinlichkeit 0 auf.

Definition 2.3.4. Es sei Q_θ eine Verteilung auf $(B, \mathcal{B})$ mit der Dichte f_θ bzgl. μ. Q_θ gehört zur *einparametrischen Exponentialklasse* ($\theta \in \Theta \subset \mathbb{R}$ offen), falls die Dichte folgende Form hat:

$$f_\theta(x) = \exp\{c(\theta) \cdot T(x) + a(\theta)\} \cdot l(x), \quad x = (x_1, \ldots, x_n) \in B,$$

wobei $c(\theta)$ eine monoton steigende Funktion ist, und $\operatorname{Var}_\theta T(X_1, \ldots, X_n) > 0$, $\forall \theta \in \Theta$.

Lemma 2.3.1. Verteilungen aus der einparametrischen Exponentialfamilie besitzen einen monotonen Dichtekoeffizienten.

Beweis. Es sei Q_θ aus der einparametrischen Exponentialfamilie mit der Dichte

$$f_\theta(x) = \exp\{c(\theta) \cdot T(x) + a(\theta)\} \cdot l(x).$$

Für $\theta < \theta'$ ist dann

$$\frac{f_{\theta'}(x)}{f_\theta(x)} = \exp\{(c(\theta') - c(\theta)) \cdot T(x) + a(\theta') - a(\theta)\}$$

monoton bezüglich T, weil $c(\theta') - c(\theta) > 0$ wegen der Monotonie von $c(\theta)$. Also besitzt f_θ einen monotonen Dichtekoeffizienten. □

Beispiel 2.3.2. 1. *Normalverteilte Stichprobenvariablen*

Es seien $X_i \sim N(\mu, \sigma_0^2)$, $i = 1, \ldots, n$, unabhängige, identisch verteile Zufallsvariablen, mit unbekanntem Parameter μ und bekannter Varianz σ_0^2. Die Dichte des Zufallsvektors $X = (X_1, \ldots, X_n)^\top$ ist gleich

$$\begin{aligned}
f_\mu(x) &= \prod_{i=1}^n g_\mu(x_i) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-\frac{(x_i-\mu)^2}{2\sigma_0^2}} \\
&= \frac{1}{(2\pi\sigma_0^2)^{n/2}} \exp\left\{-\frac{1}{2\sigma_0^2}\sum_{i=1}^n (x_i-\mu)^2\right\} \\
&= \frac{1}{(2\pi\sigma_0^2)^{n/2}} \exp\left\{-\frac{1}{2\sigma_0^2}\left(\sum_{i=1}^n x_i^2 - 2\mu\sum_{i=1}^n x_i + \mu^2 n\right)\right\} \\
&= \exp\Big(\underbrace{\frac{\mu}{\sigma_0^2}}_{c(\mu)} \cdot \underbrace{\sum_{i=1}^n x_i}_{t(x)} - \underbrace{\frac{\mu^2 n}{2\sigma_0^2}}_{a(\mu)}\Big) \cdot \underbrace{\frac{1}{(2\pi\sigma_0^2)^{n/2}} \exp\left(-\frac{\sum_{i=1}^n x_i^2}{2\sigma_0^2}\right)}_{l(x)}.
\end{aligned}$$

Also gehört $N(\mu, \sigma_0^2)$ zur einparametrischen Exponentialklasse mit $c(\mu) = \frac{\mu}{\sigma_0^2}$ und $T(x) = \sum_{i=1}^n x_i$.

2. *Binomialverteilte Stichprobenvariablen*

Es seien $X_i \sim \text{Bin}(k, p)$ unabhängig und identisch verteilt, $i = 1, \ldots, n$. Der Parameter p sei unbekannt. Die Zähldichte des Zufallsvektors $X = (X_1, \ldots, X_n)^\top$ ist

$$\begin{aligned}
f_p(x) &= \mathbb{P}_p\left(X_i = x_i, i = 1, \ldots, n\right) \\
&= \prod_{i=1}^n \binom{k}{x_i} p^{x_i}(1-p)^{k-x_i} = p^{\sum_{i=1}^n x_i} \cdot \frac{(1-p)^{nk}}{(1-p)^{\sum_{i=1}^n x_i}} \cdot \prod_{i=1}^n \binom{k}{x_i} \\
&= \exp\Big\{\Big(\underbrace{\sum_{i=1}^n x_i}_{T(x)}\Big) \cdot \underbrace{\log\left(\frac{p}{1-p}\right)}_{c(p)} + \underbrace{nk \cdot \log(1-p)}_{a(p)}\Big\} \cdot \underbrace{\prod_{i=1}^n \binom{k}{x_i}}_{l(x)},
\end{aligned}$$

also gehört $\text{Bin}(n, p)$ zur einparametrischen Exponentialklasse mit

$$c(p) = \log\left(\frac{p}{1-p}\right)$$

und

$$T(x) = \sum_{i=1}^n x_i.$$

Lemma 2.3.2. Falls φ_K der Neyman-Pearson-Test der Hypothesen $H_0 : \theta = \theta_0$ vs. $H_1 : \theta = \theta_1$ ist, dann gilt:

$$\mu(\underbrace{\{x \in B : \ f_1(x) \neq K f_0(x)\}}_{K_0 \cup K_1}) > 0.$$

Beweis. Wegen $\theta_0 \neq \theta_1$ und der eindeutigen Parametrisierung gilt $f_0 \neq f_1$ auf einer Menge mit μ-Maß > 0.

Nun sei $\mu(K_0 \cup K_1) = 0$. Daraus folgt, daß $f_1(x) = K \cdot f_0(x)$ μ-fast sicher. Das heißt

$$1 = \int_B f_1(x)dx = K \cdot \int_B f_0(x)dx,$$

woraus folgt, daß $K = 1$ und $f_1(x) = f_0(x)$ μ-fast sicher, was aber ein Widerspruch zur eindeutigen Parametrisierung ist. $\square$

Im Folgenden sei $(X_1, \ldots, X_n)$ eine Stichprobe von unabhängigen, identisch verteilten Zufallsvariablen mit $X_i \sim$ Dichte g_θ, $i = 1, \ldots, n$ und

$$(X_1, \ldots, X_n) \sim \text{ Dichte } f_\theta(x) = \prod_{i=1}^n g_\theta(x_i)$$

aus der Klasse der Verteilungen mit monotonen Dichtekoeffizienten und einer Statistik $T(X_1, \ldots, X_n)$.

Wir betrachten die Hypothesen $H_0 : \theta \leq \theta_0$ vs. $H_1 : \theta > \theta_0$ und den Neyman-Pearson-Test:

$$\varphi_{K^*}(x) = \begin{cases} 1, & \text{falls } T(x) > K^*, \\ \gamma^*, & \text{falls } T(x) = K^*, \\ 0, & \text{falls } T(x) < K^* \end{cases} \tag{2.3.6}$$

für $K^* \in \mathbb{R}$ und $\gamma^* \in [0,1]$. Die Gütefunktion von φ_{K^*} bei θ_0 ist

$$G_n(\theta_0) = \mathbb{E}_0\, \varphi_{K^*} = \mathbb{P}_0\left(T(X_1, \ldots, X_n) > K^*\right) + \gamma^* \cdot \mathbb{P}_0\left(T(X_1, \ldots, X_n) = K^*\right)$$

Satz 2.3.3.

1. Falls $\alpha = \mathbb{E}_0\, \varphi_{K^*} > 0$, dann ist der soeben definierte Neyman-Pearson-Test ein bester Test der einseitigen Hypothesen H_0 vs. H_1 zum Niveau α.
2. Zu jedem Konfidenzniveau $\alpha \in (0,1)$ gibt es ein $K^* \in \mathbb{R}$ und $\gamma^* \in [0,1]$, sodaß φ_{K^*} ein bester Test zum Umfang α ist.
3. Die Gütefunktion $G_n(\theta)$ von $\varphi_{K^*}(\theta)$ ist monoton wachsend in θ. Falls $0 < G_n(\theta) < 1$, dann ist sie sogar streng monoton wachsend.

Beweis.

1. Wähle $\theta_1 > \theta_0$ und betrachte die einfachen Hypothesen $H_0' : \theta = \theta_0$ und $H_1' : \theta = \theta_1$. Sei

$$\varphi_K(x) = \begin{cases} 1, & f_1(x) > K f_0(x), \\ \gamma, & f_1(x) = K f_0(x), \\ 0, & f_1(x) < K f_0(x) \end{cases}$$

der Neyman-Pearson-Test für H_0', H_1' mit $K > 0$. Da f_θ den monotonen Dichtekoeffizienten mit Statistik T besitzt,

$$\frac{f_1(x)}{f_0(x)} = h(T(x), \theta_0, \theta_1),$$

gilt

$$\left\{x : f_1(x)/f_0(x) \begin{array}{c} > K \\ < K \end{array}\right\} = \left\{T(x) \begin{array}{c} > K^* \\ < K^* \end{array}\right\} \qquad \text{mit } K = h(K^*, \theta_0, \theta_1).$$

φ_K ist ein bester Neyman-Pearson-Test zum Niveau $\alpha = \mathbb{E}_0\, \varphi_K = \mathbb{E}_0\, \varphi_{K^*}$. Aus $\alpha > 0$ folgt $K < \infty$, denn aus $K = \infty$ würde folgen

$$\begin{aligned} 0 < \alpha = \mathbb{E}_0\, \varphi_K &\le \mathbb{P}_0\left(T(X_1, \ldots, X_n) \ge K^*\right) \le \mathbb{P}_0\left(\frac{f_1(X_1, \ldots, X_n)}{f_0(X_1, \ldots, X_n)} = \infty\right) \\ &= \mathbb{P}_0\left(f_1(X_1, \ldots, X_n) > 0, f_0(X_1, \ldots, X_n) = 0\right) \\ &= \int_B \mathbb{I}\left(f_1(x) > 0, f_0(x) = 0\right) \cdot f_0(x)\mu(dx) = 0. \end{aligned}$$

Für den Test φ_{K^*} aus (2.3.6) gilt dann

$$\varphi^*(x) = \begin{cases} 1, & \text{falls } f_1(x)/f_0(x) > K, \\ \gamma^*(x), & \text{falls } f_1(x)/f_0(x) = K, \\ 0, & \text{falls } f_1(x)/f_0(x) < K, \end{cases}$$

wobei $\gamma^*(x) \in \{\gamma^*, 0, 1\}$. Man kann zeigen, daß $\varphi^* = \varphi_{K^*}$. Daraus folgt, daß φ_{K^*} ein bester Neyman-Pearson-Test ist, für H_0' vs. H_1' (vergleiche Bemerkung 2.3.2, 1.) und Bemerkung 2.3.3) für beliebige $\theta_1 > \theta_0$. Daraus folgt, daß φ_{K^*} ein bester Neyman-Pearson-Test für $H_0'' : \theta = \theta_0$ vs. $H_1'' : \theta > \theta_0$ ist.

Die selbe Behauptung erhalten wir aus dem Teil 3. des Satzes für $H_0 : \theta \le \theta_0$ vs. $H_1 : \theta > \theta_0$, weil dann $G_n(\theta) \le G_n(\theta_0) = \alpha$ für alle $\theta < \theta_0$.

2. Siehe Beweis zu Satz 2.3.2, 1.).

3. Wir müssen zeigen, daß $G_n(\theta)$ monoton ist. Dazu wählen wir $\theta_1 < \theta_2$ und zeigen, daß $\alpha_1 = G_n(\theta_1) \le G_n(\theta_2)$. Wir betrachten die neuen, einfachen Hypothesen $H_0'' : \theta = \theta_1$ vs. $H_1'' : \theta = \theta_2$. Der Test φ_{K^*} kann genauso wie in 1. als Neyman-Pearson-Test φ^* dargestellt werden (für die Hypothesen H_0'' und H_1''), der ein bester Test zum Niveau α_1 ist. Betrachten wir einen weiteren konstanten Test $\varphi(x) = \alpha_1$. Dann ist $\alpha_1 = \mathbb{E}_{\theta_2}\, \varphi \le \mathbb{E}_{\theta_2}\, \varphi_{K^*} = G_n(\theta_2)$. Daraus folgt, daß $G_n(\theta_1) \le G_n(\theta_2)$.

   Nun zeigen wir, daß für $G_n(\theta) \in (0, 1)$ gilt: $G_n(\theta_1) < G_n(\theta_2)$. Wir nehmen an, daß $\alpha_1 = G_n(\theta_1) = G_n(\theta_2)$ und $\theta_1 < \theta_2$ für $\alpha \in (0, 1)$. Es folgt, daß $\varphi(x) = \alpha_1$ auch ein bester Test für H_0'' und H_1'' ist. Aus Satz2.3.2, 2.) folgt

$$\mu(\{x \in B : \underbrace{\varphi(x)}_{=\alpha_1} \ne \varphi_{K^*}(x)\}) = 0 \text{ auf } K_0 \cup K_1 = \{f_1(x) \ne K f_0(x)\},$$

was ein Widerspruch zur Bauart des Tests φ_{K^*} ist, der auf $K_0 \cup K_1$ nicht gleich $\alpha_1 \in (0,1)$ sein kann.

□

Bemerkung 2.3.5. 1. Der Satz 2.3.3 ist genauso auf Neyman-Pearson-Tests der einseitigen Hypothesen

$$H_0 : \theta \geq \theta_0 \text{ vs. } H_1 : \theta < \theta_0$$

anwendbar, mit dem entsprechenden Unterschied

$$\theta \mapsto -\theta$$
$$T \mapsto -T$$

Somit existiert der beste α-Test auch in diesem Fall.

2. Man kann zeigen, daß die Gütefunktion $G_n(\varphi_{K^*}, \theta)$ des besten Neyman-Pearson-Tests auf $\Theta_0 = (-\infty, \theta_0)$ folgende Minimalitätseigenschaft besitzt:

$$G_n(\varphi_{K^*}, \theta) \leq G_n(\varphi, \theta) \quad \forall \varphi \in \Psi(\alpha),\ \theta \leq \theta_0$$

Beispiel 2.3.3. Wir betrachten eine normalverteilte Stichprobe $(X_1, \ldots, X_n)$ von unabhängigen und identisch verteilten Zufallsvariablen X_i, wobei $X_i \sim N(\mu, \sigma_0^2)$ und σ_0^2 sei bekannt. Es werden die Hypothesen

$$H_0 : \mu \leq \mu_0 \text{ vs. } H_1 : \mu > \mu_0,$$

getestet. Aus Beispiel 2.1.2 kennen wir die Testgröße

$$T(X_1, \ldots, X_n) = \sqrt{n}\frac{\overline{X}_n - \mu_0}{\sigma_0},$$

wobei unter H_0 gilt: $T(X_1, \ldots, X_n) \sim N(0,1)$. H_0 wird verworfen, falls

$$T(X_1, \ldots, X_n) > z_{1-\alpha}, \quad \text{wobei } \alpha \in (0,1).$$

Wir zeigen jetzt, daß dieser Test der beste Neyman-Pearson-Test zum Niveau α ist. Aus Beispiel 2.3.2 ist bekannt, daß die Dichte f_n von $(X_1, \ldots, X_n)$ aus der einparametrischen Exponentialklasse ist, mit

$$\tilde{T}(X_1, \ldots, X_n) = \sum_{i=1}^{n} X_i.$$

Dann gehört f_μ von $(x_1, \ldots, x_n)$ zur einparametrischen Exponentialklasse auch bezüglich der Statistik

$$T(X_1, \ldots, X_n) = \sqrt{n}\frac{\overline{X}_n - \mu}{\sigma_0}$$

Es gilt nämlich

$$
\begin{aligned}
f_\mu(x) &= \exp\Big(\underbrace{\frac{\mu}{\sigma_0^2}}_{\tilde{c}(\mu)} \cdot \underbrace{\sum_{i=1}^n x_i}_{\tilde{T}} - \underbrace{\frac{\mu^2 n}{2\sigma_0^2}}_{\tilde{a}(\mu)} \Big) \cdot l(x) \\
&= \exp\Big(\underbrace{\frac{\mu\sqrt{n}}{\sigma_0}}_{c(\mu)} \cdot \underbrace{\sqrt{n}\frac{\overline{x}_n - \mu}{\sigma_0}}_{T} + \underbrace{\frac{\mu^2 n}{2\sigma_0^2}}_{a(\mu)} \Big) \cdot l(x).
\end{aligned}
$$

Die Statistik T kann also in der Konstruktion des Neyman-Pearson-Tests (Gleichung (2.3.6)) verwendet werden:

$$
\varphi_{K^*}(x) = \begin{cases} 1, & \text{falls } T(x) > z_{1-\alpha}, \\ 0, & \text{falls } T(x) = z_{1-\alpha}, \\ 0, & \text{falls } T(x) < z_{1-\alpha} \end{cases}
$$

(mit $K^* = z_{1-\alpha}$ und $\gamma^* = 0$). Nach Satz 2.3.3 ist dieser Test der beste Neyman-Pearson-Test zum Niveau α für unsere Hypothesen:

$$
\begin{aligned}
G_n(\varphi_{K^*}, \mu_0) &= \mathbb{P}_0\left(T(X_1, \ldots, X_n) > z_{1-\alpha}\right) + 0 \cdot \mathbb{P}_0\left(T(X_1, \ldots, X_n) \leq z_{1-\alpha}\right) \\
&= 1 - \Phi(z_{1-\alpha}) = 1 - (1-\alpha) = \alpha.
\end{aligned}
$$

2.3.4 Unverfälschte zweiseitige Tests

Es sei $(X_1, \ldots, X_n)$ eine Stichprobe von unabhängigen und identisch verteilten Zufallsvariablen mit der Dichte

$$
f_\theta(x) = \prod_{i=1}^n g_\theta(x_i).
$$

Es wird ein zweiseitiger Test der Hypothesen

$$
H_0 : \theta = \theta_0 \text{ vs. } H_1 : \theta \neq \theta_0
$$

betrachtet. Für alle $\alpha \in [0,1]$ kann es jedoch keinen besten Test φ zum Niveau α für H_0 vs. H_1 geben. Denn, nehmen wir an, φ wäre der beste Test zum Niveau α für H_0 vs. H_1. Dann wäre φ der beste Test für die Hypothesen

1. $H_0' : \theta = \theta_0$ vs. $H_1' : \theta > \theta_0$
2. $H_0'' : \theta = \theta_0$ vs. $H_1'' : \theta < \theta_0$.

Dann ist nach Satz 2.3.3 die Gütefunktion

1. $G_n(\varphi, \theta) < \alpha$ auf $\theta < \theta_0$, bzw.
2. $G_n(\varphi, \theta) > \alpha$ auf $\theta < \theta_0$,

was ein Widerspruch ist!

Darum werden wir die Klasse aller möglichen Tests auf unverfälschte Tests (Definition 2.1.5) eingrenzen. Der Test φ ist unverfälscht genau dann, wenn

$$G_n(\varphi,\theta) \le \alpha \text{ für } \theta \in \Theta_0$$
$$G_n(\varphi,\theta) \ge \alpha \text{ für } \theta \in \Theta_1$$

Beispiel 2.3.4. 1. $\varphi(x) \equiv \alpha$ ist unverfälscht.

2. Der zweiseitige Gauß-Test ist unverfälscht, vergleiche Beispiel 2.1.2: $G_n(\varphi,\mu) \ge \alpha$ für alle $\mu \in \mathbb{R}$.

Im Folgenden seien X_i unabhängig und identisch verteilt. Die Dichte f_θ des Zufallsvektors $(X_1,\dots,X_n)$ gehöre zur einparametrischen Exponentialklasse:

$$f_\theta(x) = \exp\{c(\theta)\cdot T(x) + a(\theta)\}\cdot l(x),$$

wobei $c(\theta)$ und $a(\theta)$ stetig differenzierbar auf Θ sein sollen, mit $c'(\theta) > 0$ und der Varianz $\mathrm{Var}_\theta\, T(X_1,\dots,X_n) > 0$ für alle $\theta \in \Theta$.

Übungsaufgabe 2.3.1. Zeigen Sie, daß folgende Relation gilt:

$$a'(\theta) = -c'(\theta)\mathbb{E}_\theta\, T(X_1,\dots,X_n),$$

falls $a(\theta)$ differenzierbar für alle $\theta \in \Theta$ ist.

Lemma 2.3.3. Es sei φ ein unverfälschter Test zum Niveau α für

$$H_0 : \theta = \theta_0 \text{ vs. } H_1 : \theta \neq \theta_0.$$

Dann gilt:

1. $\alpha = \mathbb{E}_0\, \varphi(X_1,\dots,X_n) = G_n(\varphi,\theta_0)$
2. $\mathbb{E}_0\, [T(X_1,\dots,X_n)\varphi(X_1,\dots,X_n)] = \alpha\cdot\mathbb{E}_0\, T(X_1,\dots,X_n)$

Beweis. 1. Die Gütefunktion von φ ist

$$G_n(\varphi,\theta) = \int_B \varphi(x) f_\theta(x)\mu(dx)$$

Da f_θ aus der einparametrischen Exponentialklasse ist, ist $G_n(\varphi,\theta)$ differenzierbar (unter dem Integral) bezüglich θ. Wegen der Unverfälschtheit von φ gilt

$$G_n(\varphi,\theta_0) \le \alpha,\ G_n(\varphi,\theta) \ge \alpha, \quad \theta \neq \theta_0$$

und daraus folgt $G_n(\varphi,\theta_0) = \alpha$ und θ_0 ist ein Minimumpunkt von G_n. Somit ist 1) bewiesen.

2. Da θ_0 der Minimumpunkt von G_n ist, gilt

$$\begin{aligned}
0 = G_n'(\varphi, \theta_0) &= \int_B \varphi(x)(c'(\theta_0)T(x) + a'(\theta_0)) f_0(x)\mu(dx) \\
&= c'(\theta_0) \cdot \mathbb{E}_0 \left[\varphi(X_1, \dots X_n)T(X_1, \dots, X_n)\right] + a'(\theta) \cdot G_n(\varphi, \theta_0) \\
&= c'(\theta_0) \cdot \mathbb{E}_0 \left[\varphi(X_1, \dots, X_n)T(X_1, \dots, X_n)\right] + \alpha a'(\theta_0) \\
&\overset{\text{(Übung 2.3.1)}}{=} c'(\theta_0) \left(\mathbb{E}_0 \left(\varphi \cdot T\right) - \alpha \mathbb{E}_0 T\right)
\end{aligned}$$

Daraus folgt $\mathbb{E}_0 (\varphi T) = \alpha \mathbb{E}_0 T$ und damit ist der Satz bewiesen.

□

Wir definieren jetzt die modifizierten Neyman-Pearson-Tests für einfache Hypothesen

$$H_0 : \theta = \theta_0 \text{ vs. } H_1' : \theta = \theta_1, \quad \theta_1 \neq \theta_0.$$

Für $\lambda, K \in \mathbb{R}$, $\gamma : B \to [0,1]$ definieren wir

$$\varphi_{K,\lambda}(x) = \begin{cases} 1, & \text{falls } f_1(x) > (K + \lambda T(x)) f_0(x), \\ \gamma(x), & \text{falls } f_1(x) = (K + \lambda T(x) f_0(x)), \\ 0, & \text{falls } f_1(x) < (K + \lambda T(x)) f_0(x), \end{cases} \tag{2.3.7}$$

wobei $T(x)$ aus der einparametrischen Exponentialklasse ist.

Es sei $\tilde{\Psi}(\alpha)$ die Klasse aller Tests, die Aussagen 1) und 2) des Lemmas 2.3.3 erfüllen. Aus Lemma 2.3.3 folgt dann, daß die Menge der unverfälschten Tests zum Niveau α eine Teilmenge von $\tilde{\Psi}(\alpha)$ ist.

Satz 2.3.4. Der modifizierte Neyman-Pearson-Test $\varphi_{K,\lambda}$ ist der beste α-Test in $\tilde{\Psi}(\alpha)$ für Hypothesen H_0 vs. H_1' zum Niveau $\alpha = \mathbb{E}_0\, \varphi_{K,\lambda}$, falls $\varphi_{K,\lambda} \in \tilde{\Psi}(\alpha)$.

Beweis. Es ist zu zeigen, daß $\mathbb{E}_1\, \varphi_{K,\lambda} \geq \mathbb{E}_1\, \varphi$ für alle $\varphi \in \tilde{\Psi}(\alpha)$, bzw. $\mathbb{E}_1 \left(\varphi_{K,\lambda} - \varphi\right) \geq 0$. Es gilt

$$\begin{aligned}
\mathbb{E}_1 \left(\varphi_{K,\lambda} - \varphi\right) &= \int_B (\varphi_{K,\lambda}(x) - \varphi(x)) f_1(x) \mu(dx) \\
&\overset{\text{(Bem. 2.3.3, 2.))}}{\geq} \int_B (\varphi_{K,\lambda}(x) - \varphi(x))(K + \lambda T(x)) f_0(x)\mu(dx) \\
&= K\Big(\underbrace{\mathbb{E}_0\, \varphi_{K,\lambda}}_{=\alpha} - \underbrace{\mathbb{E}_0\, \varphi}_{=\alpha}\Big) + \lambda\Big(\underbrace{\mathbb{E}_0 \left(\varphi_{K,\lambda} \cdot T\right)}_{\alpha \mathbb{E}_0 T} - \underbrace{\mathbb{E}_0 \left(\varphi \cdot T\right)}_{=\alpha \cdot \mathbb{E}_0 T}\Big) \\
&= 0,
\end{aligned}$$

weil $\varphi, \varphi_{K,\lambda} \in \tilde{\Psi}(\alpha)$. □

Wir definieren folgende Entscheidungsregel, die später zum Testen der zweiseitigen Hypothesen

$$H_0 : \theta = \theta_0 \text{ vs. } H_1 : \theta \neq \theta_0$$

verwendet wird:

$$\varphi_c(x) = \begin{cases} 1, & \text{falls } T(x) \notin (c_1, c_2), \\ \gamma_1, & \text{falls } T(x) = c_1, \\ \gamma_2, & \text{falls } T(x) = c_2, \\ 0, & \text{falls } T(x) \in (c_1, c_2), \end{cases} \tag{2.3.8}$$

für $c_1 \leq c_2 \in \mathbb{R}$, $\gamma_1, \gamma_2 \in [0,1]$ und die Statistik $T(x)$, $x = (x_1, \ldots, x_n) \in B$, die in der Dichte aus der einparametrischen Exponentialfamilie vorkommt. Zeigen wir, daß φ_c sich als modifizierter Neyman-Pearson-Test schreiben lässt.

Lemma 2.3.4. Es sei $(X_1, \ldots, X_n)$ eine Stichprobe von unabhängigen, identisch verteilten Zufallsvariablen mit gemeinsamer Dichte $f_\theta(x), x \in B$, die zur einparametrischen Exponentialfamilie gehört. Sei $T(x)$ die dazugehörige Statistik, die im Exponenten der Dichte f_θ vorkommt. Für beliebige reelle Zahlen $c_1 \leq c_2$, $\gamma_1, \gamma_2 \in [0,1]$ und Parameterwerte $\theta_0, \theta_1 \in \Theta : \theta_0 \neq \theta_1$ läßt sich der Test φ_c aus (2.3.8) als modifizierter Neyman-Pearson-Test $\varphi_{K,\lambda}$ aus (2.3.7) mit gegebenen $K, \lambda \in \mathbb{R}$, $\gamma(x) \in [0,1]$ schreiben.

Beweis. Für die Dichte

$$f_\theta(x) = \exp\{c(\theta)T(x) + a(\theta)\} \cdot l(x)$$

wird (wie immer) vorausgesetzt, daß $l(x) > 0$, $c'(x) > 0$ und $a'(x)$ existiert für $\theta \in \Theta$. Falls wir die Bezeichnung

$$f_{\theta_i}(x) = f_i(x), \quad i = 0, 1$$

verwenden, dann gilt

$$\frac{f_1(x)}{f_0(x)} = \exp\Big\{ \underbrace{(c(\theta_1) - c(\theta_0))}_{c} T(x) + \underbrace{a(\theta_1) - a(\theta_0)}_{a} \Big\},$$

und somit

$$\{x \in B : f_1(x) > (K + \lambda T(x))\, f_0(x)\} = \{x \in B : \exp(cT(x) + a) > K + \lambda T(x)\}.$$

Finden wir solche K und λ aus $\mathbb{R}$, für die die Gerade $K + \lambda t$, $t \in \mathbb{R}$ die konvexe Kurve $\exp\{ct + a\}$ genau an den Stellen c_1 und c_2 schneidet (falls $c_1 \neq c_2$) bzw. an der Stelle $t = c_1$ berührt (falls $c_1 = c_2$). Dies ist immer möglich, siehe Abbildung 2.1. Ferner setzen wir $\gamma(x) = \gamma_i$ für $\{x \in B : T(x) = c_i\}$. Insgesamt gilt dann

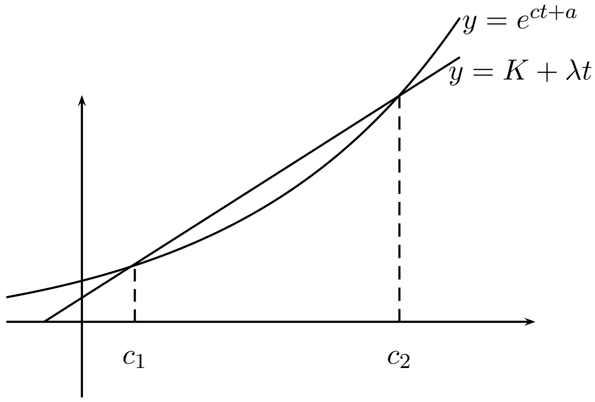
$$\{x : \exp(cT(x) + a) > K + \lambda T(x)\} = \{x : T(x) \notin [c_1, c_2]\}$$

und

$$\{x : \exp(cT(x) + a) < K + \lambda T(x)\} = \{x : T(x) \in (c_1, c_2)\}.$$

Damit ist das Lemma bewiesen. □

Abbildung 2.1:



Bemerkung 2.3.6. 1. Die Umkehrung des Lemmas stimmt nicht, denn bei vorgegebenen Kurven $y = K + \lambda t$ und $y = \exp\{ct + a\}$ muss es die Schnittpunkte c_1 und c_2 nicht unbedingt geben. So kann die Gerade vollständig unter der Kurve $y = \exp\{ct + a\}$ liegen.

2. Der Test φ_c macht von den Werten θ_0 und θ_1 nicht explizit Gebrauch. Dies unterscheidet ihn vom Test $\varphi_{K,\lambda}$, für den die Dichten f_0 und f_1 gebraucht werden.

Jetzt sind wir bereit, den Hauptsatz über zweiseitige Tests zum Prüfen der Hypothesen

$$H_0 : \theta = \theta_0 \text{ vs. } H_1 : \theta \neq \theta_0$$

zu formulieren und zu beweisen.

Satz 2.3.5 (Hauptsatz über zweiseitige Tests)**.** Unter den Voraussetzungen des Lemmas 2.3.4 sei φ_c ein Test aus (2.3.8), für den $\varphi_c \in \tilde{\Psi}(\alpha)$ gilt, $G_n(\varphi_c, \theta_0) = \alpha$. Dann ist φ_c bester unverfälschter Test zum Niveau α (und dadurch bester Test in $\tilde{\Psi}(\alpha)$) der Hypothesen

$$H_0 : \theta = \theta_0 \text{ vs. } H_1 : \theta \neq \theta_0$$

Beweis. Wählen wir ein beliebiges $\theta_1 \in \Theta$, $\theta_1 \neq \theta_0$. Nach Lemma 2.3.4 ist φ_c ein modifizierter Neyman-Pearson-Test $\varphi_{K,\lambda}$ für eine spezielle Wahl von K und $\lambda \in \mathbb{R}$. $\varphi_{K,\lambda}$ ist aber nach Satz 2.3.4 bester Test in $\tilde{\Psi}(\alpha)$ für $H_0 : \theta = \theta_0$ vs. $H_1 : \theta \neq \theta_0$. Da φ_c nicht von θ_1 abhängt, ist es bester Test in $\tilde{\Psi}(\alpha)$ für $H_1' : \theta \neq \theta_0$. Da aber unverfälschte Niveau-α-Tests in $\tilde{\Psi}(\alpha)$ liegen, müssen wir nur zeigen, daß φ_c unverfälscht ist. Da φ_c der beste Test ist, ist er nicht schlechter als der konstante unverfälschte Test $\varphi = \alpha$, das heißt

$$G_n(\varphi_c, \theta) \geq G_n(\varphi, \theta) = \alpha, \quad \theta \neq \theta_0.$$

Somit ist auch φ_c unverfälscht. Der Beweis ist beendet. □

Bemerkung 2.3.7. Wir haben gezeigt, daß φ_c der beste Test seines Umfangs ist. Es wäre jedoch noch zu zeigen, daß für beliebiges $\alpha \in (0,1)$ Konstanten $c_1, c_2, \gamma_1, \gamma_2$ gefunden werden, für die $\mathbb{E}_0\, \varphi_c = \alpha$ gilt. Da der Beweis schwierig ist, wird er hier ausgelassen. Im folgenden Beispiel jedoch wird es klar, wie die Parameter $c_1, c_2, \gamma_1, \gamma_2$ zu wählen sind.

Beispiel 2.3.5 (Zweiseitiger Gauß-Test)**.** Im Beispiel 2.1.2 haben wir folgenden Test des Erwartungswertes einer normalverteilten Stichprobe $(X_1, \ldots, X_n)$ mit unabhängigen und identisch verteilten X_i und $X_i \sim N(\mu, \sigma_0^2)$ bei bekannten Varianzen σ_0^2 betrachtet. Getestet werden die Hypothesen

$$H_0 : \mu = \mu_0 \text{ vs. } H_1 : \mu \neq \mu_0.$$

Der Test $\varphi(x)$ lautet

$$\varphi(x) = \mathbb{I}\left(x \in \mathbb{R}^n : |T(x)| > z_{1-\alpha/2}\right),$$

wobei

$$T(x) = \sqrt{n}\frac{\overline{x}_n - \mu_0}{\sigma_0}.$$

Zeigen wir, daß φ der beste Test zum Niveau α in $\tilde{\Psi}(\alpha)$ (und somit bester unverfälschter Test) ist. Nach Satz 2.3.5 müssen wir lediglich prüfen, daß φ als φ_c mit (2.3.8) dargestellt werden kann, weil die n-dimensionale Normalverteilung mit Dichte f_μ (siehe Beispiel 2.3.3) zu der einparametrischen Exponentialfamilie mit Statistik

$$T(x) = \sqrt{n}\frac{\overline{x}_n - \mu}{\sigma_0}$$

gehört. Setzen wir $c_1 = z_{1-\alpha/2}$, $c_2 = -z_{1-\alpha/2}$, $\gamma_1 = \gamma_2 = 0$. Damit ist

$$\varphi(x) = \varphi_c(x) = \begin{cases} 1, & \text{falls } |T(x)| > z_{1-\alpha/2}, \\ 0, & \text{falls } |T(x)| \leq z_{1-\alpha/2}. \end{cases}$$

und die Behauptung ist bewiesen, weil aus der in Beispiel 2.1.2 ermittelten Gütefunktion $G_n(\varphi, \theta)$ von φ ersichtlich ist, daß φ ein unverfälschter Test zum Niveau α ist (und somit $\varphi \in \tilde{\Psi}(\alpha)$).

Bemerkung 2.3.8. Bisher haben wir immer vorausgesetzt, daß nur *ein* Parameter der Verteilung der Stichprobe $(X_1, \ldots, X_n)$ unbekannt ist, um die Theorie des Abschnittes 2.3 über die besten (Neyman-Pearson-) Tests im Fall der einparametrischen Exponentialfamilie aufstellen zu können. Um jedoch den Fall weiterer unbekannten Parameter betrachten zu können (wie im Beispiel der zweiseitigen Tests des Erwartungswertes der normalverteilten Stichprobe bei unbekannter Varianz (der sog. t-Test, vergleiche Abschnitt 2.2.1, 1 (a))), bedarf es einer tiefergehenderen Theorie, die aus Zeitgründen in dieser Vorlesung nicht behandelt wird. Der interessierte Leser findet das Material dann im Buch [12].

2.4 Anpassungstests

Sei eine Stichprobe von unabhängigen, identisch verteilten Zufallsvariablen $(X_1, \ldots, X_n)$ gegeben mit $X_i \sim F$ (Verteilungsfunktion) für $i = 1, \ldots, n$. Bei den Anpassungstests wird die Hypothese

$$H_0 : F = F_0 \text{ vs. } H_1 : F \neq F_0$$

überprüft, wobei F_0 eine vorgegebene Verteilungsfunktion ist.

Einen Test aus dieser Klasse haben wir bereits in der Vorlesung Statistik I kennengelernt: den Kolmogorow-Smirnov-Test (vergleiche Bemerkung 3.3.8. 3), Vorlesungsskript Statistik I).

Jetzt werden weitere nichtparametrische Anpassungstests eingeführt. Der erste ist der χ^2-Anpassungstest von K. Pearson.

2.4.1 χ^2-Anpassungstest

Der Test von Kolmogorov-Smirnov basierte auf dem Abstand

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

zwischen der empirischen Verteilungsfunktion der Stichprobe $(X_1, \ldots, X_n)$ und der Verteilungsfunktion F_0. In der Praxis jedoch erscheint dieser Test zu feinfühlig, denn er ist zu sensibel gegenüber Unregelmäßigkeiten in den Stichproben und verwirft H_0 zu oft. Einen Ausweg aus dieser Situation stellt die Vergröberung der Haupthypothese H_0 dar, auf welcher der folgende χ^2-Anpassungstest beruht.

Man zerlegt den Wertebereich der Stichprobenvariablen X_i in r Klassen $(a_i, b_i]$, $i = 1, \ldots, n$ mit der Eigenschaft

$$-\infty \leq a_1 < b_1 = a_2 < b_2 = \ldots = a_r < b_r \leq \infty.$$

Anstelle von $X_i, i = 1, \ldots, n$ betrachten wir die sogenannten *Klassenstärken* Z_j, $j = 1, \ldots, r$, wobei

$$Z_j = \#\{i : a_j < X_i \leq b_j,\ 1 \leq i \leq n\}.$$

Lemma 2.4.1. Der Zufallsvektor $Z = (Z_1, \ldots, Z_n)^\top$ ist *multinomialverteilt* mit Parametervektor

$$p = (p_1, \ldots, p_{r-1})^\top \in [0, 1]^{r-1},$$

wobei

$$p_j = \mathbb{P}(a_j < X_1 \leq b_j) = F(b_j) - F(a_j), \quad j = 1, \ldots, r-1, \quad p_r = 1 - \sum_{j=1}^{r-1} p_j.$$

Schreibweise:

$$Z \sim M_{r-1}(n, p)$$

Beweis. Es ist zu zeigen, daß für alle Zahlen $k_1, \dots k_r \in \mathbb{N}_0$ mit $k_1 + \dots + k_r = n$ gilt:

$$\mathbb{P}(Z_i = k_i,\, i = 1, \dots, r) = \frac{n!}{k_1! \cdot \ldots \cdot k_r!} p_1^{k_1} \cdot \ldots \cdot p_r^{k_r}. \tag{2.4.1}$$

Da X_i unabhängig und identisch verteilt sind, gilt

$$\mathbb{P}\left(X_j \in (a_{i_j}, b_{i_j}],\, j = 1, \dots, n\right) = \prod_{j=1}^{n} \mathbb{P}\left(a_{i_j} < X_1 \le b_{i_j}\right) = p_1^{k_1} \cdot \ldots \cdot p_r^{k_r},$$

falls die Folge von Intervallen $(a_{i_j}, b_{i_j}]_{j=1,\dots,n}$ das Intervall $(a_i, b_i]$ k_i Mal enthält, $i = 1, \dots, r$. Die Formel (2.4.1) ergibt sich aus dem Satz der totalen Wahrscheinlichkeit als Summe über die Permutationen von Folgen $(a_{i_j}, b_{i_j}]_{j=1,\dots,n}$ dieser Art. □

Im Sinne des Lemmas 2.4.1 werden neue Hypothesen über die Beschaffenheit von F geprüft.

$$H_0 : p = p_0 \text{ vs. } H_1 : p \neq p_0,$$

wobei $p = (p_1, \dots, p_{r-1})^\top$ der Parametervektor der Multinomialverteilung von Z ist, und $p_0 = (p_{01}, \dots, p_{0,r-1})^\top \in (0,1)^{r-1}$ mit $\sum\limits_{i=1}^{r-1} p_{0i} < 1$. In diesem Fall ist

$$\Lambda_0 = \{F \in \Lambda : F(b_j) - F(a_j) = p_{0j}, \quad j = 1, \dots, r-1\},$$

$\Lambda_1 = \Lambda \setminus \Lambda_0$, wobei Λ die Menge aller Verteilungsfunktionen ist. Um H_0 vs. H_1 zu testen, führen wir die *Pearson-Teststatistik*

$$T_n(x) = \sum_{j=1}^{r} \frac{(z_j - np_{0j})^2}{np_{0j}}$$

ein, wobei $x = (x_1, \dots, x_n)$ eine konkrete Stichprobe der Daten ist und z_j, $j = 1, \dots, r$ ihre Klassenstärken sind.

Unter H_0 gilt

$$\mathbb{E}\, Z_j = np_{0j}, \quad j = 1, \dots, r,$$

somit soll H_0 abgelehnt werden, falls $T_n(X)$ ungewöhnlich große Werte annimmt.

Im nächsten Satz zeigen wir, daß $T(X_1, \dots, X_n)$ asymptotisch (für $n \to \infty$) χ^2_{r-1}-verteilt ist, was zu folgendem Anpassungstest (χ^2-Anpassungstest) führt:

$$H_0 \text{ wird verworfen, falls } T_n(x_1, \dots, x_n) > \chi^2_{r-1,1-\alpha}.$$

Dieser Test ist nach seinem Entdecker *Karl Pearson* (1857-1936) benannt worden.

Satz 2.4.1. Unter H_0 gilt

$$\lim_{n\to\infty} \mathbb{P}_{p_0}\left(T_n(X_1, \dots, X_n) > \chi^2_{r-1,1-\alpha}\right) = \alpha,\ \alpha \in (0,1),$$

das heißt, der χ^2-Pearson-Test ist ein asymptotischer Test zum Niveau α.

Beweis. Führen wir die Bezeichnung $Z_{nj} = Z_j(X_1, \ldots, X_n)$ der Klassenstärken ein, die aus der Stichprobe $(X_1, \ldots, X_n)$ entstehen. Nach Lemma 2.4.1 ist

$$Z_n = (Z_{n1}, \ldots, Z_{nr}) \sim M_{r-1}(n, p_0) \text{ unter } H_0.$$

Insbesondere soll $\mathbb{E}\, Z_{nj} = np_{0j}$ und

$$\mathrm{Cov}(Z_{ni}, Z_{nj}) = \begin{cases} np_{0j}(1 - p_{0j}), & i = j, \\ -np_{0i}p_{0j}, & i \neq j \end{cases}$$

für alle $i, j = 1, \ldots, r$ gelten. Da

$$Z_{nj} = \sum_{i=1}^{n} \mathbb{I}(a_j < X_j \leq b_j), \quad j = 1, \ldots, r,$$

ist $Z_n = (Z_{n1}, \ldots, Z_{n,r-1})$ eine Summe von n unabhängigen und identisch verteilten Zufallsvektoren $Y_i \in \mathbb{R}^{r-1}$ mit Koordinaten $Y_{ij} = \mathbb{I}(a_j < X_i \leq b_j)$, $j = 1, \ldots, r-1$. Daher gilt nach dem multivariaten Grenzwertsatz (der in Lemma 2.4.2 bewiesen wird), daß

$$Z'_n = \frac{Z_n - \mathbb{E}\, Z_n}{\sqrt{n}} = \frac{\sum\limits_{i=1}^{n} Y_i - n\mathbb{E}\, Y_1}{\sqrt{n}} \xrightarrow[n\to\infty]{d} Y \sim N(0, K),$$

mit $N(0, K)$ eine $(r-1)$-dimensionale multivariate Normalverteilung (vergleiche Vorlesungsskript WR, Beispiel 3.4.1. 3.) mit Erwartungswertvektor Null und Kovarianzmatrix $K = (\sigma^2_{ij})$, wobei

$$\sigma^2_{ij} = \begin{cases} -p_{0i}p_{0j}, & i \neq j, \\ p_{0i}(1 - p_{0j}), & i = j \end{cases}$$

für $i, j = 1, \ldots, r-1$ ist. Diese Matrix K ist invertierbar mit $K^{-1} = A = (a_{ij})$,

$$a_{ij} = \begin{cases} \frac{1}{p_{0r}}, & i \neq j, \\ \frac{1}{p_{0i}} + \frac{1}{p_{0r}}, & i = j. \end{cases}$$

Außerdem ist K (als Kovarianzmatrix) symmetrisch und positiv definit. Aus der linearen Algebra ist bekannt, daß es eine invertierbare $(r-1) \times (r-1)$-Matrix $A^{1/2}$ gibt, mit der Eigenschaft $A = A^{1/2}(A^{1/2})^\top$. Daraus folgt,

$$K = A^{-1} = ((A^{1/2})^\top)^{-1} \cdot (A^{1/2})^{-1}.$$

Wenn wir $(A^{1/2})^\top$ auf Z'_n anwenden, so bekommen wir

$$(A^{1/2})^\top \cdot Z'_n \xrightarrow[n\to\infty]{d} (A^{1/2})^\top \cdot Y,$$

wobei

$$(A^{1/2})^\top \cdot Y \sim N\left(0,\, (A^{1/2})^\top \cdot K \cdot A^{1/2}\right) = N\left(0,\, \mathcal{I}_{r-1}\right)$$

nach der Eigenschaft der multivariaten Normalverteilung, die im Kapitel 3, Satz 3.1.3 behandelt wird. Des Weiteren wurde hier der Stetigkeitssatz aus der Wahrscheinlichkeitsrechnung benutzt, daß

$$Y_n \xrightarrow[n\to\infty]{d} Y \Longrightarrow \varphi(Y_n) \xrightarrow[n\to\infty]{d} \varphi(Y)$$

für beliebige Zufallsvektoren $\{Y_n\}$, $Y \in \mathbb{R}^m$ und stetige Abbildungen $\varphi : \mathbb{R} \to \mathbb{R}$. Diesen Satz haben wir in WR für Zufallsvariablen bewiesen (Satz 6.4.3, Vorlesungsskript WR). Die erneute Anwendung des Stetigkeitssatzes ergibt

$$\left|(A^{1/2})^\top Z_n'\right|^2 \xrightarrow[n\to\infty]{d} |Y|^2 = R \sim \chi^2_{r-1}.$$

Zeigen wir, daß

$$T_n(X_1,\ldots,X_n) = \left|(A^{1/2})^\top Z_n'\right|^2.$$

Es gilt:

$$\begin{aligned}
\left|(A^{1/2})^\top Z_n'\right|^2 &= ((A^{1/2})^\top Z_n')^\top((A^{1/2})^\top Z_n') = Z_n'^\top \cdot \underbrace{A^{1/2} \cdot (A^{1/2})^\top}_{A} Z_n' = Z_n'^\top A Z_n \\
&= n\sum_{j=1}^{r-1} \frac{1}{p_{0j}}\left(\frac{Z_{nj}}{n} - p_{0j}\right)^2 + \frac{n}{p_{0r}} \sum_{i=1}^{r-1}\sum_{j=1}^{r-1} \left(\frac{Z_{ni}}{n} - p_{0i}\right)\left(\frac{Z_{nj}}{n} - p_{0j}\right) \\
&= \sum_{j=1}^{r-1} \frac{(Z_{nj} - np_{0j})^2}{np_{0j}} + \frac{n}{p_{0r}} \left(\sum_{j=1}^{r-1}\left(\frac{Z_{nj}}{n} - p_{0j}\right)\right)^2 \\
&= \sum_{j=1}^{r-1} \frac{(Z_{nj-np_{0j}})^2}{np_{0j}} + \frac{n}{p_{0r}}\left(\frac{Z_{nr}}{n} - p_{0r}\right)^2 \\
&= \sum_{j=1}^{r} \frac{(Z_{nj} - np_{0j})^2}{np_{0j}} = T_n(X_1,\ldots,X_n),
\end{aligned}$$

weil

$$\sum_{j=1}^{r-1} Z_{nj} = n - Z_{nr},$$
$$\sum_{j=1}^{r-1} p_{0j} = 1 - p_{0r}.$$

□

Lemma 2.4.2 (Multivariater zentraler Grenzwertsatz)**.** Sei $\{Y_n\}_{n\in\mathbb{N}}$ eine Folge von unabhängigen und identisch verteilten Zufallsvektoren, mit $\mathbb{E}\, Y_1 = \mu$ und Kovarianzmatrix K. Dann gilt

$$\frac{\sum\limits_{i=1}^{n} Y_i - n\mu}{\sqrt{n}} \xrightarrow[n\to\infty]{d} Y \sim N(0,K). \tag{2.4.2}$$

Beweis. Sei $Y_j = (Y_{j1}, \ldots, Y_{jm})^\top$. Nach dem Stetigkeitssatz für charakteristische Funktionen ist die Konvergenz (2.4.2) äquivalent zu

$$\varphi_n(t) \underset{n\to\infty}{\longrightarrow} \varphi(t) \quad t \in \mathbb{R}^m, \tag{2.4.3}$$

wobei

$$\varphi_n(t) = \mathbb{E}\, e^{itS_n} = \mathbb{E}\, \exp\left\{ i \sum_{j=1}^{m} t_j \frac{Y_{1j} + \ldots + Y_{nj} - n\mu_j}{\sqrt{n}} \right\}$$

die charakteristische Funktion vom Zufallsvektor

$$S_n = \frac{\sum\limits_{i=1}^{n} Y_i - n\mu}{\sqrt{n}}$$

und

$$\varphi(t) = e^{-t^\top K t/2}$$

die charakteristische Funktion der $N(0,K)$-Verteilung ist. Die Funktion $\varphi_n(t)$ kann in der Form

$$\varphi_n(t) = \mathbb{E}\, \exp\left\{ i \sum_{i=1}^{n} \frac{\sum\limits_{j=1}^{m} t_j(Y_{ij} - \mu_j)}{\sqrt{n}} \right\}, \quad t = (t_1, \ldots, t_m)^\top \in \mathbb{R}^m$$

umgeschrieben werden, wobei für die Zufallsvariable

$$L_i := \sum_{j=1}^{m} t_j (Y_{ij} - \mu_j)$$

gilt:

$$\begin{aligned} \mathbb{E}\, L_i &= 0, \\ \operatorname{Var} L_i &= \mathbb{E}\left[\sum_{k,j=1}^{m} t_j (Y_{ij} - \mu_j)(Y_{ik-\mu_k}) t_k \right] = t^\top K t, \quad i \in \mathbb{N}. \end{aligned}$$

Falls $t^\top Kt = 0$, dann gilt $L_i = 0$ fast sicher, für alle $i \in \mathbb{N}$. Hieraus folgt $\varphi_n(t) = \varphi(t) = 1$, also gilt die Konvergenz 2.4.2.

Falls jedoch $t^\top Kt > 0$, dann kann $\varphi_n(t)$ als charakteristische Funktion der Zufallsvariablen

$$\sum_{i=1}^n L_i/\sqrt{n}$$

an Stelle 1, und $\varphi(t)$ als charakteristische Funktion der eindimensionalen Normalverteilung $N(0, t^\top Kt)$ an Stelle 1 interpretiert werden. Aus dem zentralen Grenzwertsatz für eindimensionale Zufallsvariablen (vergleiche Satz 7.2.1, Vorlesungsskript WR) gilt

$$\sum_{i=1}^n \frac{L_i}{\sqrt{n}} \xrightarrow[n\to\infty]{d} L \sim N(0, t^\top Kt)$$

und somit

$$\varphi_n(t) = \varphi_{\left(\sum_{i=1}^n L_i/\sqrt{n}\right)}(1) \xrightarrow[n\to\infty]{} \varphi_L(1) = \varphi(t).$$

Somit ist die Konvergenz (2.4.2) bewiesen. □

Bemerkung 2.4.1. 1. Die im letzten Beweis verwendete Methode der Reduktion einer mehrdimensionalen Konvergenz auf den eindimensionalen Fall mit Hilfe von Linearkombinationen von Zufallsvariablen trägt den Namen von *Cramér-Wold*.

2. Der χ^2-Pearson-Test ist asymptotisch, also für große Stichprobenumfänge, anzuwenden. Aber welches n ist groß genug? Als „Faustregel“gilt: np_{0j} soll größer gleich a sein, $a \in (2, \infty)$. Für eine größere Klassenanzahl $r \geq 10$ kann sogar $a = 1$ verwendet werden. Wir zeigen jetzt, daß der χ^2-Anpassungstest konsistent ist.

Lemma 2.4.3. Der χ^2-Pearson-Test ist konsistent, das heißt

$$\forall p \in [0,1]^{r-1},\ p \neq p_0 \text{ gilt: } \lim_{n\to\infty} \mathbb{P}_p\left(T_n(X_1,\ldots,X_n) > \chi^2_{r-1,1-\alpha}\right) = 1$$

Beweis. Unter H_1 gilt

$$Z_{nj}/n = \frac{\sum_{i=1}^n \mathbb{I}(a_j < X_i \leq b_j)}{n} \xrightarrow[n\to\infty]{f.s} \underbrace{\mathbb{E}\,\mathbb{I}(a_j < X_1 \leq b_j)}_{=p_j}$$

nach dem starken Gesetz der großen Zahlen. Wir wählen j so, daß $p_j \neq p_{0j}$. Es gilt

$$T_n(X_1,\ldots,X_n) \geq \frac{(Z_{nj} - np_{0j})^2}{np_{0j}} \geq \underbrace{n\left(\frac{Z_{nj}}{n} - p_{0j}\right)^2}_{\sim n(p_j - p_{0j})^2} \xrightarrow[n\to\infty]{f.s} \infty.$$

Somit ist auch

$$\mathbb{P}_p\left(T_n(X_1,\ldots,X_n) > \chi^2_{r-1,1-\alpha}\right) \xrightarrow[n\to\infty]{f.s.} 1.$$

□

2.4.2 χ^2-Anpassungstest von Pearson-Fisher

Es sei $(X_1, \ldots, X_n)$ eine Stichprobe von unabhängigen und identisch verteilten Zufallsvariablen X_i, $i = 1, \ldots, n$. Wir wollen testen, ob die Verteilungsfunktion F von X_i zu einer parametrischen Familie

$$\Lambda_0 = \{F_\theta : \theta \in \Theta\}$$

gehört. Seien die Zahlen a_i, b_i, $i = 1, \ldots, r$ vorgegeben mit der Eigenschaft

$$-\infty \leq a_1 < b_1 = a_2 < b_2 = \ldots = a_r < b_r \leq \infty$$

und

$$\begin{aligned} Z_j &= \#\{X_i,\, i = 1, \ldots, n : \, a_j < X_i \leq b_j\}, \quad j = 1, \ldots, r, \\ Z &= (Z_1, \ldots, Z_r)^\top . \end{aligned}$$

Nach Lemma 2.4.1 gilt: $Z \sim M_{r-1}(n, p)$, $p = (p_0, \ldots, p_{r-1})^\top \in [0,1]^{r-1}$. Unter der Hypothese $H_0 : F \in \Lambda_0$ gilt: $p = p(\theta)$, $\theta \in \Theta \subset \mathbb{R}^m$. Wir vergröbern die Hypothese H_0 und wollen folgende neue Hypothese testen:

$$H_0 : p \in \{p(\theta) : \theta \in \Theta\} \text{ vs. } H_1 : p \notin \{p(\theta) : \theta \in \Theta\} .$$

Um dieses Hypothesenpaar zu testen, wird der χ^2-*Pearson-Fisher-Test* wie folgt aufgebaut:

1. Ein (schwach konsistenter) Maximum-Likelihood-Schätzer $\hat{\theta}_n = \hat{\theta}(X_1, \ldots, X_n)$ für θ wird gefunden: $\hat{\theta}_n \underset{n\to\infty}{\xrightarrow{P}} \theta$. Dabei muß $\{\hat{\theta}_n\}_{n\in\mathbb{N}}$ asymptotisch normalverteilt sein.

2. Es wird der Plug-In-Schätzer $p(\hat{\theta}_n)$ für $p(\theta)$ gebildet.

3. Die Testgröße

$$\hat{T}_n(X_1, \ldots, X_n) = \sum_{j=1}^{r} \frac{\left(Z_{nj} - np_j(\hat{\theta})\right)^2}{np_j(\hat{\theta})} \underset{n\to\infty}{\xrightarrow{P}} \varphi \sim \chi^2_{r-m-1}$$

   unter H_0 und gewissen Voraussetzungen.

4. H_0 wird verworfen, falls $\hat{T}_n(X_1, \ldots, X_n) > \chi^2_{r-m-1,1-\alpha}$. Dies ist ein asymptotischer Test zum Niveau α.

Bemerkung 2.4.2. 1. Bei einem χ^2-Pearson-Fisher-Test wird vorausgesetzt, daß die Funktion $p(\theta)$ explizit bekannt ist, θ jedoch unbekannt. Das bedeutet, daß für jede Klasse von Verteilungen Λ_0 die Funktion $p(\cdot)$ berechnet werden soll.

2. Warum kann $\hat{T}_n$ die Hypothese H_0 von H_1 unterscheiden? Nach dem Gesetz der großen Zahlen gilt

$$\frac{1}{n}Z_{nj} - p_j(\hat{\theta}_n) = \underbrace{\frac{1}{n}Z_{nj} - p_j(\theta)}_{\xrightarrow{P} 0} - \underbrace{(p_j(\hat{\theta}_n) - p_j(\theta))}_{\xrightarrow{P} 0} \xrightarrow[n\to\infty]{P} 0,$$

falls $\hat{\theta}_n$ schwach konsistent ist und $p_j(\cdot)$ eine stetige Funktion für alle $j = 1, \ldots, r$ ist.

Das heißt, unter H_0 soll $\hat{T}_n(X_1, \ldots, X_n)$ relativ kleine Werte annehmen. Eine signifikante Abweichung von diesem Verhalten soll zur Ablehnung von H_0 führen, vergleiche Punkt 4.

Für die Verteilung F_θ von X_i gelten folgende Regularitätsvoraussetzungen (vergleiche Satz 3.4.2, Vorlesungsskript Statistik I).

1. Die Verteilungsfunktion F_θ ist entweder diskret oder absolut stetig für alle $\theta \in \Theta$.
2. Die Parametrisierung ist eindeutig, das heißt: $\theta \neq \theta_1 \Leftrightarrow F_\theta \neq F_{\theta_1}$.
3. Der Träger der Likelihood-Funktion

$$L(x,\theta) = \begin{cases} P_\theta(X_1 = x), & \text{im Falle von diskreten } F_\theta, \\ f_\theta(x), & \text{im absolut stetigen Fall.} \end{cases}$$

$$\mathrm{Supp} L(x,\theta) = \{x \in \mathbb{R} : L(x,\theta) > 0\} \text{ hängt nicht von } \theta \text{ ab.}$$

4. $L(x,\theta)$ sei 3 Mal stetig differenzierbar, und es gelte für $k = 1 \ldots 3$ und $i_1, \ldots, i_k \in \{1 \ldots m\}$, daß

$$\left(\sum\right) \int \frac{\partial^k L(x,\theta)}{\partial\theta_{i_1} \cdot \ldots \cdot \partial\theta_{i_k}} dx = \frac{\partial^k}{\partial\theta_{i_1} \cdot \ldots \cdot \partial\theta_{i_k}} \left(\sum\right) \int L(x,\theta) dx = 0.$$

5. Für alle $\theta_0 \in \Theta$ gibt es eine Konstante c_{θ_0} und eine messbare Funktion g_{θ_0} : $\mathrm{Supp}L \to \mathbb{R}_+$, sodaß

$$\left| \frac{\partial^3 \log L(x,\theta)}{\partial\theta_{i_1}\partial\theta_{i_2}\partial\theta_{i_3}} \right| \leq g_{\theta_0}(x), |\theta - \theta_0| < c_{\theta_0}$$

und

$$\mathbb{E}_{\theta_0}\, g_{\theta_0}(X_1) < \infty.$$

Wir definieren die *Informationsmatrix von Fisher* durch

$$I(\theta) = \left(\mathbb{E} \left[\frac{\partial \log L(X_1,\theta)}{\partial\theta_i} \frac{\partial \log L(X_1,\theta)}{\partial\theta_j} \right] \right)_{i,j=1,\ldots,m}. \tag{2.4.4}$$

Satz 2.4.2 (asymptotische Normalverteiltheit von konsistenten ML-Schätzern $\hat{\theta}_n$, multivariater Fall $m > 1$)**.** Es seien $X_1, \ldots, X_n$ unabhängig und identisch verteilt mit Likelihood-Funktion L, die den Regularitätsbedingungen 1-5 genügt. Sei $I(\theta)$ positiv definit für alle $\theta \in \Theta \subset \mathbb{R}^m$. Sei $\hat{\theta}_n = \hat{\theta}(X_1, \ldots, X_n)$ eine Folge von schwach konsistenten Maximum-Likelihood-Schätzern für θ. Dann gilt:

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow[n\to\infty]{d} N(0, I^{-1}(\theta)).$$

Ohne Beweis; siehe den Beweis des Satzes 3.4.2, Vorlesungsskript Statistik I.

Für unsere vergröberte Hypothese $H_0 : p \in \{p(\theta), \theta \in \Theta\}$ stellen wir folgende, stückweise konstante, Likelihood-Funktion auf:

$$L(x, \theta) = p_j(\theta), \text{ falls } x \in (a_j, b_j].$$

Dann ist die Likelihood-Funktion der Stichprobe $(x_1, \ldots, x_n)$ gleich

$$\begin{aligned} L(x_1, \ldots, x_n, \theta) &= \prod_{j=1}^{n} p_j(\theta)^{Z_j(x_1,\ldots,x_n)} \\ \Rightarrow \log L(x_1, \ldots, x_n, \theta) &= \sum_{j=1}^{r} Z_j(x_1, \ldots, x_n) \cdot \log p_j(\theta). \\ \hat{\theta}_n = \hat{\theta}(x_1, \ldots, x_n) &= \underset{\theta\in\Theta}{\operatorname{argmax}} \log L(x_1, \ldots, x_n, \theta) \\ \Rightarrow \sum_{j=1}^{r} Z_j(x_1, \ldots, x_n) \frac{\partial p_j(\theta)}{\partial \theta_i} \cdot \frac{1}{p_j(\theta)} &= 0, \quad i = 1, \ldots, m. \end{aligned}$$

Aus $\sum_{i=1}^{r} p_j(\theta) = 1$ folgt

$$\sum_{j=1}^{r} \frac{\partial p_j(\theta)}{\partial \theta_i} = 0 \Rightarrow \sum_{j=1}^{r} \frac{Z_j(x_1, \ldots, x_n) - np_j(\theta)}{p_j(\theta)} \cdot \frac{\partial p_j(\theta)}{\partial \theta_i} = 0, \quad i = 1, \ldots, m.$$

Lemma 2.4.4. Im obigen Fall gilt $I(\theta) = C^\top(\theta) \cdot C(\theta)$, wobei $C(\theta)$ eine $(r \times m)$-Matrix mit Elementen

$$c_{ij}(\theta) = \frac{\partial p_i(\theta)}{\partial \theta_j} \cdot \frac{1}{\sqrt{p_i(\theta)}} \quad \text{ist.}$$

Beweis.

$$\begin{aligned} &\mathbb{E}_0 \left[\frac{\partial \log L(X_1, \theta)}{\partial \theta_i} \cdot \frac{\partial \log L(X_1, \theta)}{\partial \theta_j} \right] = \sum_{k=1}^{r} \frac{\partial \log p_k(\theta)}{\partial \theta_i} \cdot \frac{\partial \log p_k(\theta)}{\partial \theta_j} \cdot p_k(\theta) \\ &= \sum_{k=1}^{r} \frac{\partial p_k(\theta)}{\partial \theta_i} \frac{1}{p_k(\theta)} \cdot \frac{\partial p_k(\theta)}{\partial \theta_j} \cdot \frac{1}{p_k(\theta)} \\ &= \Big(C^\top(\theta) \cdot C(\theta) \Big)_{ij}, \end{aligned}$$

$$\text{denn } \log L(X_1, \theta) = \sum_{i=1}^{r} \log p_j(\theta) \cdot \mathbb{I}\left(x \in (a_j, b_j]\right).$$

□

Deshalb gilt die Folgerung aus Satz 2.4.2:

Folgerung 2.4.1. Sei $\hat{\theta}_n = \hat{\theta}(X_1, \ldots, X_n)$ ein Maximum-Likelihood-Schätzer von θ im vergröberten Modell, der schwach konsistent ist und den obigen Regularitätsbedingungen genügt. Sei die Informationsmatrix von Fisher $I(\theta) = C^\top(\theta) \cdot C(\theta)$ für alle $\theta \in \Theta$ positiv definit. Dann ist $\hat{\theta}$ asymptotisch normalverteilt:

$$\sqrt{n}\left(\hat{\theta}_n - \theta\right) \xrightarrow[n\to\infty]{d} Y \sim N\left(0, I^{-1}(\theta)\right)$$

Satz 2.4.3. Es sei $\hat{\theta}_n$ ein Maximum-Likelihood-Schätzer im vergröberten Modell für θ, für den alle Voraussetzungen der Folgerung 2.4.1 erfüllt sind. Die Teststatistik

$$\hat{T}_n(X_1, \ldots, X_n) = \sum_{j=1}^{r} \frac{(Z_j(X_1, \ldots, X_n) - np_j(\hat{\theta}_n))^2}{np_j(\hat{\theta}_n)}$$

ist unter H_0 asymptotisch χ^2_{r-m-1}-verteilt:

$$\lim_{n\to\infty} \mathbb{P}_\theta \left(\hat{T}_n(X_1, \ldots, X_n) > \chi^2_{r-m-1,1-\alpha}\right) = \alpha.$$

ohne Beweis (siehe [13]).

Aus diesem Satz folgt, daß der χ^2-Pearson-Fisher-Test ein asymptotischer Test zum Niveau α ist.

Beispiel 2.4.1. 1. *χ^2-Pearson-Fisher-Test der Normalverteilung*

Sei $(X_1, \ldots, X_n)$ eine Zufallsstichprobe. Es soll geprüft werden, ob $X_i \sim N(\mu, \sigma^2)$. Es gilt

$$\theta = (\mu, \sigma^2) \in \Theta = \mathbb{R} \times \mathbb{R}_+.$$

Sei $(a_j, b_j]_{j=1,\ldots,r}$ eine beliebige Aufteilung von $\mathbb{R}$ in r disjunkte Intervalle. Sei

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

die Dichte der $N(\mu, \sigma^2)$-Verteilung.

$$p_j(\theta) = \mathbb{P}_0\left(a_j < X_1 \le b_j\right) = \int_{a_j}^{b_j} f_\theta(x)dx, \quad j = 1, \ldots, r$$

mit den Klassenstärken

$$Z_j = \#\left\{i : X_i \in (a_j, b_j]\right\}.$$

Wir suchen den Maximum-Likelihood-Schätzer im vergröberten Modell:

$$\begin{aligned}
\frac{\partial p_j(\theta)}{\partial \mu} &= \int_{a_j}^{b_j} \frac{\partial}{\partial \mu} f_\theta(x)dx = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \int_{a_j}^{b_j} \frac{x-\mu}{\sigma^2} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx \\
\frac{\partial p_j(\theta)}{\partial \sigma^2} &= \int_{a_j}^{b_j} \frac{\partial}{\partial \sigma^2} f_\theta(x)dx \\
&= \frac{1}{\sqrt{2\pi}} \int_{a_j}^{b_j} \left[-\frac{1}{2} \cdot \frac{1}{(\sigma^2)^{3/2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} + \frac{1}{\sqrt{\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \cdot \left(\frac{(x-\mu)^2}{2(\sigma^2)^2} \right) \right] dx \\
&= -\frac{1}{2}\frac{1}{\sigma^2} \int_{a_j}^{b_j} f_\theta(x)dx + \frac{1}{2(\sigma^2)^2} \int_{a_j}^{b_j} (x-\mu)^2 f_\theta(x)dx
\end{aligned}$$

Die notwendigen Bedingungen des Maximums sind:

$$\sum_{i=1}^{r} Z_j \frac{\int\limits_{a_j}^{b_j} x f_\theta(x)dx}{\int\limits_{a_j}^{b_j} f_\theta(x)dx} - \mu \underbrace{\sum_{j=1}^{r} Z_j}_{=n} = 0,$$

$$\frac{1}{\sigma^2} \sum_{j=1}^{r} Z_j \frac{\int\limits_{a_j}^{b_j} (x-\mu)^2 f_\theta(x)dx}{\int\limits_{a_j}^{b_j} f_\theta(x)dx} - \underbrace{\sum_{j=1}^{r} Z_j}_{=n} = 0.$$

Daraus folgen die Maximum-Likelihood-Schätzer $\hat{\mu}$ und $\hat{\sigma}^2$ für μ und σ^2:

$$\hat{\mu} = \frac{1}{n} \sum_{j=1}^{r} Z_j \frac{\int\limits_{a_j}^{b_j} x f_\theta(x)dx}{\int\limits_{a_j}^{b_j} f_\theta(x)dx},$$

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{j=1}^{r} Z_j \frac{\int\limits_{a_j}^{b_j} (x-\mu)^2 f_\theta(x)dx}{\int\limits_{a_j}^{b_j} f_\theta(x)dx}.$$

Wir konstruieren eine Näherung zu $\hat{\mu}$ und $\hat{\sigma}^2$ für $r \to \infty$. Falls $r \to \infty$ (und somit auch $n \to \infty$), dann ist $b_j - a_j$ klein und nach der einfachen Quadraturregel gilt:

$$\int_{a_j}^{b_j} x f_\theta(x)dx \approx (b_j - a_j)\, y_j f_\theta(y_j),$$

$$\int_{a_j}^{b_j} f_\theta(x)dx \approx (b_j - a_j)\, f_\theta(y_j),$$

wobei $y_1 = b_1$, $y_r = b_{r-1} = a_r$,

$$y_j = (b_{j+1} + b_j)/2, \quad j = 2, \ldots, r-1.$$

Daraus folgen für die Maximum-Likelihood-Schätzer $\hat{\mu}$ und $\hat{\sigma}^2$:

$$\hat{\mu} \approx \frac{1}{n} \sum_{j=1}^{r} y_j - Z_j = \tilde{\mu}$$

$$\hat{\sigma}^2 \approx \frac{1}{n} \sum_{j=1}^{r} (y_j - \tilde{\mu})^2 Z_j = \tilde{\sigma}^2,$$

$$\tilde{\theta} = \left(\tilde{\mu}, \tilde{\sigma}^2\right).$$

Der χ^2-Pearson-Fisher-Test lautet dann: H_0 wird abgelehnt, falls

$$\hat{T}_n = \frac{\sum_{j=1}^{r} \left(Z_j - np_j(\tilde{\theta})\right)^2}{np_j(\tilde{\theta})} > \chi^2_{r-3,1-\alpha}.$$

2. χ^2-*Pearson-Fisher-Test der Poissonverteilung*

Es sei $(X_1, \ldots, X_n)$ eine Stichprobe von unabhängigen und identisch verteilen Zufallsvariablen. Wir wollen testen, ob $X_i \sim \text{Poisson}(\lambda)$, $\lambda > 0$. Es gilt $\theta = \lambda$ und $\Theta = (0, +\infty)$. Die Vergröberung von Θ hat die Form

$$-\infty = a_1 < \underbrace{b_1}_{=0} = a_2 < \underbrace{b_2}_{=1} = a_3 < \ldots < \underbrace{b_{r-1}}_{=r-2} = a_r < b_r = +\infty.$$

Dann ist

$$\begin{aligned}
p_j(\lambda) &= \mathbb{P}_\lambda\left(X_1 = j-1\right) = e^{-\lambda} \frac{\lambda^{j-1}}{(j-1)!}, \quad j = 1, \ldots, r-1, \\
p_r(\lambda) &= \sum_{i=r-1}^{\infty} e^{-\lambda} \frac{\lambda^i}{i!}, \\
\frac{dp_j(\lambda)}{d\lambda} &= e^{-\lambda} \frac{\lambda^{j-1}}{(j-1)!} + (j-1) \frac{\lambda^{j-1}}{(j-1)!} e^{-\lambda} = e^{-\lambda} \frac{\lambda^{j-1}}{(j-1)! \left(\frac{j-1}{\lambda} - 1\right)} \\
&= p_j(\lambda) \cdot \left(\frac{j-1}{\lambda} - 1\right), \quad j = 1, \ldots, r-1 \\
\frac{dp_r(\lambda)}{d\lambda} &= \sum_{i \geq r} p_i(\lambda) \left(\frac{i-1}{\lambda} - 1\right).
\end{aligned}$$

Die Maximum-Likelihood-Gleichung lautet

$$0 = \sum_{j=1}^{r-1} Z_j \cdot \left(\frac{j-1}{\lambda} - 1\right) + Z_r \frac{\sum_{i \geq r} p_i(\lambda) \frac{i-1}{\lambda} - 1}{p_r(\lambda)}$$

Falls $r \longrightarrow \infty$, so findet sich $r(n)$ für jedes n, für das $Z_r = 0$. Deshalb gilt für $r > r(n)$:

$$\sum_{j=1}^{r-1}(j-1)Z_j - \lambda \underbrace{\sum_{j=1}^{n} Z_j}_{=n} = 0,$$

woraus der Maximum-Likelihood-Schätzer

$$\frac{1}{n}\sum_{j=1}^{r-1}(j-1)Z_j = \frac{1}{n}\sum_{j=1}^{n} X_j = \overline{X}_n$$

folgt. Der χ^2-Pearson-Fisher-Test lautet: H_0 wird verworfen, falls

$$\hat{T}_n = \sum_{j=1}^{r} \frac{\left(Z_j - np_\lambda(\overline{X}_n)\right)^2}{\left(np_j(\overline{X}_n)\right)^2} > \chi^2_{r-2,1-\alpha}.$$

2.4.3 Anpassungstest von Shapiro

Es sei $(X_1, \ldots, X_n)$ eine Stichprobe von unabhängigen, identisch verteilten Zufallsvariablen, $X_i \sim F$. Getestet werden soll die Hypothese

$$H_0 : F \in \{N(\mu, \sigma^2) : \mu \in \mathbb{R},\ \sigma^2 > 0\} \text{ vs. } H_1 : F \notin \{N(\mu, \sigma^2), \mu \in \mathbb{R},\ \sigma^2 > 0\}.$$

Die in den Abschnitten 2.4.1 - 2.4.2 vorgestellten χ^2-Tests sind asymptotisch; deshalb können sie für relativ kleine Stichprobenumfänge nicht verwendet werden.

Der folgende Test wird diese Lücke füllen und eine Testentscheidung über H_0 selbst bei kleinen Stichproben ermöglichen.

Man bildet Ordnungsstatistiken $X_{(1)}, \ldots, X_{(n)}$, $X_{(1)} \leq X_{(2)} \leq \ldots, \leq X_{(n)}$ und vergleicht ihre Korreliertheit mit den Mittelwerten der entsprechenden Ordnungsstatistiken der $N(0,1)$-Verteilung. Sei $(Y_1, \ldots, Y_n)$ eine Stichprobe von unabhängigen und identisch verteilten Zufallsvariablen, $Y_1 \sim N(0,1)$. Es sei $a_i = \mathbb{E}\, Y_{(i)}$, $i = 1, \ldots, n$. Falls der empirische Korrelationskoeffizient ρ_{aX} zwischen $(a_1, \ldots, a_n)$ und $(X_{(1)}, \ldots, X_{(n)})$ bei 1 liegt, dann ist die Stichprobe normalverteilt. Formalisieren wir diese Heuristik:

Es sei b_i der Erwartungswert der i-ten Ordnungsstatistik in einer Stichprobe von $N(\mu, \sigma^2)$-verteilten, unabhängigen Zufallsvariablen Z_i: $b_i = \mathbb{E}\, Z_{(i)}$, $i = 1, \ldots, n$. Es gilt: $b_i = \mu + \sigma a_i$, $i = 1, \ldots, n$. Betrachten wir den Korrelationskoeffizienten

$$\rho_{bX} = \frac{\sum_{i=1}^{n}\left(b_i - \bar{b}_n\right)\left(X_{(i)} - \overline{X}_n\right)}{\sqrt{\sum_{i=1}^{n}\left(b_i - \bar{b}_n\right)^2 \sum_{i=1}^{n}\left(X_{(i)} - \overline{X}_n\right)^2}}. \tag{2.4.5}$$

Da ρ invariant bezüglich Lineartransformationen ist und
$\sum_{i=1}^n a_i = \sum_{i=1}^n \mathbb{E}\, Y_{ii} = \mathbb{E}\left(\sum_{i=1}^n Y_i\right) = 0,$ gilt:

$$\rho_{bX} \overset{\text{(Statistik I)}}{=} \rho_{aX} = \frac{\sum\limits_{i=1}^n a_i \left(X_{(i)} - \overline{X}_n\right)}{\sqrt{\sum\limits_{i=1}^n a_i^2 \sum\limits_{i=1}^n \left(X_i - \overline{X}_n\right)^2}} = \frac{\sum\limits_{i=1}^n X_{(i)} - \overline{X}_n \overbrace{\sum_{i=1}^n a_i}^{=0}}{\sqrt{\sum\limits_{i=1}^n a_i^2 \sum\limits_{i=1}^n \left(X_i - \overline{X}_n\right)^2}}$$

$$= \frac{\sum\limits_{i=1}^n a_i X_{(i)}}{\sqrt{\sum\limits_{i=1}^n a_i^2 \cdot \sum\limits_{i=1}^n \left(X_i - \overline{X}_n\right)^2}}$$

Die Testgröße lautet:

$$T_n = \frac{\sum\limits_{i=1}^n a_i X_{(i)}}{\sum\limits_{i=1}^n a_i^2 \sum\limits_{i=1}^n \left(X_i - \overline{X}_n\right)} \qquad \text{(Shapiro-Francia-Test)}$$

Die Werte a_i sind bekannt und können den Tabellen bzw. der Statistik-Software entnommen werden. Es gilt: $|T_n| \leq 1$.

H_0 wird abgelehnt, falls $T_n \leq q_{n,\alpha}$, wobei $q_{n,\alpha}$ das α-Quantil der Verteilung von T_n ist. Diese Quantile sind aus den Tabellen bekannt, bzw. können durch Monte-Carlo-Simulationen berechnet werden.

Bemerkung 2.4.3. Einen anderen, weit verbreiteten Test dieser Art bekommt man, wenn man die Lineartransformation $b_i = \mu + \sigma a_i$ durch eine andere Lineartransformation ersetzt:

$$\left(a'_1, \ldots, a'_n\right)^\top = K^{-1} \cdot (a_1, \ldots, a_n),$$

wobei $K = (k_{ij})_{j=1}^n$, die Kovarianzmatrix von $\left(Y_{(1)}, \ldots, Y_{(n)}\right)$ ist:

$$\underset{i,j=1,\ldots,n}{k_{ij}} = \mathbb{E}\left(Y_{(i)} - a_i\right)\left(Y_{(j)} - a_j\right)$$

Der so konstruierte Test trägt den Namen *Shapiro-Wilk-Test*.

2.5 Weitere, nicht parametrische Tests

2.5.1 Binomialtest

Es sei $(X_1, \ldots, X_n)$ eine Stichprobe von unabhängigen, identisch verteilten Zufallsvariablen, wobei $X_i \sim \text{Bernoulli}(p)$. Getestet werden soll:

$$H_0 : p = p_0 \text{ vs. } H_1 : p \neq p_0$$

Die Teststatistik lautet

$$T_n = \sum_{i=1}^{n} X_i \underset{H_0}{\sim} \mathrm{Bin}(n, p_0),$$

und die Entscheidungsregel ist: H_0 wird verworfen, falls

$$T_n \notin [\mathrm{Bin}(n, p_0)_{\alpha/2}, \mathrm{Bin}(n, p_0)_{1-\alpha/2}],$$

wobei $\mathrm{Bin}(n, p)_\alpha$ das α-Quantil der $\mathrm{Bin}(n, p)$-Verteilung ist.

Für andere H_0, wie zum Beispiel $p \leq p_0$ ($p \geq p_0$) muss der Ablehnungsbereich entsprechend angepasst werden.

Die Quantile $\mathrm{Bin}(n, p)_\alpha$ erhält man aus Tabellen oder aus Monte-Carlo-Simulationen. Falls n groß ist, können diese Quantile durch die Normalapproximation berechnet werden:

Nach dem zentralen Grenzwertsatz von DeMoivre Laplace gilt:

$$\mathbb{P}(T_n \leq x) = \mathbb{P}\left(\frac{T_n - np_0}{np_0(1-p_0)} \leq \frac{x - np_0}{np_0(1-p_0)}\right) \underset{n\to\infty}{\approx} \Phi\left(\frac{x - np_0}{\sqrt{np_0(1-p_0)}}\right).$$

Daraus folgt:

$$z_\alpha \approx \frac{\mathrm{Bin}(n, p_0)_\alpha - np_0}{\sqrt{np_0(1-p_0)}}$$
$$\Rightarrow \mathrm{Bin}(n, p_0)_\alpha \approx \sqrt{np_0(1-p_0)} \cdot z_\alpha + np_0$$

Nach der Poisson-Approximation (für $n \to \infty, np_0 \to \lambda_0$) gilt:

$$\mathrm{Bin}(n, p_0)_{\alpha/2} \approx \mathrm{Poisson}(\lambda_0)_{\alpha/2},$$
$$\mathrm{Bin}(n, p_0)_{1-\alpha/2} \approx \mathrm{Poisson}(\lambda_0)_{1-\alpha/2}, \quad \text{wobei } \lambda_0 = np_0.$$

Zielstellung: Wie kann mit Hilfe des oben beschriebenen Binomialtests die Symmetrieeigenschaft einer Verteilung getestet werden?

Es sei $(Y_1, \ldots, Y_n)$ eine Stichprobe von unabhängigen und identisch verteilen Zufallsvariablen mit Verteilungsfunktion F. Getestet werden soll:

$$H_0 : F \text{ ist symmetrisch vs. } H_1 : \text{F ist nicht symmetrisch.}$$

Eine symmetrische Verteilung besitzt den Median bei Null. Deswegen vergröbern wir die Hypothese H_0 und testen:

$$H_0' : F^{-1}(0,5) = 0 \text{ vs. } H_1' : F^{-1}(0,5) \neq 0.$$

Noch allgemeiner: Für ein $\alpha \in [0, 1]$:

$$H_0'' : F^{-1}(\alpha) = \gamma_\alpha \text{ vs. } H_1'' : F^{-1}(\alpha) \neq \gamma_\alpha.$$

H_0'' vs. H_1'' wird mit Hilfe des Binomialtests wie folgt getestet: Sei $X_i = \mathbb{I}(Y_i \leq \gamma_\alpha)$. Unter H_0'' gilt:

$$X_i \sim \text{Bernoulli}(F(\gamma_\alpha)) = \text{Bernoulli}(\alpha).$$

Seien $a_1 = -\infty$, $b_1 = \gamma_\alpha$, $a_2 = b_1$, $b_2 = +\infty$ zwei disjunkte Klassen $(a_1, b_2]$, $(a_2, b_2]$ in der Sprache des χ^2-Pearson-Tests. Die Testgröße ist:

$$T_n = \sum_{i=1}^{n} X_i = \# \{Y_i : Y_i \leq \gamma_\alpha\}$$

Die Hypothese $F^{-1}(\alpha) = \gamma_\alpha$ ist äquivalent zu $H_0''' : p = \alpha$. Die Entscheidungsregel lautet dann: H_0''' wird verworfen, falls $T_n \notin \left[\text{Bin}(n, \alpha)_{\beta/2}, \text{Bin}(n, \alpha)_{1-\beta/2}\right]$. Dies ist ein Test zum Niveau β.

2.5.2 Iterationstests auf Zufälligkeit

In manchen Fragestellungen der Biologie untersucht man eine Folge von 0 oder 1 auf ihre „Zufälligkeit" bzw. Vorhandensein von größeren Clustern von 0 oder 1. Diese Hypothesen kann man mit Hilfe der sogenannten *Iterationstests* statistisch überprüfen.

Sei eine Stichprobe X_i, $i = 1, \ldots, n$ gegeben, $X_i \in \{0, 1\}$, $\sum_{i=1}^{n} X_i = n_1$ die Anzahl der Einsen, $n_1 = n - n_2$ die Anzahl der Nullen, n_1, n_2 vorgegeben. Eine Realisierung von $(X_1, \ldots, X_n)$ wäre zum Beispiel

$$x = (0, 0, 0, 1, 1, 1, 0, 1, 1, 0, 1, 1, 1, 0, 1, 1, 1, 1)$$

Es soll getestet werden, ob

$$\begin{aligned} &H_0 : \text{jede Folge } x \text{ ist gleichwahrscheinlich vs.} \\ &H_1 : \text{Es gibt bevorzugte Folgen (Cluserbildung)} \end{aligned}$$

stimmt.

Sei

$$\Omega = \left\{ x = (x_1, \ldots, x_n),\, x_i = 0 \text{ oder } 1, \quad i = 1, \ldots, n, \sum_{i=1}^{n} x_i = n_1 \right\}$$

der Stichprobenraum. Dann ist der Raum $(\Omega, \mathcal{F}, \mathbb{P})$ mit $\mathcal{F} = \mathcal{P}(\Omega)$,

$$\mathbb{P}(x) = \frac{1}{|\Omega|} = \frac{1}{\binom{n}{n_1}}$$

ein Laplacescher Wahrscheinlichkeitsraum.

Sei $T_n(X) = \#\{\text{Iterationen in } X\} = \#\{\text{ Teilfolgen der Nullen oder Einsen }\} = \#\{\text{Wechselstellen von 0 auf 1 oder von 1 auf 0}\} + 1$.

Zum Beispiel ist für $x = (0, 1, 1, 1, 0, 0, 0, 1, 0, 1, 1, 1, 0, 0)$, $T_n(x) = 7 = 6 + 1$.

$T_n(X)$ wird folgendermaßen als Teststatistik für H_0 vs. H_1 benutzt. H_0 wird abgelehnt, falls $T(x)$ klein ist, das heißt, falls $T_n(x) < F_{T_n}^{-1}(\alpha)$. Dies ist ein Test zum Niveau α. Wie berechnen wir die Quantile $F_{T_n}^{-1}$?

Satz 2.5.1. Unter H_0 gelten folgende Aussagen:

1.
$$\mathbb{P}(T_n = k) = \begin{cases} \frac{2\binom{n_1-1}{i-1}\binom{n_2-1}{i-1}}{\binom{n}{n_1}}, & \text{falls } k = 2i, \\ \frac{\binom{n_1-1}{i}\cdot\binom{n_2-1}{i-1}+\binom{n_1-1}{i-1}\cdot\binom{n_2-1}{i}}{\binom{n}{n_1}}, & \text{falls } k = 2i+1. \end{cases}$$

2.
$$\mathbb{E}\, T_n = 1 + \frac{2n_1n_2}{n}$$

3.
$$\mathrm{Var}\,(T_n) = \frac{2n_1n_2(2n_1n_2 - n)}{n^2(n-1)}$$

Beweis. 1. Wir nehmen an, daß $k = 2i$ (der ungerade Fall ist analog). Wie können i Klumpen von Einsen gewählt werden? Die Anzahl dieser Möglichkeiten = die Anzahl der Möglichkeiten, wie n_1 Teilchen auf i Klassen verteilt werden.

$$0|00|\dots|0|\,(n_1)$$

Dies ist gleich der Anzahl an Möglichkeiten, wie $i-1$ Trennwände auf $n_1 - 1$ Positionen verteilt werden können $= \binom{n_1-1}{i-1}$. Das selbe gilt für die Nullen.

2. Sei $Y_j = \mathbb{I}\{X_{j-1} \neq X_j\}_{j=2,\dots,n}$.

$$\Rightarrow \mathbb{E}\, T_n(X) = 1 + \sum_{j=2}^{n} \mathbb{E}\, Y_j = 1 + \sum_{j=2}^{n} \mathbb{P}(X_{j-1} \neq X_j).$$

$$\begin{aligned}\mathbb{P}(X_{j-1} \neq X_j) &= \frac{2\binom{n-2}{n_1-1}}{\binom{n}{n_1}} = 2 \cdot \frac{\frac{(n-2)!}{(n-2-(n_1-1))!(n_1-1)!}}{\frac{n!}{(n-n_1)!n_1!}} \\ &= \frac{2n_1(n-n_1)}{(n-1)n} \\ &= \frac{2n_1n_2}{n(n-1)}.\end{aligned}$$

Daraus folgt

$$\mathbb{E}\, T_n = 1 + (n-1)\frac{2n_1n_2}{n(n-1)} = 1 + 2\frac{n_1n_2}{n}.$$

3.

Übungsaufgabe 2.5.1. Beweisen Sie Punkt 3.

□

Beispiel 2.5.1 (*Test von Wald-Wolfowitz*)**.** Seien $Y = (Y_1, \dots, Y_n)$, $Z = (Z_1, \dots, Z_n)$ unabhängige Stichproben von unabhängigen und identisch verteilten Zufallsvariablen, $Y_i \sim F$, $Z_i \sim G$. Getestet werden soll:

$$H_0 : F = G \text{ vs. } H_1 : F \neq G.$$

Sei $(Y, Z) = (Y_1, \dots, Y_n, Z_1, \dots, Z_n)$ und seien X'_i Stichprobenvariablen von (Y, Z), $i = 1, \dots, n$, $n = n_1 + n_2$. Wir bilden die Ordnungsstatistiken $X'_{(i)}$, $i = 1, \dots, n$ und setzen

$$X_i = \begin{cases} 1, & \text{falls } X'_{(i)} = Y_j \text{ für ein } j = 1, \dots, n_1, \\ 0, & \text{falls } X'_{(i)} = Z_j \text{ für ein } j = 1, \dots, n_2. \end{cases}$$

Unter H_0 sind die Stichprobenwerte in (Y, Z) gut gemischt, das heißt jede Kombination von 0 und 1 in $(X_1, \dots, X_n)$ ist gleichwahrscheinlich. Darum können wir den Iterationstest auf Zufälligkeit anwenden, um H_0 vs. H_1 zu testen: H_0 wird verworfen, falls $T_n(x) \leq F^{-1}(\alpha)$, $x = (x_1, \dots, x_n)$.

Wie können die Quantile von F_{T_n} für große n berechnet werden? Falls

$$\frac{n_1}{n_1 + n_2} \underset{n\to\infty}{\longrightarrow} p \in (0, 1),$$

dann ist T_n asymptotisch normalverteilt.

Satz 2.5.2. Unter der obigen Voraussetzung gilt:

$$\begin{aligned} \lim_{n\to\infty} \frac{\mathbb{E}\, T_n}{n} &= 2p(1-p) \\ \lim_{n\to\infty} \frac{1}{n} \operatorname{Var} T_n &= 4p^2(1-p)^2 \\ \frac{T_n - 2p(1-p)}{2\sqrt{n}p(1-p)} &\xrightarrow[n\to\infty]{d} Y \sim N(0,1), \quad \text{falls } \frac{n_1}{n_1+n_2} \longrightarrow p \in (0,1). \end{aligned}$$

So können Quantile von T_n näherungsweise für große n folgendermaßen berechnet werden:

$$\begin{aligned} \alpha = \mathbb{P}\left(T_n \leq F_{T_n}^{-1}(\alpha)\right) &= \mathbb{P}\left(\frac{T_n - 2np(1-p)}{2\sqrt{n}p(1-p)} \leq \frac{x - 2np(1-p)}{2\sqrt{n}p(1-p)}\right)\Bigg|_{x=F_{T_n}^{-1}(\alpha)} \\ &\approx \Phi\left(\frac{F_{T_n}^{-1}(\alpha) - 2np(1-p)}{2\sqrt{n}p(1-p)}\right) \\ \Rightarrow z_\alpha &\approx \frac{F_{T_n}^{-1}(\alpha) - 2np(1-p)}{2\sqrt{n}p(1-p)} \end{aligned}$$

Damit erhalten wir für die Quantile:

$$F_{T_n}^{-1}(\alpha) \approx 2np(1-p) + 2\sqrt{n}p(1-p) \cdot z_\alpha$$

In der Praxis setzt man $\hat{p} = \frac{n_1}{n_1+n_2}$ für p ein.

3 Lineare Regression

In Statistik I betrachteten wir die einfache lineare Regression der Form

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = 1, \ldots, n.$$

In Matrix-Form schreiben wir $Y = X\beta + \varepsilon$, wobei $Y = (Y_1, \ldots, Y_n)^\top$ der Vektor der Zielzufallsvariablen ist,

$$X = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}$$

eine $(n \times 2)$-Matrix, die die Ausgangsvariablen $x_i, i = 1, \ldots, n$ enthält und deshalb *Design-Matrix* genannt wird, $\beta = (\beta_0, \beta_1)^\top$ der Parametervektor und $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)^\top$ der Vektor der Störgrößen. Bisher waren oft $\varepsilon_i \sim N(0, \sigma^2)$ für $i = 1, \ldots, n$ und $\varepsilon \sim N(0, \mathcal{I} \cdot \sigma^2)$ multivariat normalverteilt.

Die multivariate (das bedeutet, nicht einfache) lineare Regression lässt eine beliebige $(n \times m)$-Design-Matrix

$$X = (x_{ij})_{\substack{i=1,\ldots,n \\ j=1,\ldots,m}}$$

und einen m-dimensionalen Parametervektor $\beta = (\beta_1, \ldots, \beta_m)^\top$ zu, für $m \geq 2$. Das heißt, es gibt

$$Y = X\beta + \varepsilon, \tag{3.0.1}$$

wobei $\varepsilon \sim N(0, K)$ ein multivariat normalverteilter Zufallsvektor der Störgrößen mit Kovarianzmatrix K ist, die im Allgemeinen nicht unabhängig voneinander sind, das heißt

$$K \neq \operatorname{diag}\left(\sigma_1^2, \ldots, \sigma_n^2\right).$$

Das Ziel dieses Kapitels ist es, Schätzer und Tests für β zu entwickeln. Zuvor müssen jedoch die Eigenschaften der multivariaten Normalverteilung untersucht werden.

3.1 Multivariate Normalverteilung

Im Vorlesungsskript Wahrscheinlichkeitsrechnung wurde die multivariate Normalverteilung in Beispiel 3.4.1 folgendermaßen eingeführt:

Definition 3.1.1. Es sei $X = (X_1, \ldots, X_n)^\top$ ein n-dimensionaler Zufallsvektor, $\mu \in \mathbb{R}^n$, K eine symmetrische, positiv definite $(n \times n)$-Matrix. X ist multivariat normalverteilt mit den Parametern μ und K ($X \sim N(\mu, K)$), falls X absolut stetig verteilt ist mit der Dichte

$$f_X(x) = \frac{1}{(2\pi)^{n/2}} \frac{1}{\sqrt{\det(K)}} \exp\left\{-\frac{1}{2}(x-\mu)^\top K^{-1}(x-\mu)\right\},\ x = (x_1, \ldots, x_n)^\top \in \mathbb{R}^n.$$

Wir geben drei weitere Definitionen von $N(\mu, K)$ an und wollen die Zusammenhänge zwischen ihnen untersuchen:

Definition 3.1.2. Der Zufallsvektor $X = (X_1, \ldots, X_n)^\top$ ist multivariat normalverteilt ($X \sim N(\mu, K)$) mit Parametern $\mu \in \mathbb{R}^n$ und K (eine symmetrische, nicht-negativ definite $(n \times n)$-Matrix), falls die charakteristische Funktion $\varphi_X(t) = \mathbb{E}\, e^{i(t,X)}$, $t \in \mathbb{R}^n$, gegeben ist durch

$$\varphi_X(t) = \exp\left\{it^\top \mu - \frac{1}{2} t^\top K t\right\}, \quad t \in \mathbb{R}^n.$$

Definition 3.1.3. Der Zufallsvektor $X = (X_1, \ldots, X_n)^\top$ ist multivariat normalverteilt ($X \sim N(\mu, K)$) mit Parametern $\mu \in \mathbb{R}^n$ und einer symmetrischen, nicht negativ definiten $(n \times n)$-Matrix K, falls

$$\forall a \in \mathbb{R}^n: \text{ die Zufallsvariable } (a, X) = a^\top X \sim N(a^\top \mu, a^\top K a)$$

eindimensional normalverteilt ist.

Definition 3.1.4. Es sei $\mu \in \mathbb{R}^n$, K eine nicht-negativ definite, symmetrische $(n \times n)$-Matrix. Ein Zufallsvektor $X = (X_1, \ldots, X_n)^\top$ ist multivariat normalverteilt mit Parametern μ und K ($X \sim N(\mu, K)$), falls

$$X \stackrel{d}{=} \mu + C \cdot Y,$$

wobei C eine $(n \times m)$ - Matrix mit $\operatorname{rang}(C) = m$, $K = C \cdot C^\top$ und $Y \sim N(0, \mathcal{I}) \in \mathbb{R}^m$ ein m-dimensionaler Vektor mit unabhängigen und identisch verteilten Koordinaten $Y_j \sim N(0,1)$ ist, $j = 1, \ldots, m$.

Bemerkung: Dies ist das Analogon im eindimensionalen Fall: $Y \sim N(\mu, \sigma^2) \Leftrightarrow Y = \mu + \sigma X$ mit $X \sim N(0,1)$.

Übungsaufgabe 3.1.1. Prüfen Sie, daß die in Definition 3.1.1 angegebene Dichte

$$f_X(x) = \frac{1}{(2\pi)^{n/2}} \frac{1}{\sqrt{\det(K)}} \exp\left\{-\frac{1}{2}(x-\mu)^\top K^{-1}(x-\mu)\right\},\ x \in \mathbb{R}^n$$

tatsächlich eine Verteilungsdichte darstellt.

Lemma 3.1.1. Es seien X und Y n-dimensionale Zufallsvektoren mit charakteristischen Funktionen

$$\varphi_X(t) = \mathbb{E}\, e^{i(t,X)} = \mathbb{E}\, e^{it^\top X}$$
$$\varphi_Y(t) = \mathbb{E}\, e^{i(t,Y)} = \mathbb{E}\, e^{it^\top Y}$$

für $t \in \mathbb{R}^n$. Es gelten folgende Eigenschaften:

1. *Eindeutigkeitssatz:*

$$X \stackrel{d}{=} Y \Leftrightarrow \varphi_X(t) = \varphi_Y(t), \quad t \in \mathbb{R}^n$$

2. Falls X und Y unabhängig sind, dann gilt:

$$\varphi_{X+Y}(t) = \varphi_X(t) \cdot \varphi_Y(t), \quad t \in \mathbb{R}^n.$$

ohne Beweis: vergleiche den Beweis des Satzes 5.1.1 (5), Folgerung 5.1.1, Vorlesungsskript WR.

Satz 3.1.1. 1. Die Definitionen 3.1.2 - 3.1.4 der multivariaten Normalverteilung sind äquivalent.

2. Die Definitionen 3.1.1 und 3.1.4 sind im Falle $n = m$ äquivalent.

Bemerkung 3.1.1. 1. Falls die Matrix K in Definition 3.1.4 den vollen Rang n besitzt, so besitzt sie die Dichte aus Definition 3.1.1. Sie wird in dem Fall *regulär* genannt.

2. Falls $\text{Rang}(K) = m < n$, dann ist die Verteilung $N(\mu, K)$ laut Definition 3.1.4 auf dem m-dimensionalen linearen Unterraum

$$\{y \in \mathbb{R}^n : y = \mu + Cx,\, x \in \mathbb{R}^m\}$$

konzentriert. $N(\mu, K)$ ist in diesem Fall offensichtlich nicht absolutstetig verteilt und wird daher *singulär* genannt.

Beweis. Wir beweisen: Definition 3.1.3 $\Leftrightarrow$ 3.1.2 $\Leftrightarrow$ 3.1.4.

1. a) Wir zeigen: Die Definitionen 3.1.2 und 3.1.3 sind äquivalent. Dazu ist zu zeigen: Für die Zufallsvariable X mit der charakteristischen Funktion

$$\varphi_X(t) = \exp\{it^\top \mu - \frac{1}{2} t^\top K t\} \Leftrightarrow \forall a \in \mathbb{R}^n : a^\top X \sim N(a^\top \mu, a^\top K a).$$

Es gilt:

$$\varphi_{t^\top X}(1) = \mathbb{E}\, e^{it^\top X \cdot 1} \stackrel{\varphi_{N(\mu,\sigma^2)}}{=} \exp\{it^\top \mu - \frac{1}{2} t^\top K t\} = \varphi_X(t) \quad \forall t \in \mathbb{R}.$$

(Dies nennt man das *Verfahren von Cramér-Wold*, vergleiche den multivariaten zentralen Grenzwertsatz).

b) Wir zeigen: Die Definitionen 3.1.3 und 3.1.4 sind äquivalent. Dazu ist zu zeigen: $X = \mu + C \cdot Y$ (mit μ, C, und Y wie in Definition 3.1.4) $\Leftrightarrow \varphi_X(t) = \exp\{it^\top \mu - \frac{1}{2} t^\top K t\}$, wobei $K = C \cdot C^\top$. Es gilt:

$$\begin{aligned}
\varphi_{\mu+CY}(t) &= \mathbb{E}\, e^{i(t,\mu+CY)} = \mathbb{E}\, e^{it^\top\mu + it^\top CY} = e^{it^\top\mu} \cdot \mathbb{E}\, e^{i(\overbrace{C^\top t}^{y},Y)} \\
&\overset{Y\sim N(0,\mathcal{I})}{=} e^{it^\top\mu} \cdot \exp\left(-\frac{1}{2} y^\top \cdot y\right) = \exp\left\{it^\top\mu - \frac{1}{2} t^\top C \cdot C^\top t\right\} \\
&= \exp\left\{it^\top\mu - \frac{1}{2} t^\top K t\right\},\ t \in \mathbb{R}^n.
\end{aligned}$$

2. Zu zeigen ist: Aus $X \sim N(\mu, K)$ im Sinne von Definition 3.1.4, $Y \sim N(\mu, K)$ im Sinne der Definition 3.1.1, $\mathrm{Rang}(K) = n$ folgt, daß $\varphi_X = \varphi_Y$.

   Aus der Definition 3.1.2 (die äquivalent zu Definition 3.1.4 ist) folgt, daß

$$\varphi_X(t) = \exp\left\{it^\top\mu - \frac{1}{2} t^\top K t\right\}, \quad t \in \mathbb{R}^n,$$

$$\begin{aligned}
\varphi_Y(t) &= \mathbb{E}\, e^{it^\top Y} = \int_{\mathbb{R}^n} e^{it^\top y} \frac{1}{(2\pi)^{n/2}} \frac{1}{\sqrt{\det K}} \cdot \exp\Big\{-\frac{1}{2} \overbrace{(y-\mu)}^{x}{}^\top K^{-1} \overbrace{(y-u)}^{x}\Big\} dy \\
&= e^{it^\top\mu} \cdot \int_{\mathbb{R}^n} \frac{1}{(2\pi)^{n/2}\sqrt{\det K}} \cdot \exp\left\{it^\top x - \frac{1}{2} x^\top K^{-1} x\right\} dx
\end{aligned}$$

   Wir diagonalisieren K : $\exists$ orthogonale $(n \times n)$-Matrix $V : V^\top = V^{-1}$ und $V^\top K V = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$, wobei $\lambda_i > 0$, $i = 1, \ldots, n$. Mit der neuen Substitution: $x = Vz$, $t = Vs$ erhalten wir:

$$\begin{aligned}
\varphi_Y(t) &= \frac{e^{it^\top\mu}}{(2\pi)^{n/2}\sqrt{\det K}} \cdot \int_{\mathbb{R}^n} \exp\left\{is^\top V^\top V z - \frac{1}{2} z^\top V^\top K^{-1} V z\right\} dz \\
&= \frac{e^{it^\top\mu}}{\sqrt{(2\pi)^n \lambda_1 \cdot \ldots \cdot \lambda_n}} \cdot \int_{\mathbb{R}} \ldots \int_{\mathbb{R}} \exp\left\{is^\top z - \frac{1}{2} \sum_{i=1}^n \frac{z_i^2}{\lambda_i}\right\} dz_1 \ldots dz_n \\
&= e^{it^\top\mu} \prod_{i=1}^n \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi\lambda_i}} e^{is_i z_i - \frac{z_i^2}{(2\lambda_i)}} dz_i = e^{it^\top\mu} \cdot \prod_{i=1}^n \varphi_{N(0,\lambda_i)}(s_i) = e^{it^\top\mu} \prod_{i=1}^n e^{\frac{-s_i^2\lambda_i}{2}} \\
&= \exp\left\{it^\top\mu - \frac{1}{2} s^\top \mathrm{diag}(\lambda_1, \ldots, \lambda_n) s\right\} = \exp\left\{it^\top\mu - \frac{1}{2} (V^\top t)^\top V^\top K V V^\top t\right\} \\
&= \exp\left\{it^\top\mu - \frac{1}{2} t^\top \underbrace{VV^\top}_{\mathcal{I}} K \underbrace{VV^\top}_{\mathcal{I}} t\right\} = \exp\left\{it^\top\mu - \frac{1}{2} t^\top K t\right\},\ t \in \mathbb{R}^n.
\end{aligned}$$

$\square$

3.1.1 Eigenschaften der multivariaten Normalverteilung

Satz 3.1.2. Es sei $X = (X_1, \ldots, X_n) \sim N(\mu, K)$, $\mu \in \mathbb{R}^n$, K symmetrisch und nicht-negativ definit. Dann gelten folgende Eigenschaften:

1. μ ist der *Erwartungswertvektor* von X:

$$\mathbb{E}\, X = \mu \quad \text{, das heißt: } \mathbb{E}\, X_i = \mu_i,\ i = 1, \ldots, n.$$

   K ist die *Kovarianzmatrix* von X:

$$K = (k)_{ij}, \text{ mit } k_{ij} = \operatorname{Cov}(X_i, X_j).$$

2. Jeder Teilvektor $X' = (X_{i_1}, \ldots, X_{i_k})^\top$ $(1 \leq i_1 < \ldots < i_k \leq n)$ von X ist ebenso multivariat normalverteilt, $X' \sim N(\mu', K')$, wobei $\mu' = (\mu_{i_1}, \ldots, \mu_{i_k})^\top$, $K' = (k'_{jl}) = (\operatorname{Cov}(X_{i_j}, X_{i_l})$, $j, l = 1, \ldots, k$. Insbesondere sind $X_i \sim N(\mu_i, k_{ii})$, wobei $k_{ii} = \operatorname{Var} X_i$, $i = 1, \ldots, n$.

3. Zwei Teilvektoren von X sind unabhängig, genau dann, wenn entsprechende Elemente k_{ij} von K, die ihre Kreuzkovarianzen darstellen, Null sind, das heißt: $X' = (X_1, \ldots, X_k)^\top$, $X'' = (X_{k+1}, \ldots, X_n)$ unabhängig (wobei die Reihenfolge nur wegen der Einfachheit so gewählt wurde, aber unerheblich ist) $\Leftrightarrow k_{ij} = 0$ für $1 \leq i \leq k$, $j > k$ oder $i > k$, $1 \leq j \leq k$.

$$K = \left(\begin{array}{c|c} K' & 0 \\ \hline 0 & K'' \end{array}\right)$$

   K' und K'' sind Kovarianzmatrizen von X' bzw. X''.

4. *Faltungsstabilität:* Falls X und Y unabhängige, n-dimensionale Zufallsvektoren mit $X \sim N(\mu_1, K_1)$ und $Y \sim N(\mu_2, K_2)$ sind, dann ist

$$X + Y \sim N(\mu_1 + \mu_2, K_1 + K_2).$$

Übungsaufgabe 3.1.2. Beweisen Sie Satz 3.1.2.

Satz 3.1.3 (*Lineare Transformation von* $N(\mu, K)$)**.** Sei $X \sim N(\mu, K))$ ein n-dimensionaler Zufallsvektor, A eine $(m \times n)$-Matrix mit $\operatorname{Rang}(A) = m \leq n$, $b \in \mathbb{R}^n$. Dann ist der Zufallsvektor $Y = AX + b$ multivariat normalverteilt:

$$Y \sim N(A\mu + b, AKA^\top).$$

Beweis. Ohne Beschränkung der Allgemeinheit setzen wir $\mu = 0$ und $b = 0$, weil $\varphi_{Y-a}(t) = e^{-it^\top a} \cdot \varphi_Y(t)$, für $a = A\mu + b$. Es ist zu zeigen:

$$Y = AX,\ X \sim N(0, K) \Rightarrow Y \sim N(0, AKA^\top)$$

Es ist

$$\begin{aligned}
\varphi_Y(t) &= \varphi_{AX}(t) = \mathbb{E}\, e^{it^\top AX} = \mathbb{E}\, e^{i(X,\overbrace{A^\top t}^{:=s})} \\
&\overset{\text{(Def. 3.1.2)}}{=} \exp\left\{-\frac{1}{2}s^\top K s\right\} = \exp\left\{-\frac{1}{2}t^\top AKA^\top t\right\},\ t \in \mathbb{R}^n \\
&\Rightarrow Y \sim N\left(0, AKA^\top\right).
\end{aligned}$$

□

3.1.2 Lineare und quadratische Formen von normalverteilten Zufallsvariablen

Definition 3.1.5. Seien $X = (X_1, \ldots, X_n)^\top$ und $Y = (Y_1, \ldots, Y_n)^\top$ Zufallsvektoren auf $(\Omega, \mathcal{F}, P)$, A eine $(n \times n)$-Matrix aus $\mathbb{R}^{n^2}$, die symmetrisch ist.

1. $Z = AX$ heißt *lineare Form* von X mit Matrix A.
2. $Z = Y^\top AX$ heißt *bilineare Form* von X und Y mit Matrix A.

$$Z = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij} X_j Y_i$$

3. Die Zufallsvariable $Z = X^\top AX$ (die eine bilineare Form aus 2. mit $Y = X$ ist) heißt *quadratische Form* von X mit Matrix A.

Satz 3.1.4. Sei $Z = Y^\top AX$ eine bilineare Form von Zufallsvektoren $X, Y \in \mathbb{R}^n$ bzgl. der symmetrischen Matrix A. Falls $\mu_x = \mathbb{E}\, X$, $\mu_Y = \mathbb{E}\, Y$ und $K_{XY} = (\operatorname{Cov}(X_i, Y_j))_{i,j=1,\ldots,n}$ die Kreuzkovarianzmatrix von X und Y ist, dann gilt:

$$\mathbb{E}\, Z = \mu_Y^\top A \mu_X + \operatorname{Spur}(AK_{XY}).$$

Beweis.

$$\begin{aligned}
\mathbb{E}\, Z &= \mathbb{E}\operatorname{Spur}(Z) = \mathbb{E}\operatorname{Spur}(Y^\top AX) \quad (\text{wegen } \operatorname{Spur}(AB) = \operatorname{Spur}(BA)) \\
&= \mathbb{E}\operatorname{Spur}(AXY^\top) = \operatorname{Spur}(A\mathbb{E}\,(XY^\top)) \quad (\text{wobei } XY^\top = (X_iY_j)_{i,j=1,\ldots,n}) \\
&= \operatorname{Spur}\left(A\mathbb{E}\left((X-\mu_X)\cdot(Y-\mu_Y)^\top + \mu_X Y^\top + X\mu_Y^\top - \mu_X\mu_Y^\top\right)\right) \\
&= \operatorname{Spur}\left(A(K_{XY} + \mu_X\mu_Y^\top + \mu_X\mu_Y^\top - \mu_X\mu_Y^\top)\right) = \operatorname{Spur}\left(AK_{XY} + A\mu_X\mu_Y^\top\right) \\
&= \operatorname{Spur}(AK_{XY}) + \operatorname{Spur}\left(A\mu_X \cdot \mu_Y^\top\right) \\
&= \operatorname{Spur}\left(\mu_Y^\top A\mu_X\right) + \operatorname{Spur}\left(AK_{XY}\right) = \mu_Y^\top A\mu_X + \operatorname{Spur}\left(AK_{XY}\right).
\end{aligned}$$

□

Folgerung 3.1.1. Für quadratische Formen gilt

$$\mathbb{E}\,(X^\top AX) = \mu_X^\top A\mu_X + \mathrm{Spur}(A\cdot K),$$

wobei $\mu_X = \mathbb{E}\,X$ und K die Kovarianzmatrix von X ist.

Satz 3.1.5 (*Kovarianz quadratischer Formen*)**.** Es sei $X \sim N(\mu, K)$ ein n-dimensionaler Zufallsvektor und $A, B \in \mathbb{R}^{n^2}$ zwei symmetrische $(n\times n)$-Matrizen. Dann gilt Folgendes:

$$\mathrm{Cov}\left(X^\top AX, X^\top BX\right) = 4\mu^\top AB\mu + 2\cdot\mathrm{Spur}(AKBK).$$

Lemma 3.1.2 (gemischte Momente)**.** Es sei $Y = (Y_1,\dots,Y_n)^\top \sim N(0,K)$ ein Zufallsvektor. Dann gilt Folgendes:

$$\begin{aligned}\mathbb{E}\,(Y_iY_jY_k) &= 0,\\ \mathbb{E}\,(Y_iY_jY_kY_l) &= k_{ij}\cdot k_{kl} + k_{ik}\cdot k_{jl} + k_{jk}\cdot k_{il}, \quad 1\le i,j,k,l\le n,\end{aligned}$$

wobei $K = (k_{ij})_{i,j=1,\dots,n}$ die Kovarianzmatrix von Y ist.

Übungsaufgabe 3.1.3. Beweisen Sie dieses Lemma.

Beweis von Satz 3.1.5.

$$\begin{aligned}
\mathrm{Cov}(X^\top AX, X^\top BX) &= \mathbb{E}\left(X^\top AX\cdot X^\top BX\right) - \mathbb{E}\left(X^\top AX\right)\cdot\mathbb{E}\left(X^\top BX\right)\\
&\overset{\text{(Folgerung 3.1.1)}}{=} \mathbb{E}\left((\overbrace{X-\mu}^{:=Y}+\mu)^\top A(\overbrace{X-\mu}^{=Y}+\mu)\cdot(\overbrace{X-\mu}^{=Y}+\mu)^\top B(\overbrace{X-\mu}^{=Y}+\mu)\right)\\
&\quad - \left(\mu^\top A\mu + \mathrm{Spur}(AK)\right)\left(\mu^\top B\mu + \mathrm{Spur}(BK)\right)\\
&= \mathbb{E}\left[\left(Y^\top AY + 2\mu^\top AY + \mu^\top A\mu\right)\left(Y^\top BY + 2\mu^\top BY + Y^\top B\mu\right)\right]\\
&\quad - \mu^\top A\mu\cdot\mu^\top B\mu - \mu^\top A\mu\cdot\mathrm{Spur}(BK) - \mu^\top B\mu\cdot\mathrm{Spur}(AK)\\
&\quad - \mathrm{Spur}(AK)\cdot\mathrm{Spur}(BK)\\
&= \mathbb{E}\left(Y^\top AY\cdot Y^\top BY\right) + 2\mathbb{E}\left(Y^\top AY\cdot\mu^\top BY\right) + \mathbb{E}\left(Y^\top AY\right)\cdot\mu^\top B\mu\\
&\quad + 2\mathbb{E}\left(\mu^\top AY\cdot Y^\top BY\right) + 4\mathbb{E}\left(\mu^\top AY\cdot\mu^\top BY\right) + 2\underbrace{\mathbb{E}\left(\mu^\top AY\right)}_{=0}\mu^\top B\mu\\
&\quad + \mu^\top A\mu\cdot\mathbb{E}\left(Y^\top BY\right) + 2\mu^\top A\mu\cdot\underbrace{\mathbb{E}\,\mu^\top BY}_{=0}\mu^\top A\mu\cdot\mu^\top B\mu - \mu^\top A\mu\cdot\mu^\top B\mu\\
&\quad - \mu^\top A\mu\cdot\mathrm{Spur}(BK) - \mu^\top B\mu\cdot\mathrm{Spur}(AK) - \mathrm{Spur}(AK)\cdot\mathrm{Spur}(BK)\\
&= \mathbb{E}\left(Y^\top AY\cdot Y^\top BY\right) + 2\mu^\top B\overbrace{\mathbb{E}\left(Y\cdot Y^\top AY\right)}^{=0\text{ (Lemma 3.1.2)}} + \mu^\top B\mu\cdot\mathrm{Spur}(AK)\\
&\quad + 2\mu^\top A\overbrace{\mathbb{E}\left(Y\cdot Y^\top BY\right)}^{=0} + 4\mu^\top A\overbrace{\mathbb{E}\left(YY^\top\right)}^{=K}B\mu + \mu^\top A\mu\cdot\mathrm{Spur}(BK)\\
&\quad - \mu^\top A\mu\cdot\mathrm{Spur}(BK) - \mu^\top B\mu\cdot\mathrm{Spur}(AK) - \mathrm{Spur}(AK)\mathrm{Spur}(BK)\\
&= \mathbb{E}\left(Y^\top AY\cdot Y^\top BY\right) + 4\mu^\top AKB\mu - \mathrm{Spur}(AK)\cdot\mathrm{Spur}(BK).
\end{aligned}$$

Wegen

$$\mathbb{E}\left(Y^\top AY \cdot Y^\top BY\right) = \mathbb{E}\left(\sum_{i,j=1}^{n} a_{ij}Y_iY_j \cdot \sum_{k,l=1}^{n} b_{kl}Y_kY_l\right) = \sum_{i,j,k,l=1}^{n} a_{ij}b_{kl}\mathbb{E}\left(Y_iY_jY_kY_l\right)$$
$$\overset{\text{(Lemma 3.1.2)}}{=} \sum_{i,j,k,l=1}^{n} a_{ij}b_{kl}\left(k_{ij}\cdot k_{kl} + k_{ik}\cdot k_{jl} + k_{jk}\cdot k_{il}\right)$$
$$= \sum_{i,j=1}^{n} a_{ij}k_{ij} \cdot \sum_{k,l=1}^{n} b_{kl}\cdot k_{kl} + 2\sum_{i,j,k,l=1}^{n} a_{ij}\cdot k_{jl}\cdot b_{lk}\cdot k_{ki}$$
$$= 2\cdot \text{Spur}\,(AKBK) + \text{Spur}\,(AK)\cdot \text{Spur}\,(BK)$$

folgt:

$$\begin{aligned}&\text{Cov}\left(X^\top AX, X^\top BX\right)\\ &\quad = 2\cdot \text{Spur}\,(AKBK) + \text{Spur}\,(AK)\cdot\text{Spur}\,(BK) + 4\mu^\top AKB\mu\\ &\qquad - \text{Spur}\,(AK)\cdot\text{Spur}\,(BK) = 4\mu^\top AKB\mu + 2\cdot\text{Spur}(AKBK).\end{aligned}$$

□

Folgerung 3.1.2.

$$\text{Var}\left(X^\top AX\right) = 4\mu^\top AKA\mu + 2\cdot\text{Spur}\left((AK)^2\right)$$

Satz 3.1.6. Es sein $X \sim N(\mu, K)$ und $A, B \in \mathbb{R}^{n^2}$ zwei symmetrische Matrizen. Dann gilt:

$$\text{Cov}(BX, X^\top AX) = 2BKA\mu$$

Beweis.

$$\text{Cov}(BX, X^\top AX) \overset{\text{(Folgerung 3.1.1)}}{=} \mathbb{E}\left[(BX - B\mu)(X^\top AX - \mu^\top A\mu - \text{Spur}(AK))\right]$$
$$= \mathbb{E}\left[B(X-\mu)\left((X-\mu)^\top A(X-\mu) + 2\mu^\top AX - 2\mu^\top A\mu - \text{Spur}(AK)\right)\right],$$

denn

$$(X-\mu)^\top A(X-\mu) = X^\top AX - \mu^\top AX - X^\top A\mu + \mu^\top A\mu$$

und mit der Substitution $Z = X - \mu$ (und damit $\mathbb{E}\,Z = 0$)

$$\begin{aligned}\text{Cov}(BX, X^\top AX) &= \mathbb{E}\left[(BZ(Z^\top AZ + 2\mu^\top AZ - \text{Spur}(AK)))\right]\\ &= \mathbb{E}\,(BZ\cdot Z^\top AZ) + 2\mathbb{E}\,(BZ\cdot\mu^\top AZ) - \text{Spur}(AK)\cdot\overbrace{\mathbb{E}\,(BZ)}^{=B\mathbb{E}\,Z=0}\\ &= 2\mathbb{E}\,(BZ\cdot Z^\top A\mu) + \mathbb{E}\,(BZZ^\top AZ) = \underbrace{2B\mathbb{E}\,(ZZ^\top)}_{\text{Cov}X=K}A\mu\\ &\quad + B\cdot\underbrace{\mathbb{E}\,(ZZ^\top AZ)}_{=0} = 2BKA\mu,\end{aligned}$$

wegen $Z \sim N(0, K)$ und Lemma 3.1.2 und dem Beweis von Satz 3.1.5. □

Definition 3.1.6. Es seien $X_i \sim N(\mu_i, 1)$, $i = 1, \ldots, n$ unabhängig. Dann besitzt die Zufallsvariable

$$Y = X_1^2 + \ldots + X_n^2$$

die sogenannte *nicht-zentrale* $\chi^2_{n,\mu}$*-Verteilung* mit n *Freiheitsgraden* und dem *Nichtzentralitätsparameter*

$$\mu = \sum_{i=1}^n \mu_i^2.$$

(in Statistik I betrachteten wir den Spezialfall der zentralen χ^2_n-Verteilung mit $\mu = 0$).

In Bemerkung 5.2.1, Vorlesungsskript WR, haben wir momenterzeugende Funktionen von Zufallsvariablen eingeführt. Jetzt benötigen wir für den Beweis des Satzes 3.1.7 folgenden Eindeutigkeitssatz:

Lemma 3.1.3 (*Eindeutigkeitssatz für momenterzeugende Funktionen*)**.** Es seien X_1 und X_2 zwei absolutstetige Zufallsvariablen mit momenterzeugenden Funktionen

$$M_{X_i}(t) = \mathbb{E}\, e^{tX_i}, \quad i = 1, 2,$$

die auf einem Intervall (a, b) definiert sind. Falls f_1 und f_2 die Dichten der Verteilung von X_1 und X_2 sind, dann gilt

$$f_1(x) = f_2(x) \text{ für fast alle } x \in \mathbb{R} \Leftrightarrow M_{X_1}(t) = M_{X_2}(t),\ \forall t \in (a, b).$$

Ohne Beweis.

Satz 3.1.7. Die Dichte einer $\chi^2_{n,\mu}$-verteilten Zufallsvariable X (mit $n \in \mathbb{N}$ und $\mu > 0$) ist gegeben durch die Mischung der Dichten von χ^2_{n+2J}-Verteilungen mit Mischungsvariable $J \sim \text{Poisson}(\mu/2)$:

$$f_X(x) = \begin{cases} \sum\limits_{j=0}^{\infty} e^{-\mu/2} \frac{(\mu/2)^j}{j!} \cdot \frac{e^{-x/2} x^{\frac{n+2j}{2}-1}}{\Gamma\left(\frac{n+2j}{2}\right) \cdot 2^{\frac{n+2j}{2}}}, & x \geq 0, \\ 0, & x < 0. \end{cases} \tag{3.1.1}$$

Beweis. 1. Wir berechnen zuerst $M_X(t)$, $X \sim \chi^2_{n,\mu}$:

$$\begin{aligned} M_X(t) &= \mathbb{E}\,(e^{tX}) = \mathbb{E}\, \exp\left\{ t \sum_{i=1}^n X_i^2 \right\} \\ &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi}} \cdot \int_{-\infty}^{\infty} e^{tx_i^2} \cdot e^{-\frac{(x_i - \mu_i)^2}{2}} dx_i \qquad \left(t < \frac{1}{2},\ X_i \sim N(\mu_i, 1) \right) \end{aligned}$$

Es gilt:

$$\begin{aligned}
tx_i^2 - \frac{(x_i-\mu_i)^2}{2} &= \frac{1}{2}(2tx_i^2 - x_i^2 + 2x_i\mu_i - \mu_i^2)\\
&= -\frac{1}{2}\left(x_i^2(1-2t) - 2x_i\mu_i + \frac{\mu_i^2}{(1-2t)} - \frac{\mu_i^2}{(1-2t)} + \mu_i^2\right)\\
&= -\frac{1}{2}\left(\left(x_i\cdot\sqrt{1-2t} - \frac{\mu}{\sqrt{1-2t}}\right)^2 + \mu_i^2\left(1-\frac{1}{1-2t}\right)\right)\\
&= -\frac{1}{2}\left(\frac{(x_i(1-2t)-\mu_i)^2}{1-2t} - \mu_i^2\cdot\frac{2t}{1-2t}\right)
\end{aligned}$$

Wir substituieren

$$y_i = \frac{(x_i\cdot(1-2t)-\mu_i)}{\sqrt{1-2t}}$$

und erhalten

$$\begin{aligned}
M_X(t) &= (1-2t)^{-\frac{n}{2}}\prod_{i=1}^n \exp\left\{\mu_i^2\cdot\left(\frac{t}{1-2t}\right)\right\}\cdot\underbrace{\frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{\infty} e^{-\frac{y_i^2}{2}}\,dy_i}_{=1}\\
&= (1-2t)^{-\frac{n}{2}}\cdot\exp\left\{\frac{t}{1-2t}\cdot\sum_{i=1}^n\mu_i^2\right\} = \frac{1}{(1-2t)^{n/2}}\cdot\exp\left\{\frac{\mu t}{1-2t}\right\},\quad t<\frac{1}{2}.
\end{aligned}$$

2. Es sei Y eine Zufallsvariable mit der Dichte (3.1.1). Wir berechnen $M_Y(t)$:

$$\begin{aligned}
M_Y(t) &= \sum_{j=0}^\infty e^{-\frac{\mu}{2}}\frac{(\mu/2)^j}{j!}\cdot\underbrace{\int\limits_0^\infty e^{xt}\cdot\frac{e^{-\frac{x}{2}}\cdot x^{\frac{n+2j}{2}-1}}{\Gamma\left(\frac{n+2j}{2}\right)\cdot\frac{n+2j}{2}}}_{=M_{\chi^2_{n+2j}}(t)=\frac{1}{(1-2t)^{(n+2j)/2}}\ \text{(Statistik I, Satz 3.2.1)}}\\
&= \frac{e^{-\frac{\mu}{2}}}{(1-2t)^{\frac{n}{2}}}\cdot\sum_{j=1}^\infty\left(\frac{\mu}{2(1-2t)}\right)^j\cdot\frac{1}{j!}\\
&= \frac{1}{(1-2t)^{\frac{n}{2}}}\cdot\exp\left\{-\frac{\mu}{2}+\frac{\mu}{2(1-2t)}\right\} = \frac{1}{(1-2t)^{\frac{n}{2}}}\cdot\exp\left\{\frac{\mu\cdot(1-(1-2t))}{2\cdot(1-2t)}\right\}\\
&= (1-2t)^{-\frac{n}{2}}\cdot\exp\left\{\frac{\mu t}{1-2t}\right\}\\
&\Longrightarrow\quad M_X(t) = M_Y(t),\quad \forall t<\frac{1}{2}
\end{aligned}$$

Nach Lemma 3.1.3 gilt dann, $f_X(x) = f_Y(x)$ für fast alle $x\in\mathbb{R}$.

$\square$

Bemerkung 3.1.2. 1. Die Definition 3.1.6 kann in folgender Form umgeschrieben werden:

Falls $X \sim N(\vec{\mu}, \mathcal{I})$, $\vec{\mu} = (\mu_1, \dots, \mu_n)^\top$, dann gilt $|X|^2 = X^\top X \sim \chi^2_{n,\mu}$, wobei $\mu = |\vec{\mu}|^2$.

2. Die obige Eigenschaft kann auf $X \sim N(\vec{\mu}, K)$, mit einer symmetrischen, positiv definiten $(n \times n)$-Matrix K verallgemeinert werden:

$$X^\top K^{-1} X \sim \chi^2_{n,\tilde{\mu}}, \quad \text{wobei } \tilde{\mu} = \vec{\mu}^\top K^{-1} \vec{\mu},$$

denn weil K positiv definit ist, gibt es ein $K^{\frac{1}{2}}$, sodaß $K = K^{\frac{1}{2}} K^{\frac{1}{2}\top}$. Dann gilt

$$Y = K^{-\frac{1}{2}} X \sim N(K^{-\frac{1}{2}}\mu, \mathcal{I}), \text{ weil } K^{-\frac{1}{2}} K K^{-\frac{1}{2}\top} = K^{-\frac{1}{2}} \cdot K^{\frac{1}{2}} \cdot K^{\frac{1}{2}\top} \cdot K^{-\frac{1}{2}\top} = \mathcal{I}$$

und daher

$$Y^\top Y \overset{\text{Punkt 1}}{\sim} \chi^2_{n,\tilde{\mu}}, \quad \text{mit } \tilde{\mu} = \left(K^{-\frac{1}{2}}\vec{\mu}\right)^\top K^{-\frac{1}{2}}\vec{\mu} = \vec{\mu}^\top K^{-\frac{1}{2}\top} K^{-\frac{1}{2}} \vec{\mu} = \vec{\mu}^\top K^{-1} \vec{\mu}.$$

Satz 3.1.8. Es sei $X \sim N(\mu, K)$, wobei K eine symmetrische, positiv definite $(n \times n)$-Matrix ist, und sei A eine weitere symmetrische $(n \times n)$-Matrix mit der Eigenschaft $AK = (AK)^2$ (Idempotenz) und $\mathrm{Rang}(A) = r \le n$. Dann gilt:

$$X^\top A X \sim \chi^2_{r,\tilde{\mu}}, \text{ wobei } \tilde{\mu} = \mu^\top A \mu.$$

Beweis. Wir zeigen, daß A nicht negativ definit ist.

$$\begin{aligned}
AK = (AK)^2 = AK \cdot AK \quad & | \, K^{-1} \\
\Longrightarrow A = AKA \Rightarrow \forall x \in \mathbb{R}^n : \; & x^\top A x = x^\top AKA x \\
= (\underbrace{Ax}_{=y})^\top K (\underbrace{Ax}_{=y}) \ge 0 & \text{ wegen der positiven Definitheit von } K. \\
\Longrightarrow & \text{ A ist nicht negativ definit.} \\
\Longrightarrow \exists H : & \text{ eine } (n \times n)\text{-Matrix mit } \mathrm{Rang}(H) = r : \; A = HH^\top
\end{aligned}$$

Somit gilt

$$X^\top A X = X^\top H \cdot H^\top X = (\underbrace{H^\top X}_{=Y})^\top \cdot H^\top X = Y^\top Y$$

Es gilt: $Y \sim N(H^\top \mu, \mathcal{I})$, denn nach Satz 3.1.3 ist $Y \sim N(H^\top \mu, H^\top K H)$ und $\mathrm{Rang}(H) = r$. Das heißt, $H^\top H$ ist eine invertierbare $(n \times n)$-Matrix, und

$$\begin{aligned}
H^\top K H &= (H^\top H)^{-1} (H^\top \underbrace{H \cdot H^\top K H \cdot (H^\top}_{=AKA=A} H)(H^\top H)^{-1} \\
&= (H^\top H)^{-1} H^\top \cdot \underbrace{A}_{=HH^T} \cdot H (H^\top H)^{-1} \\
&= \mathcal{I}
\end{aligned}$$

Dann ist

$$X^\top AX = |\, Y \,|^2 \sim \chi^2_{r,\tilde{\mu}} \text{ mit } \tilde{\mu} = (H^\top \mu)^2 = \mu^\top H \cdot H^\top \mu = \mu^\top A\mu.$$

□

Satz 3.1.9 (Unabhängigkeit)**.** Es sei $X \sim N(\mu, K)$ und K eine symmetrische, nicht-negativ definite $(n \times n)$-Matrix.

1. Es seien A, B $(r_1 \times n)$ bzw. $(r_2 \times n)$-Matrizen, $r_1, r_2 \leq n$ mit $AKB^\top = 0$. Dann sind die Vektoren AX und BX unabhängig.
2. Sei ferner C eine symmetrische, nicht-negativ definite $(n \times n)$-Matrix mit der Eigenschaft $AKC = 0$. Dann sind AX und $X^\top CX$ unabhängig.

Beweis. 1. Nach Satz 3.1.2, 3) gilt: AX und BX sind unabhängig $\Longleftrightarrow \varphi_{(AX,BX)}(t) = \varphi_{AX}(t) \cdot \varphi_{BX}(t)$, $t = (t_1, t_2)^\top \in \mathbb{R}^{r_1+r_2}$, $t_1 \in \mathbb{R}^{r_1}$, $t_2 \in \mathbb{R}^{r_2}$. Es ist zu zeigen:

$$\varphi_{(AX,BX)}(t) = \mathbb{E}\, e^{\left(it_1^\top A + t_2^\top B\right)\cdot X} \stackrel{!}{=} \mathbb{E}\, e^{it_1^\top AX} \cdot \mathbb{E}\, e^{it_2^\top BX}$$

Es gilt

$$\varphi_{(AX,BX)}(t) = \mathbb{E}\, e^{i\left(t_1^\top A + t_2^\top B\right)\cdot X} \stackrel{(Def. 3.1.2)}{=} e^{i\left(t_1^\top A + t_2^\top B\right)\cdot \mu - \frac{1}{2}\cdot\left(t_1^\top A + t_2^\top B\right)\cdot K \cdot \left(t_1^\top A + t_2^\top B\right)^\top},$$

und mit

$$\begin{aligned}
&\left(t_1^\top A + t_2^\top B\right) \cdot K \cdot \left(t_1^\top A + t_2^\top B\right)^\top \\
&= \left(t_1^\top A\right) K \left(t_1^\top A\right)^\top + \left(t_1^\top A\right)^\top K \left(t_2^\top B\right) + \left(t_2^\top B\right) K \left(t_1^\top A\right)^\top + \left(t_2^\top B\right) K \left(t_2^\top B\right)^\top \\
&= t_1^\top AKA^\top t_1 + t_1^\top \cdot \underbrace{AKB^\top}_{=0} \cdot t_2 + t_2^\top \cdot \underbrace{BKA^\top}_{=(AKB^\top)^\top = 0} \cdot t_1 + t_2^\top BKB^\top t_2
\end{aligned}$$

ist

$$\begin{aligned}
\varphi_{(AX,BX)}(t) &= e^{it^\top A - \frac{1}{2} t_1^\top AKA^\top t_1} \cdot e^{it_2^\top B - \frac{1}{2} t_2^\top BKB^\top t_2} \\
&= \varphi_{AX}(t_1) \cdot \varphi_{BX}(t_2), \quad t_1 \in \mathbb{R}^{r_1},\ t_2 \in \mathbb{R}^{r_2}
\end{aligned}$$

2. C ist symmetrisch, nicht-negativ definit $\Longrightarrow$ Es gibt eine $(n \times r)$-Matrix H mit $\text{Rang}(H) = r < n$ und $C = HH^\top$, $\Longrightarrow H^\top H$ hat Rang r und ist somit invertierbar. Dann gilt:

$$X^\top C = X^\top HH^\top X = (H^\top X)^\top \cdot H^\top X = |H^\top X|^2.$$

Falls AX und $H^\top X$ unabhängig sind, dann sind auch AX und $X^\top CX = |H^\top X|^2$ unabhängig, nach dem Transformationssatz für Zufallsvektoren. Nach 1) sind AX und $H^\top X$ unabhängig, falls $AK(H^\top)^\top = AKH = 0$. Da nach Voraussetzung

$$AKC = AKH \cdot H^\top = 0 \Longrightarrow AKH \cdot H^\top H = 0,$$

da aber $\exists (H^\top H)^{-1}$, folgt, daß

$$\begin{aligned} 0 = AKH \cdot H^\top H \cdot (H^\top H)^{-1} = AKH &\Longrightarrow AKH = 0 \\ \Longrightarrow AX &\text{ und } H^\top X \text{ sind unabhängig} \\ \Longrightarrow AX &\text{ und } X^\top CX \text{ sind unabhängig.} \end{aligned}$$

□

3.2 Multivariate lineare Regressionsmodelle mit vollem Rang

Die *multivariate lineare Regression* hat die Form

$$Y = X\beta + \varepsilon,$$

wobei $Y = (Y_1, \ldots, Y_n)^\top$ der Zufallsvektor der Zielvariablen ist,

$$X = (x_{ij})_{\substack{i=1,\ldots,n \\ j=1,\ldots,m}}$$

ist eine deterministische *Design-Matrix* mit vollem Rang, $\operatorname{Rang}(X) = r = m \le n$, $\beta = (\beta_1, \ldots, \beta_m)^\top$ ist der *Parametervektor* und $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)^\top$ ist der Zufallsvektor der *Störgrößen*, mit $\mathbb{E}\,\varepsilon_i = 0$, $\operatorname{Var}\varepsilon_i = \sigma^2 > 0$. Das Ziel dieses Abschnittes wird sein, β und σ^2 geeignet zu schätzen.

3.2.1 Methode der kleinsten Quadrate

Sei $X = (X_1, \ldots, X_m)$, wobei die deterministischen Vektoren $X_j = (x_{1j}, x_{2j}, \ldots, x_{nj})^\top$, $j = 1, \ldots, m$ einen m-dimensionalen linearen Unterraum $L_X =< X_1, \ldots, X_m >$ aufspannen. Sei

$$e(\beta) = \frac{1}{n}|Y - X\beta|^2 = \frac{1}{n}\sum_{i=1}^{n}(Y_i - x_{i1}\beta_1 - \ldots - x_{im}\beta_m)^2$$

die mittlere quadratische Abweichung zwischen Y und $X\beta$.

Der *MKQ-Schätzer* $\hat\beta$ für β ist definiert durch

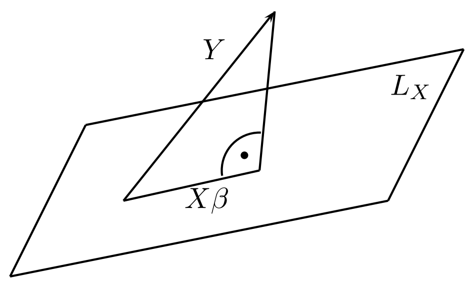
$$\hat\beta = \operatorname{argmin}(e(\beta)) \tag{3.2.1}$$

Warum existiert eine Lösung $\beta \in \mathbb{R}^m$ des quadratischen Optimierungsproblems (3.2.1)? Geometrisch kann $X\hat\beta$ als die orthogonale Projektion des Datenvektors Y auf den linearen Unterraum L_X interpretiert werden. Formal zeigen wir die Existenz der Lösung mit folgendem Satz.

Satz 3.2.1. Unter den obigen Voraussetzungen existiert der eindeutig bestimmte MKQ-Schätzer $\hat\beta$, der die Lösung der sogenannten *Normalengleichung* ist:

$$X^\top X\beta = X^\top Y. \tag{3.2.2}$$

Abbildung 3.1: Projektion auf den linearen Unterraum L_X



Daher gilt:

$$\hat{\beta} = \left(X^\top X\right)^{-1} X^\top Y$$

Beweis. Die notwendige Bedingung für die Existenz des Minimums ist $e'(\beta) = 0$, das heißt

$$e'(\beta) = \left(\frac{\partial e(\beta)}{\partial \beta_1}, \dots, \frac{\partial e(\beta)}{\partial \beta_m}\right)^\top = 0$$

Es gilt:

$$e'(\beta) = \frac{2}{n}\left(X^\top X\beta - X^\top Y\right)$$

$\Longrightarrow$ $\hat{\beta}$ ist eine Lösung der Normalengleichung $X^\top X\beta = X^\top Y$. Wir zeigen die hinreichende Bedingung des Minimums:

$$e''(\beta) = \left(\frac{\partial^2 e(\beta)}{\partial \beta_i \partial \beta_j}\right)_{i,j=1,\dots,m} = \frac{2}{n} X^\top X$$

$X^\top X$ ist symmetrisch und positiv definit, weil X einen vollen Rang hat:

$$\forall y \neq 0,\, y \in \mathbb{R}^m : \quad y^\top X^\top X y = (Xy)^\top X y = |Xy|^2 > 0$$
$$\Longrightarrow X^\top X \text{ ist invertierbar.}$$

und aus $y \neq 0 \Longrightarrow Xy \neq 0$, folgt, daß $e''(\beta) > 0$ wegen der positiven Definitheit. Das heißt, $\hat{\beta}$ ist der Minimumpunkt von $e(\beta)$. Den Schätzer $\hat{\beta} = \left(X^\top X\right)^{-1} X^\top Y$ bekommt man, indem man die Normalengleichung $X^\top X\beta = X^\top Y$ von links mit $\left(X^\top X\right)^{-1}$ multipliziert. $\square$

Beispiel 3.2.1. 1. *Einfache lineare Regression*

$$X = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \qquad m = 2,\ \beta = (\beta_1, \beta_2)^\top,\ Y = X\beta + \varepsilon$$

$\hat\beta = \left(\hat\beta_1, \hat\beta_2\right)$ ergibt den MKQ-Schätzer aus der Statistik I.

$$\hat\beta_2 = \frac{S^2_{XY}}{S^2_{XX}}, \quad \hat\beta_1 = \overline{Y}_n - \overline{X}_n \hat\beta_2,$$

wobei

$$\begin{aligned} \overline{X}_n &= \frac{1}{n}\sum_{i=1}^n X_i, \quad \overline{Y}_n = \frac{1}{n}\sum_{i=1}^n Y_i \\ S^2_{XY} &= \frac{1}{n-1}\sum_{i=1}^n \left(X_i - \overline{X}_n\right)\left(Y_i - \overline{Y}_n\right) \\ S^2_{XX} &= \frac{1}{n-1}\sum_{i=1}^n \left(X_i - \overline{X}_n\right)^2 \end{aligned}$$

Übungsaufgabe 3.2.1. Beweisen Sie dies!

2. *Multiple lineare Regression*
   $Y = X\beta + \varepsilon$ mit Designmatrix

$$X = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1m} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & \cdots & x_{nm} \end{pmatrix} \text{ für } \beta = (\beta_0, \beta_1, \ldots, \beta_m)^\top.$$

Der MKQ-Schätzer $\hat\beta = (X^\top X)^{-1} X^\top Y$ ist offensichtlich ein linearer Schätzer bezüglich Y.

Wir werden jetzt zeigen, daß $\hat\beta$ der *beste lineare, erwartungstreue Schätzer* von β (im Englischen *BLUE = best linear unbiased estimator*) in der Klasse

$$\mathcal{L} = \left\{\tilde\beta = AY + b : \mathbb{E}\,\tilde\beta = \beta\right\}$$

aller linearen erwartungstreuen Schätzer ist.

Satz 3.2.2 (*Güteeigenschaften des MKQ-Schätzers* $\hat\beta$)**.** Es sei $Y = X\beta + \varepsilon$ ein multivariates lineares Regressionsmodell mit vollem Rang m und Störgrößen $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)^\top$, die folgende Voraussetzungen erfüllen:

$$\mathbb{E}\,\varepsilon = 0, \quad \operatorname{Cov}(\varepsilon_i, \varepsilon_j) = \sigma^2 \delta_{ij},\ i,j = 1, \ldots, n \text{ für ein } \sigma^2 \in (0, \infty).$$

Dann gilt Folgendes:

1. Der MKQ-Schätzer $\hat{\beta} = \left(X^\top X\right)^{-1} X^\top Y$ ist erwartungstreu: $\mathbb{E}\,\hat{\beta} = \beta$.

2. $\mathrm{Cov}(\hat{\beta}) = \sigma^2 \left(X^\top X\right)^{-1}$

3. $\hat{\beta}$ besitzt die minimale Varianz:

$$\forall \tilde{\beta} \in \mathcal{L}: \quad \mathrm{Var}\,\tilde{\beta}_j \geq \mathrm{Var}\,\hat{\beta}_j, \quad j = 1, \ldots, m.$$

Beweis. 1. Es gilt:

$$\begin{aligned}
\mathbb{E}\,\hat{\beta} &= \mathbb{E}\left[\left(X^\top X\right)^{-1} X^\top (X\beta + \varepsilon)\right] \\
&= \left(X^\top X\right)^{-1} \cdot X^\top X \cdot \beta + \left(X^\top X\right)^{-1} X^\top \cdot \underbrace{\mathbb{E}\,\varepsilon}_{=0} \\
&= \beta \quad \forall \beta \in \mathbb{R}^m.
\end{aligned}$$

2. Für alle $\tilde{\beta} = AY + b \in \mathcal{L}$ gilt:

$$\begin{aligned}
&\beta = \mathbb{E}\,\tilde{\beta} = A\mathbb{E}\,Y + b = AX\beta + b \quad \forall \beta \in \mathbb{R}^m. \\
&\Longrightarrow b = 0, \quad AX = \mathcal{I}. \\
&\Longrightarrow \tilde{\beta} = AY = A\,(X\beta + \varepsilon) = AX\beta + A\varepsilon \\
&= \beta + A\varepsilon.
\end{aligned}$$

Für

$$\hat{\beta} = \underbrace{\left(X^\top X\right)^{-1} X^\top}_{=A} Y$$

gilt:

$$\begin{aligned}
\mathrm{Cov}\hat{\beta} &= \left(\mathbb{E}\left(\left(\hat{\beta}_i - \beta_i\right)\left(\hat{\beta}_j - \beta_j\right)\right)\right)_{i,j=1,\ldots,m} \\
&= \mathbb{E}\left(A\varepsilon \cdot (A\varepsilon)^\top\right) = \mathbb{E}\left(A\varepsilon\varepsilon^\top A^\top\right) = A\mathbb{E}\left(\varepsilon\varepsilon^\top\right) \cdot A^\top \\
&= A \cdot \sigma^2 \mathcal{I} A^\top = \sigma^2 AA^\top = \sigma^2 \left(X^\top X^{-1}\right) X^\top \left(\left(X^\top X\right)^{-1} X^\top\right)^\top \\
&= \sigma^2 \left(X^\top X\right)^{-1} X^\top X \left(X^\top X\right)^{-1} = \sigma^2 \left(X^\top X\right)^{-1}.
\end{aligned}$$

3. Sei $\tilde{\beta} \in \mathcal{L}$, $\tilde{\beta} = \beta + A\varepsilon$. Zu zeigen ist, daß

$$\left(\mathrm{Cov}(\tilde{\beta})\right)_{ii} = \sigma^2 (AA^\top)_{ii} \geq \left(\mathrm{Cov}(\hat{\beta})\right)_{ii} = \sigma^2 (X^\top X)^{-1}_{ii}, \quad i = 1, \ldots, m.$$

Sei $D = A - (X^\top X)^{-1}X^\top$, dann folgt: $A = D + (X^\top X)^{-1}X^\top$,

$$\begin{aligned} AA^\top &= \left(D + \left(X^\top X\right)^{-1} X^\top\right)\left(D^\top + X\left(X^\top X\right)^{-1\top}\right) \\ &= DD^\top + \left(X^\top X\right)^{-1}, \text{ weil} \end{aligned}$$

$$\begin{aligned} DX\left(X^\top X\right)^{-1} &= \Big(\underbrace{AX}_{=\mathcal{I}} - \underbrace{\left(X^\top X\right)^{-1} X^\top X}_{=\mathcal{I}}\Big)\left(X^\top X\right)^{-1} = 0 \\ \left(X^\top X\right)^{-1} X^\top D &= \left(X^\top X\right)^{-1} X^\top \left(A^\top - X\left(X^\top X\right)^{-1\top}\right) \\ &= \left(X^\top X\right)^{-1}\Big(\underbrace{(AX)^\top}_{=\mathcal{I}} - \underbrace{X^\top X\left(X^\top X\right)^{-1}}_{=\mathcal{I}}\Big) = 0. \end{aligned}$$

$$\begin{aligned} &\Longrightarrow \left(AA^\top\right)_{ii} = \underbrace{\left(DD^\top\right)_{ii}}_{\geq 0} + \left(X^\top X\right)^{-1}_{ii} \geq \left(X^\top X\right)^{-1}_{ii} \\ &\Longrightarrow \operatorname{Var}\hat{\beta}_i \leq \operatorname{Var}\tilde{\beta}_i, \quad i = 1, \ldots, m. \end{aligned}$$

□

Satz 3.2.3. Es sei $\hat{\beta}_n$ der MKQ-Schätzer im oben eingeführten multivariaten linearen Regressionsmodell. Sei $\{a_n\}_{n\in\mathbb{N}}$ eine Zahlenfolge mit $a_n \neq 0$, $n \in \mathbb{N}$, $a_n \to \infty$ $(n \to \infty)$. Es wird vorausgesetzt, daß eine invertierbare $(m \times m)$-Matrix Q existiert mit

$$Q = \lim_{n\to\infty} a_n\left(X_n^\top X_n\right).$$

Dann ist $\hat{\beta}_n$ schwach konsistent:

$$\hat{\beta}_n \xrightarrow[n\to\infty]{p} \beta.$$

Beweis.

$$\hat{\beta}_n \xrightarrow[n\to\infty]{p} \beta \Longleftrightarrow \mathbb{P}\left(\left|\hat{\beta}_n - \beta\right| > \varepsilon\right) \xrightarrow[n\to\infty]{} 0 \quad \forall \varepsilon > 0.$$

$$
\begin{aligned}
\mathbb{P}\left(\left|\hat{\beta}_n - \beta\right| > \varepsilon\right) &= \mathbb{P}\left(\left|\hat{\beta}_n - \beta\right|^2 > \varepsilon^2\right) \\
&= \mathbb{P}\left(\sum_{i=1}^{m}\left|\hat{\beta}_{in} - \beta_i\right|^2 > \varepsilon^2\right) \leq \mathbb{P}\left(\bigcup_{i=1}^{m}\left\{\left|\hat{\beta}_{in} - \beta_i\right|^2 > \frac{\varepsilon^2}{m}\right\}\right) \\
&\leq \sum_{i=1}^{m}\mathbb{P}\left(\left|\hat{\beta}_{in} - \beta_i\right| > \frac{\varepsilon}{\sqrt{m}}\right) \\
&\leq m\sum_{i=1}^{m}\frac{\operatorname{Var}\hat{\beta}_{in}}{\varepsilon^2} \underset{n\to\infty}{\longrightarrow} 0, \quad \text{(aus der Ungleichung von Tschebyschew)} \\
&\text{falls } \operatorname{Var}\hat{\beta}_{in} \underset{n\to\infty}{\longrightarrow} 0, \quad i = 1, \ldots, m.
\end{aligned}
$$

$\operatorname{Var}\hat{\beta}_{in}$ ist ein Diagonaleintrag von der Matrix

$$
\operatorname{Cov}\hat{\beta}_n \overset{(Satz 3.2.2)}{=} \sigma^2\left(X_n^\top X_n\right)^{-1}.
$$

Wenn wir zeigen, daß $\operatorname{Cov}\hat{\beta}_n \underset{n\to\infty}{\longrightarrow} 0$, ist der Satz bewiesen.

Es existiert

$$
Q^{-1} = \lim_{n\to\infty}\frac{1}{a_n}\left(X_n^\top X_n\right)^{-1}
$$

und damit gilt:

$$
\begin{aligned}
\lim_{n\to\infty}\operatorname{Cov}\hat{\beta}_n &= \sigma^2\lim_{n\to\infty}\left(X_n^\top X_n\right)^{-1} = \sigma^2\lim_{n\to\infty}a_n\cdot\frac{1}{a_n}\left(X_n^\top X_n\right)^{-1} \\
&= 0\cdot Q^{-1}\cdot\sigma^2 = 0.
\end{aligned}
$$

□

3.2.2 Schätzer der Varianz σ^2

Wir führen den Schätzer $\hat{\sigma}^2$ für die Varianz σ^2 der Störgrößen ε_i folgendermaßen ein:

$$
\hat{\sigma}^2 = \frac{1}{n-m}\left|Y - X\hat{\beta}\right|^2. \tag{3.2.3}
$$

Dies ist eine verallgemeinerte Version des Varianzschätzers aus der einfachen linearen Regression, die wir bereits in Statistik I kennenlernten. Dabei ist $\hat{Y} = Y - X\hat{\beta}$ der Vektor der Residuen.

Satz 3.2.4 (*Erwartungstreue*)**.** Der Varianzschätzer

$$
\hat{\sigma}^2 = \frac{1}{n-m}\left|Y - X\hat{\beta}\right|^2
$$

ist erwartungstreu. Das heißt,

$$
\mathbb{E}\,\hat{\sigma}^2 = \sigma^2.
$$

Beweis.

$$\begin{aligned}
\hat{\sigma}^2 &= \frac{1}{n-m}\left(Y - X\hat{\beta}\right)^\top \left(Y - X\hat{\beta}\right)\\
&= \left(Y - X(X^\top X)^{-1}X^T Y\right)^\top \left(Y - X\left(X^\top X\right)^{-1} X^\top Y\right)\\
&= (DY)^\top DY
\end{aligned}$$

wobei $D = \mathcal{I} - X(X^\top X)^{-1}X^\top$ eine $(n \times n)$-Matrix ist. Dann ist

$$\hat{\sigma}^2 = \frac{1}{n-m} Y^\top D^\top D Y = \frac{1}{n-m} Y^\top D^2 Y = \frac{1}{n-m} Y^\top D Y, \text{ falls}$$

$D^\top = D$ und $D^2 = D$ (das heißt, daß D symmetrisch und idempotent ist). Tatsächlich gilt:

$$\begin{aligned}
D^\top &= \mathcal{I} - \left(X^\top\right)^\top \left(X^\top X\right)^{\top -1} X^\top = \mathcal{I} - X\left(X^\top X\right)^{-1} X^\top = D.\\
D^2 &= \left(\mathcal{I} - X(X^\top X)^{-1}X^T\right)\left(\mathcal{I} - X\left(X^\top X\right)^{-1} X^\top\right)\\
&= \mathcal{I} - 2X\left(X^\top X\right)^{-1} X^\top + X\left(X^\top X\right)^{-1} X^\top X \left(X^\top X\right)^{-1} X^\top\\
&= \mathcal{I} - X\left(X^\top X\right)^{-1} X^\top = D.
\end{aligned}$$

$\square$

Weiterhin gilt:

$$\begin{aligned}
\hat{\sigma}^2 &= \frac{1}{n-m} \cdot \operatorname{Spur}\left(Y^\top D Y\right) = \frac{1}{n-m} \cdot \operatorname{Spur}\left(D Y Y^\top\right)\\
\Longrightarrow \mathbb{E}\,\hat{\sigma}^2 &= \frac{1}{n-m} \cdot \operatorname{Spur}\left(D\mathbb{E}\left(YY^\top\right)\right) = \frac{\sigma^2}{n-m} \cdot \operatorname{Spur}(D),
\end{aligned}$$

denn

$$\operatorname{Spur}\left(D \cdot \mathbb{E}\left(YY^\top\right)\right) = \operatorname{Spur}\Big(D(X\beta)(X\beta)^\top + DX\beta \underbrace{\mathbb{E}\,\varepsilon^\top}_{=0} + D\underbrace{\mathbb{E}\,\varepsilon}_{=0}(X\beta)^\top + D \cdot \underbrace{\mathbb{E}\,\varepsilon\varepsilon^\top}_{= \operatorname{Cov}\varepsilon = \sigma^2 \cdot \mathcal{I}}\Big)$$

und

$$\begin{aligned}
DX &= \left(\mathcal{I} - X\left(X^\top X\right)^{-1} X^T\right) X\\
&= X - X\left(X^\top X\right)^{-1} X^\top X = X - X = 0.
\end{aligned}$$

Es bleibt zu zeigen, daß $\operatorname{Spur}(D) = n - m$:

$$\begin{aligned}
\operatorname{Spur}(D) &= \operatorname{Spur}\left(\mathcal{I} - X\left(X^\top X\right)^{-1} X^\top\right) = \operatorname{Spur}(\mathcal{I}) - \operatorname{Spur}\left(X\left(X^\top X\right)^{-1} X^\top\right)\\
&= n - \operatorname{Spur}\Big(\underbrace{X^\top X \cdot \left(X^\top X\right)^{-1}}_{\text{eine } (m \times m)\text{-Matrix}}\Big) = n - m.
\end{aligned}$$

3.2.3 Maximum-Likelihood-Schätzer für β und σ^2

Um Maximum-Likelihood-Schätzer für β und σ^2, bzw. Verteilungseigenschaften der MKQ-Schätzer von $\hat{\beta}$ und $\hat{\sigma}^2$ herleiten zu können, muß die Verteilung von ε bzw. Y präzisiert werden. Wir werden ab sofort normalverteilte Störgrößen betrachten, die unabhängig und identisch verteilt sind:

$$\varepsilon \sim N\left(0, \sigma^2 \mathcal{I}\right), \quad \sigma^2 > 0.$$

Daraus folgt:

$$Y \sim N\left(X\beta, \sigma^2 \mathcal{I}\right).$$

Wie sieht die Verteilung der MKQ-Schätzer $\hat{\beta}$ und $\hat{\sigma}^2$ aus? Da $\hat{\beta} = \left(X^\top X\right)^{-1} X^\top Y$ linear von Y abhängt, erwartungstreu ist und die $\operatorname{Cov}\hat{\beta} = \hat{\sigma}^2 \left(X^\top X\right)^{-1}$ besitzt, gilt:

$$\hat{\beta} \sim N\left(\beta, \sigma^2 \left(X^\top X\right)^{-1}\right)$$

Berechnen wir nun Maximum-Likelihood-Schätzer für β und σ^2, und zwar $\tilde{\beta}$ und $\tilde{\sigma}^2$. Dann zeigen wir, daß sie im Wesentlichen mit den MKQ-Schätzern übereinstimmen.

$$\begin{aligned} \tilde{\beta} &= \hat{\beta}, \\ \tilde{\sigma}^2 &= \frac{n-m}{m}\hat{\sigma}^2. \end{aligned}$$

Betrachten wir zunächst die Likelihood-Funktion von Y:

$$L(y, \beta, \sigma^2) = f_Y(y) = \frac{1}{\left(\sqrt{2\pi}\sigma\right)^n} \cdot \exp\left\{-\frac{1}{2\sigma^2}(y - X\beta)^\top (y - X\beta)\right\}$$

und die Log-Likelihood-Funktion

$$\operatorname{logL}(y, \beta, \sigma^2) = -\frac{n}{2}\log(2\pi) \underbrace{-\frac{n}{2}\log\left(\sigma^2\right) - \frac{1}{2\sigma^2}|y - X\beta|^2}_{:=g}.$$

Die Maximum-Likelihood-Schätzer sind dann

$$\left(\tilde{\beta}, \tilde{\sigma}^2\right) = \underset{\beta \in \mathbb{R}^m,\, \sigma^2 > 0}{\operatorname{argmax}} \operatorname{logL}(y, \beta, \sigma^2),$$

sofern sie existieren.

Satz 3.2.5 (*Maximum-Likelihood-Schätzung von $\tilde{\beta}$ und $\tilde{\sigma}^2$*)**.** Es existieren eindeutig bestimmte Maximum-Likelihood-Schätzer für β und σ^2, die folgendermaßen aussehen:

$$\begin{aligned} \tilde{\beta} &= \hat{\beta} = \left(X^\top X\right)^{-1} X^\top Y \\ \tilde{\sigma}^2 &= \frac{n-m}{n}\hat{\sigma}^2 = \frac{1}{n}\left|Y - X\tilde{\beta}\right|^2. \end{aligned}$$

Beweis. Wir fixieren $\sigma^2 > 0$ und suchen

$$\tilde{\beta} = \underset{\beta \in \mathbb{R}^m}{\operatorname{argmax}} \operatorname{logL}(Y, \beta, \sigma^2) = \underset{\beta \in \mathbb{R}^m}{\operatorname{argmin}} |Y - X\beta|^2,$$

woraus folgt, daß $\tilde{\beta}$ mit dem bekannten MKQ-Schätzer $\hat{\beta} = \left(X^\top X\right)^{-1} X^\top Y$ identisch ist, der nicht von σ^2 abhängt. Berechnen wir jetzt

$$\tilde{\sigma}^2 = \underset{\sigma^2 > 0}{\operatorname{argmax}} \operatorname{logL}\left(Y, \tilde{\beta}, \sigma^2\right) = \underset{\sigma^2 > 0}{\operatorname{argmax}}\, g(\sigma^2).$$

Es gilt

$$g\left(\sigma^2\right) \underset{\sigma^2 \to +\infty}{\longrightarrow} -\infty, \quad g\left(\sigma^2\right) \underset{\sigma^2 \to 0}{\longrightarrow} -\infty,$$

weil $|Y - X\beta|^2 \neq 0$, dadurch, daß $Y \sim N\left(X\beta, \sigma^2 \mathcal{I}\right) \in \{Xy : y \in \mathbb{R}^m\}$ mit Wahrscheinlichkeit Null. Da

$$g'(\sigma^2) = -\frac{n}{2}\frac{1}{\sigma^2} + \frac{|Y - X\beta|}{2\left(\sigma^2\right)^2} = 0, \text{ ist } \tilde{\sigma}^2 = \frac{1}{n}\left|Y - X\tilde{\beta}\right|^2$$

ein Maximumpunkt von $g(\sigma^2)$, das heißt, $\tilde{\sigma}^2$ ist ein Maximum-Likelihood-Schätzer für σ^2. □

Satz 3.2.6. Unter den obigen Voraussetzungen gilt:

1. $\mathbb{E}\,\tilde{\sigma}^2 = \frac{n-m}{m}\sigma^2$, das heißt, $\tilde{\sigma}^2$ ist nicht erwartungstreu; allerdings ist er asymptotisch unverzerrt.
2. $\frac{n}{\sigma^2}\tilde{\sigma}^2 \sim \chi^2_{n-m}$, $\frac{n-m}{\sigma^2}\hat{\sigma}^2 \sim \chi^2_{n-m}$.

Beweis. 1. Trivial (vergleiche den Beweis von Satz 3.2.4)

2. Wir zeigen den Satz nur für $\hat{\sigma}^2$.

$$\begin{aligned}
\frac{n-m}{\sigma^2}\hat{\sigma}^2 &= \frac{1}{\sigma^2}\left|Y - X\hat{\beta}\right|^2 \\
&= \frac{1}{\sigma^2} Y^\top \underbrace{D}_{=D^2} Y \quad \text{(nach dem Beweis von Satz 3.2.4)} \\
&= \frac{1}{\sigma^2}(DY)^\top DY = \frac{1}{\sigma^2}\left(D\left(X\beta + \varepsilon\right)\right)^\top \cdot D\left(X\beta + \varepsilon\right) \\
&= \frac{1}{\sigma^2}(D\varepsilon)^\top D\varepsilon = \left(\frac{\varepsilon}{\sigma}^\top\right) D \left(\frac{\varepsilon}{\sigma}\right),
\end{aligned}$$

wobei

$$\left(\frac{\varepsilon}{\sigma}\right) \sim N\left(0, \mathcal{I}\right).$$

Nach Satz 3.1.8 gilt

$$\frac{\varepsilon^\top}{\sigma} D \frac{\varepsilon}{\sigma} \sim \chi^2_r,$$

wobei $r = \text{Rang}(D)$, weil $D\mathcal{I} = D$ idempotent ist. Falls $r = n-m$, dann ist $\frac{n-m}{\sigma^2} \sim \chi^2_{n-m}$. Zeigen wir, daß $\text{Rang}(D) = r = n-m$. Aus der linearen Algebra ist bekannt, daß $\text{Rang}(X) = n - \dim(\text{Kern}(X))$. Wir zeigen, daß $\text{Kern}(D) = \{Xx : x \in \mathbb{R}^n\}$ und damit $\dim(\text{Kern}(D)) = m$, weil $\text{Rang}(X) = m$. Es ist $\{Xx : x \in \mathbb{R}^n\} \subseteq \text{Kern}(D)$, da

$$DX = (\mathcal{I} - X(X^\top X)^{-1}X^\top)X = X - (X^\top X)^{-1}X^\top X = 0.$$

und $\text{Kern}(D) \subseteq \{Xx : x \in \mathbb{R}^n\}$, weil

$$\begin{aligned} \forall y \in \text{Kern}(D) : \quad & Dy = 0 \Longleftrightarrow (\mathcal{I} - X(X^\top X)^{-1}X^\top)y = 0 \\ \Longleftrightarrow y = X \cdot & \underbrace{(X^\top X)^{-1}X^\top Y}_{x} = Xx \in \{Xx : x \in \mathbb{R}^n\}. \end{aligned}$$

□

Satz 3.2.7. Sei $Y = X\beta + \varepsilon$ ein multivariates lineares Regressionsmodell mit $Y = (Y_1, \ldots, Y_n)^\top$, Designmatrix X mit $\text{Rang}(X) = m$, $\beta = (\beta_1, \ldots, \beta_m)^\top$, $\varepsilon \sim N(0, \sigma^2\mathcal{I})$. Dann sind die Schätzer $\hat\beta = (X^\top X)^{-1}X^\top Y$ für β bzw. $\hat\sigma^2 = \frac{1}{n-m}|Y - X\hat\beta|^2$ für σ^2 unabhängig voneinander.

Beweis. Dieser Satz ist eine Verallgemeinerung des Lemmas 3.3.2 über die Unabhängigkeit der Schätzer $\overline{X}_n$ und S_n^2 der einfachen linearen Regression aus Statistik I. In diesem Beweis verwenden wir den Satz 3.1.9, für dessen Anwendung wir $\hat\beta$ als lineare und $\hat\sigma^2$ als quadratische Form von ε darstellen. Es ist in den Beweisen der Sätze 3.2.2 und 3.2.6 gezeigt worden, daß

$$\begin{aligned} \hat\beta &= \beta + \underbrace{(X^\top X)^{-1}X^\top}_{=A}\varepsilon, \\ \hat\sigma^2 &= \frac{1}{n-m}\varepsilon^\top D\varepsilon, \text{ wobei } D = \mathcal{I} - (X^\top X)^{-1}X^\top. \end{aligned}$$

Zusätzlich gilt $AD = 0$, weil nach dem Beweis des Satzes 3.2.4

$$(AD)^\top = D^\top A^\top = \underbrace{D \cdot X}_{=0}((X^\top X)^{-1})^\top = 0.$$

Da $\varepsilon \sim N(0, \sigma^2\mathcal{I})$, folgt daraus

$$A\sigma^2\mathcal{I}D = 0.$$

Deshalb sind die Voraussetzungen des Satzes 3.1.9 erfüllt, und $\hat\beta$ und $\hat\sigma^2$ sind unabhängig.

□

3.2.4 Tests für Regressionsparameter

In diesem Abschnitt wird zunächst die Hypothese

$$H_0 : \beta = \beta_0 \text{ vs. } H_1 : \beta \neq \beta_0$$

für ein $\beta_0 \in \mathbb{R}^m$ getestet. Dafür definieren wir die Testgröße

$$T = \frac{\left(\hat{\beta} - \beta_0\right)^\top X^\top X \left(\hat{\beta} - \beta_0\right)}{m\hat{\sigma}^2}.$$

Man kann zeigen (vergleiche Satz 3.2.8), daß unter H_0 gilt:

$$T \sim F_{m,n-m}.$$

Daraus folgt, daß H_0 abgelehnt werden soll, falls $T > F_{m,n-m,1-\alpha}$, wobei $F_{m,n-m,1-\alpha}$ das $(1-\alpha)$-Quantil der $F_{m,n-m}$-Verteilung darstellt. Dies ist ein Test zum Niveau $\alpha \in (0,1)$.

Spezialfall: Der Fall $\beta_0 = 0$ beschreibt einen *Test auf Zusammenhang* ; das heißt, man testet, ob die Parameter $\beta_1, \ldots, \beta_m$ für die Beschreibung der Daten Y relevant sind.

Bemerkung 3.2.1. 1. Wie kann man verstehen, daß die Testgröße T tatsächlich H_0 von H_1 unterscheiden soll? Führen wir die Bezeichnung

$$\tilde{Y} = Y - \underbrace{X\hat{\beta}}_{:=\hat{Y}}$$

ein; dabei gilt:

$$\hat{\sigma^2} = \frac{1}{n-m} \left|\tilde{Y}\right|^2$$

und $\tilde{Y}$ ist der Vektor der *Residuen*.

Ohne Beschränkung der Allgemeinheit setzen wir $\beta_0 = 0$. Falls H_0 nicht gelten soll, dann ist $\beta \neq 0$, und somit

$$|X\beta|^2 = (X\beta)^\top X\beta = \beta^\top X^\top X\beta > 0,$$

weil X den vollen Rang hat. Daraus folgt, daß H_0 abgelehnt werden soll, falls

$$\left|\hat{Y}\right|^2 = \left|X\hat{\beta}\right|^2 = \hat{\beta}^\top X^\top X\hat{\beta} \gg 0.$$

In der Testgröße $|X\hat{\beta}|^2$ sind allerdings die Schwankungen der Schätzung von β nicht berücksichtigt. Deswegen teilt man $|X\hat{\beta}|^2$ durch $\hat{\sigma}^2$:

$$T = \frac{\hat{\beta}^\top X^\top X\hat{\beta}}{m \cdot \hat{\sigma}^2} = \frac{\left|\hat{Y}\right|^2}{\frac{m}{n-m}\left|Y - \hat{Y}\right|^2},$$

Der Satz von Pythagoras liefert

$$|Y|^2 = \left|\tilde{Y}\right|^2 + \left|\hat{Y}\right|^2,$$

wobei unter H_0

$$\begin{aligned}
&\mathbb{E}\,|\hat{Y}^2| = \mathbb{E}\,|Y^2| - \mathbb{E}\,|Y-\hat{Y}|^2 = n\sigma^2 - \mathbb{E}\,|\tilde{Y}|^2 \quad \text{gilt, und somit} \\
&\frac{\mathbb{E}\,|\hat{Y}|^2}{\mathbb{E}\left(\frac{m}{n-m}\left|\tilde{Y}\right|^2\right)} \overset{(H_0)}{=} \frac{n\sigma^2 - \mathbb{E}\,|\tilde{Y}|^2}{\frac{m}{n-m}\mathbb{E}\,|\tilde{Y}|^2} = \frac{n-m}{m}\left(\frac{n\sigma^2}{\mathbb{E}\,|\tilde{Y}|^2} - 1\right), \\
&\text{weil } \mathbb{E}\,|Y|^2 = \mathbb{E}\left(Y^\top Y\right) = \sigma^2 \cdot n, \quad \text{wegen } Y \sim N(0, \sigma^2 \mathcal{I}).
\end{aligned}$$

$\Longrightarrow$ Die Testgröße T ist sensibel gegenüber Abweichungen von H_0.

2. Die Größe

$$\left|\tilde{Y}\right|^2 = \left|Y - \hat{Y}\right|^2$$

wird *Reststreuung* genannt. Mit deren Hilfe kann der Begriff des *Bestimmtheitsmaßes* R^2 aus der Statistik I wie folgt verallgemeinert werden:

$$R^2 = 1 - \frac{|\tilde{Y}|^2}{\left|Y - \overline{Y}_n \cdot e\right|^2},$$

wobei $e = (1, \ldots, 1)^\top$, $\overline{Y}_n = \frac{1}{n}\sum\limits_{i=1}^{n} Y_i$.

Satz 3.2.8. Unter $H_0 : \beta = \beta_0$ gilt

$$T = \frac{\left(\hat{\beta} - \beta_0\right)^\top X^\top X \left(\hat{\beta} - \beta_0\right)}{m\hat{\sigma}^2} \sim F_{m,n-m}.$$

Beweis. Es gilt

$$\begin{aligned}
&\hat{\beta} \sim N\left(\beta_0, \sigma^2 \left(X^\top X\right)^{-1}\right) \\
&\Longrightarrow \hat{\beta} - \beta_0 \sim N\Big(0, \underbrace{\sigma^2 (X^\top X)^{-1}}_{:=K}\Big).
\end{aligned}$$

Falls $A = \frac{X^\top X}{\sigma^2}$, dann ist $AK = \mathcal{I}$ idempotent. Dann gilt nach Satz 3.1.8

$$\left(\hat{\beta} - \beta_0\right)^\top A \left(\hat{\beta} - \beta_0\right) \overset{H_0}{\sim} \chi^2_m$$

(Zur Information: Unter H_1 wäre $(\hat{\beta} - \beta_0)^\top A(\hat{\beta} - \beta_0)$ nicht-zentral χ^2-verteilt).

Es gilt zusätzlich:

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

Aus Satz 3.2.7 folgt die Unabhängigkeit von $(\hat{\beta}-\beta_0)^\top A(\hat{\beta}-\beta_0)$ und $\frac{n-m}{\sigma^2}\hat{\sigma}^2$.

$$\Longrightarrow T = \frac{(\hat{\beta}-\beta_0)^\top (X^\top X)(\hat{\beta}-\beta_0)/m}{(n-m)\hat{\sigma}^2/(n-m)} \sim F_{m,n-m}$$

nach der Definition der F-Verteilung.

□

Jetzt wird die Relevanz der einzelnen Parameter β_j getestet:

$$H_0 : \beta_j = \beta_{0j} \text{ vs. } H_1 : \beta_j \neq \beta_{0j}.$$

Satz 3.2.9. Unter $H_0 : \beta_j = \beta_{0j}$ gilt:

$$T_j = \frac{\hat{\beta}_j - \beta_{0j}}{\hat{\sigma}\sqrt{x^{jj}}} \sim t_{n-m}, \text{ wobei}$$
$$\left(X^\top X\right)^{-1} = \left(x^{ij}\right)_{i,j=1,\dots,m}.$$

Beweis. Aus $\hat{\beta} \overset{H_0}{\sim} N(\beta_0, \sigma^2(X^\top X)^{-1})$ folgt $\hat{\beta}_j \overset{H_0}{\sim} N(\beta_{0j}, \sigma^2 x^{jj})$ und somit $\hat{\beta}_j - \beta_{0j} \sim N(0, \sigma^2 x^{jj})$. Dann ist $\frac{\hat{\beta}_j - \beta_{0j}}{\sigma\sqrt{x^{jj}}} \sim N(0,1)$. Zusätzlich gilt: $\frac{(n-m)\hat{\sigma}^2}{\sigma^2} \overset{H_0}{\sim} \chi^2_{n-m}$, und nach Satz 3.2.7 sind beide Größen unabhängig. Daraus folgt:

$$T_j = \frac{\frac{\hat{\beta}_j - \beta_{0j}}{\sigma\sqrt{x^{jj}}}}{\sqrt{\frac{(n-m)\hat{\sigma}^2}{(n-m)\sigma^2}}} \sim t_{n-m}.$$

□

Somit wird $H_0 : \beta_j = \beta_{j0}$ abgelehnt, falls $|T| > t_{n-m,1-\alpha/2}$. Dies ist ein Test von H_0 vs. H_1 zum Niveau α.

Sei nun

$$H_0 : \beta_{j_1} = \beta_{0j_1}, \dots, \beta_{j_l} = \beta_{0j_l} \text{ vs. } H_1 : \exists i \in \{1,\dots,l\} : \beta_{j_i} \neq \beta_{0j_i}$$

die zu testende Hypothese.

Übungsaufgabe 3.2.2. Zeigen Sie, daß unter H_0 folgende Verteilungsaussage gilt:

$$T = \frac{(\hat{\beta}' - \beta'_0)^\top K'(\hat{\beta}' - \beta'_0)}{(m-l+1)\hat{\sigma}^2} \sim F_{m-l+1,n-m},$$

wobei

$$\hat{\beta}' = (\hat{\beta}_{j_1}, \ldots, \hat{\beta}_{j_l}),$$
$$\beta_0' = (\beta_{0j_1}, \ldots, \beta_{0j_l}),$$
$$K' = \begin{pmatrix} x^{j_1 j_1} & \ldots & x^{j_1 j_l} \\ \vdots & \vdots & \vdots \\ x^{j_l j_1} & \ldots & x^{j_l j_l} \end{pmatrix}^{-1}.$$

Konstruieren Sie den dazugehörigen F-Test!

Test auf Linearkombination von Parametern

Sei nun

$$H_0 : H\beta = c \text{ vs. } H_1 : H\beta \neq c,$$

wobei H eine $(r \times m)$-Matrix und $c \in \mathbb{R}^r$ sind.

Satz 3.2.10. Unter H_0 gilt

$$T = \frac{(H\hat{\beta} - c)^\top (H(X^\top X)^{-1} H^\top)^{-1} (H\hat{\beta} - c)}{r\hat{\sigma^2}} \sim F_{r,n-m}.$$

Deshalb wird $H_0 : H\beta = c$ abgelehnt, falls $T > F_{r,n-m,1-\alpha}$.

Übungsaufgabe 3.2.3. Beweisen Sie Satz 3.2.10!

3.2.5 Konfidenzbereiche

1. *Konfidenzintervall für β_j*

   Im Satz 3.2.9 haben wir gezeigt, daß

   $$\frac{\hat{\beta}_j - \beta_j}{\hat{\sigma} \cdot \sqrt{x^{jj}}} \sim t_{n-m},$$

   wobei $(X^\top X) = (x^{ij})_{i,j=1,\ldots,m}$. Daraus kann mit den üblichen Überlegungen folgendes Konfidenzintervall für β_j zum Niveau $1 - \alpha$ abgeleitet werden:

   $$\mathbb{P}\left(\hat{\beta}_j - t_{n-m,1-\alpha/2} \cdot \hat{\sigma}\sqrt{x^{jj}} \;\le\; \beta_j \le \hat{\beta}_j + t_{n-m,1-\alpha/2} \cdot \hat{\sigma}\sqrt{x^{jj}}\right) = 1 - \alpha.$$

2. *Simultaner Konfidenzbereich für $\beta = (\beta_1, \ldots, \beta_m)^\top$*

   Falls A_j wie unten definiert ist, dann erhält man mit Hilfe folgender *Bonferroni-Ungleichung*

   $$\mathbb{P}\left(\bigcup_{j=1}^{m} A_j\right) \geq \sum_{j=1}^{m} \mathbb{P}(A_j) - (m-1),$$

daß

$$\mathbb{P}\Big(\underbrace{\hat{\beta}_j - t_{n-m,1-\alpha/(2m)} \cdot \hat{\sigma}\sqrt{x^{jj}} \le \beta_j \le \hat{\beta}_j + t_{n-m,1-\alpha/(2m)} \cdot \hat{\sigma}\sqrt{x^{jj}}, \quad j = 1, \ldots, m}_{:=A_j} \Big)$$

$$\overset{\text{(Bonferroni)}}{\ge} \sum_{j=1}^{m} \mathbb{P}\left(A_j\right) - (m-1) = m \cdot \left(1 - \frac{\alpha}{m}\right) - m + 1.$$

Daraus folgt, daß

$$\left\{ \beta - (\beta_1, \ldots, \beta_m)^\top : \beta_j \in \left[\hat{\beta}_j - t_{n-m,1-\alpha/(2m)} \cdot \hat{\sigma}\sqrt{x^{jj}},\ \hat{\beta} + t_{n-m,1-\alpha/(2m)} \cdot \hat{\sigma}\sqrt{x^{jj}} \right] \right\}$$

ein simultaner Konfidenzbereich für β zum Niveau $1 - \alpha$ ist.

3. *Konfidenzellipsoid für β.*

   In Satz 3.2.8 haben wir bewiesen, daß

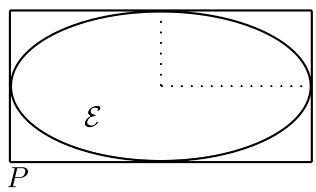
$$T = \frac{(\hat{\beta} - \beta)^\top (X^\top X)(\hat{\beta} - \beta)}{m\hat{\sigma}^2} \sim F_{m,n-m}.$$

   Daraus folgt, daß

$$\mathbb{P}\left(T \le F_{m,n-m,1-\alpha}\right) = 1 - \alpha \quad \text{und}$$
$$\mathcal{E} = \left\{ \beta \in \mathbb{R}^m : \frac{(\hat{\beta} - \beta)^\top (X^\top X)(\hat{\beta} - \beta)}{m\hat{\sigma}^2} \le F_{m,n-m,1-\alpha} \right\}$$

   ein Konfidenzellipsoid zum Niveau $1 - \alpha$ ist.

Abbildung 3.2: Konfidenzellipsoid



Da ein Ellipsoid in das minimale Parallelepipet P eingebettet werden kann, sodaß die Seitenlängen von P gleich $2\times$ der Halbachsenlängen von $\mathcal{E}$ sind, ergibt sich folgender simultaner Konfidenzbereich für β:

$$P = \left\{ \beta = (\beta_1, \ldots, \beta_m)^\top : \hat{\beta}_j - \hat{\sigma}\sqrt{m x^{jj} F_{m,n-m,1-\alpha}} \le \beta_j \le \hat{\beta}_j + \hat{\sigma}\sqrt{m x^{jj} F_{m,n-m,1-\alpha}} \right\}$$

für $j = 1, \ldots, m$.

4. *Konfidenzintervall für den erwarteten Zielwert* $x_{01}\beta_1 + \ldots + x_{0m}\beta_m$.

   Sei $Y_0 = x_{01}\beta_1 + \ldots + x_{0m}\beta_m + \varepsilon_0$ eine neue Zielvariable mit $\mathbb{E}\,\varepsilon_0 = 0$. Dann ist

$$\mathbb{E}\,Y_0 = \sum_{i=1}^{n} x_{0i}\beta_i.$$

   Wir konstruieren ein Konfidenzintervall für $\mathbb{E}\,Y_0$. Dazu verwenden wir die Beweisidee des Satzes 3.2.9 kombiniert mit Satz 3.2.10 mit $H = (x_{01}, \ldots, x_{0m}) = x_0^\top$, $r = 1$. Dann ist

$$T = \frac{\sum\limits_{i=1}^{m} \hat{\beta}_i x_{0i} - \sum\limits_{i=1}^{m} \beta_i x_{0i}}{\hat{\sigma}\sqrt{x_0^\top (X^\top X)^{-1} x_0}} \sim t_{n-m}.$$

   Darum ist

$$\Big\{\beta = (\beta_1, \ldots, \beta_m)^\top : \sum_{i=1}^{m} x_{0i}\hat{\beta}_i - \hat{\sigma}\sqrt{x_0^\top (X^\top X)^{-1} x_0} \cdot t_{n-m,1-\alpha/2} \\ \leq \sum_{i=1}^{m} x_{0i}\beta_i \leq \sum_{i=1}^{m} x_{0i}\hat{\beta}_i + \hat{\sigma}\sqrt{x_0^\top (X^\top X)^{-1} x_0} \cdot t_{n-m,1-\alpha/2}\Big\}$$

   ein Konfidenzintervall für $\sum\limits_{i=1}^{m} x_{0i}\beta_i$ zum Niveau $1 - \alpha$.

5. *Prognoseintervall für die Zielvariable* Y_0.

   Für $Y_0 = \sum\limits_{i=1}^{m} x_{0i}\beta_i + \varepsilon_0$ mit $\varepsilon_0 \sim N(0, \sigma^2)$, ε_0 unabhängig von $\varepsilon_1, \ldots, \varepsilon_n$, gilt:

$$\begin{aligned} & x_0^\top \hat{\beta} - Y_0 \sim N(0, \sigma^2(1 + x_0^\top (X^\top X)^{-1} x_0)) \\ \Longrightarrow\ & \frac{x_0^\top \hat{\beta} - Y_0}{\sigma\sqrt{1 + x_0^\top (X^\top X)^{-1} x_0}} \sim N(0,1) \\ \Longrightarrow\ & \frac{x_0^\top \hat{\beta} - Y_0}{\hat{\sigma}\sqrt{1 + x_0^\top (X^\top X)^{-1}} x_0} \sim t_{n-m} \end{aligned}$$

   Also ist

$$\Big(x_0^\top \hat{\beta} + c,\ x_0^\top \hat{\beta} - c\Big)$$

$$\text{mit } c = \hat{\sigma}\sqrt{1 + x_0^\top (X^\top X)^{-1} \cdot x_0} \cdot t_{n-m,1-\alpha/2}$$

   ein Prognoseintervall für die Zielvariable Y_0 zum Niveau $1 - \alpha$.

6. *Konfidenzband für die Regressionsebene* $y = \beta_1 + \sum\limits_{i=2}^{m} x_i\beta_i$ *im multiplen Regressionsmodell.*

   Es sei $Y = X\beta + \varepsilon$, wobei

$$X = \begin{pmatrix} 1 & x_{12} & \cdots & x_{1m} \\ 1 & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n2} & \cdots & x_{nm} \end{pmatrix} \quad \text{und } \varepsilon \sim N(0,\, \sigma^2 \cdot \mathcal{I}).$$

   Wir wollen ein zufälliges Konfidenzband $B(x)$ für y angeben. Es gilt

$$\mathbb{P}\left(y = \beta_1 + \sum_{i=2}^{m} \beta_i x_i \in B(x)\right) = 1 - \alpha \quad \forall x \in \mathbb{R}_1^{m-1}, \quad \text{wobei}$$
$$R_1^{m-1} = \left\{(1, x_2, \ldots, x_m)^\top \in \mathbb{R}^m\right\}.$$

   Satz 3.2.11. Es gilt:

$$\mathbb{P}\left(\max_{x \in \mathbb{R}_1^{m-1}} \frac{\left(x^T\hat{\beta} - \overbrace{\left(\beta_1 + \sum_{i=2}^{m} \beta_i x_i\right)}^{=y}\right)^2}{\hat{\sigma}^2 x^\top (X^\top X)^{-1} x} \leq m \cdot F_{m,n-m,1-\alpha}\right) = 1 - \alpha.$$

   ohne Beweis.

3.3 Multivariate lineare Regression mit Rang$(X) < m$

Es sei $Y = X\beta + \varepsilon$, $Y \in \mathbb{R}^n$, wobei X eine $(n \times m)$-Matrix mit $\text{Rang}(X) = r < m$ ist, $\beta = (\beta_1, \ldots, \beta_m)^\top$, $\varepsilon \in \mathbb{R}^n$, $\mathbb{E}\,\varepsilon = 0$, $\mathbb{E}\,(\varepsilon_i\varepsilon_j) = \delta_{ij}\sigma^2$, $\forall i, j = 1, \ldots, n$, $\sigma^2 > 0$.

Der MKQ-Schätzer $\hat{\beta}$ ist nach wie vor eine Lösung der Normalengleichung

$$\left(X^\top X\right)\beta = X^\top Y.$$

$X^\top X$ ist aber nicht mehr invertierbar, weil

$$\text{Rang}(X^\top X) \leq \min\left\{\text{Rang}(X), \text{Rang}(X^\top)\right\} = r < m.$$

Um $\hat{\beta}$ aus der Normalengleichung zu gewinnen, sollen beide Seiten der Gleichung mit der sogenannten *verallgemeinerten Inversen* von $X^\top X$ multipliziert werden.

3.3.1 Verallgemeinerte Inverse

Definition 3.3.1. Sei A eine $(n \times m)$-Matrix. Eine $(m \times n)$-Matrix A^- heißt *verallgemeinerte Inverse* von A, falls

$$AA^-A = A \quad \text{gilt.}$$

Die Matrix A^- ist nicht eindeutig bestimmt, was die folgenden Hilfssätze zeigen.

Lemma 3.3.1. Sei A eine $(n \times m)$-Matrix, $m \leq n$ mit $\mathrm{Rang}(A) = r \leq m$. Es existieren invertierbare Matrizen P $(n \times n)$ und Q $(m \times m)$, sodaß

$$PAQ = \mathrm{diag}(\underbrace{1, \ldots, 1}_{r \text{ Mal}}, 0, \ldots, 0) \tag{3.3.1}$$

Folgerung 3.3.1. Für eine beliebige $(n \times m)$-Matrix A mit $n \geq m$, $r = \mathrm{Rang}(A) \leq m$ gilt

$$A^- = Q \begin{pmatrix} \mathcal{I}_r & A_2 \\ A_1 & A_3 \end{pmatrix} P, \tag{3.3.2}$$

wobei P und Q Matrizen aus der Darstellung (3.3.1) sind, $\mathcal{I}_r = \mathrm{diag}(\overbrace{1, \ldots, 1}^{r \text{ Mal}})$, und A_1, A_2, A_3 beliebige $((m-r) \times r)$, $(r \times (n-r))$ bzw. $((m-r) \times (n-r))$-Matrizen sind.

Insbesondere kann

$$\begin{aligned} A_1 &= 0, \\ A_2 &= 0, \\ A_3 &= \mathrm{diag}(\underbrace{1, \ldots, 1}_{s-r \text{ Mal}}, 0, \ldots, 0), \\ s &\in \{r, \ldots, m\} \end{aligned}$$

gewählt werden, das heißt, $\mathrm{Rang}(A^-) = s \in \{r, \ldots, m\}$ für

$$A^- = Q \begin{pmatrix} \mathcal{I}_s & 0 \\ 0 & 0 \end{pmatrix} P.$$

Beweis. Zeigen wir, daß für A^- wie in (3.3.2) gegeben, $AA^-A = A$ gilt. Aus Lemma 3.3.1 folgt, daß

$$\begin{aligned} A &= P^{-1} \cdot \mathrm{diag}(1, \ldots, 1, 0, \ldots, 0) \cdot Q^{-1} \quad \text{und somit} \\ AA^-A &= P^{-1} \begin{pmatrix} \mathcal{I}_r & 0 \\ 0 & 0 \end{pmatrix} Q^{-1} Q \cdot \begin{pmatrix} \mathcal{I}_r & A_2 \\ A_1 & A_3 \end{pmatrix} PP^{-1} \begin{pmatrix} \mathcal{I}_r & 0 \\ 0 & 0 \end{pmatrix} Q^{-1} \\ &= P^{-1} \begin{pmatrix} \mathcal{I}_r & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \mathcal{I}_r & A_2 \\ A_1 & A_3 \end{pmatrix} \begin{pmatrix} \mathcal{I}_r & 0 \\ 0 & 0 \end{pmatrix} Q^{-1} = P^{-1} \begin{pmatrix} \mathcal{I}_r & 0 \\ 0 & 0 \end{pmatrix} Q^{-1} \\ &= A. \end{aligned}$$

$\square$

Lemma 3.3.2. Sei A eine beliebige $(n \times m)$-Matrix mit $\mathrm{Rang}(A) = r \le m$, $m \le n$.

1. Falls $(A^\top A)^-$ eine verallgemeinerte Inverse von $A^\top A$ ist, dann ist $\left((A^\top A)^-\right)^\top$ ebenfalls eine verallgemeinerte Inverse von $A^\top A$.
2. Es gilt die Darstellung
$$\begin{aligned}(A^\top A)(A^\top A)^- A^\top &= A^\top \quad \text{bzw.}\\ A(A^\top A)^-(A^\top A) &= A.\end{aligned}$$

Beweis. 1. $A^\top A$ ist symmetrisch, also
$$\underbrace{\left(A^\top A(A^\top A)^- A^\top A\right)^\top}_{=A^\top A\left((A^\top A)^-\right)^\top A^\top A} = \left(A^\top A\right)^\top = A^\top A.$$

Also ist $\left((A^\top A)^-\right)^\top$ eine verallgemeinerte Inverse von $A^\top A$.

2. Es sei $B = (A^\top A)(A^\top A)^- A^\top - A^\top$. Wir zeigen, daß $B = 0$, indem wir zeigen, daß $BB^\top = 0$.
$$\begin{aligned}BB^\top &= \left((A^\top A)(A^\top A)^- A^\top - A^\top\right)\left(A\left((A^\top A)^-\right)^\top A^\top A - A\right)\\ &= A^\top A(A^\top A)^- A^\top A\left((A^\top A)^-\right)^\top A^\top A - \underbrace{A^\top A(A^\top A)^- A^\top A}_{=A^\top A}\\ &\quad - \underbrace{A^\top A\left((A^\top A)^-\right)^\top}_{=A^\top A} \cdot A^\top A + A^\top A = A^\top A - 2A^\top A + A^\top A = 0.\end{aligned}$$

Die Aussage $A(A^\top A)^- A^\top A = A$ erhält man, indem man die Matrizen an beiden Seiten der Gleichung $A^\top A(A^\top A)^- A^\top = A^\top$ transponiert.

□

3.3.2 MKQ-Schätzer für β

Satz 3.3.1. Es sei X eine $(n \times m)$-Designmatrix mit $\mathrm{Rang}(X) = r \le m$ in der linearen Regression $Y = X\beta + \varepsilon$. Die allgemeine Lösung der Normalengleichung
$$\left(X^\top X\right)\beta = X^\top Y$$

sieht folgendermaßen aus:
$$\beta = \left(X^\top X\right)^- X^\top Y + \left(\mathcal{I}_m - \left(X^\top X\right)^- X^\top X\right) z, \quad z \in \mathbb{R}^m. \tag{3.3.3}$$

Beweis. 1. Zeigen wir, daß β wie in (3.3.3) angegeben, eine Lösung der Normalengleichung darstellt.

$$X^\top X\beta = \underbrace{(X^\top X)(X^\top X)^- X^\top}_{=X^\top (\text{ Lemma 3.3.2, 2.})} Y + \left(X^\top X - \underbrace{X^\top X(X^\top X)^- X^\top X}_{=X^\top X}\right) z$$
$$= X^\top Y$$

2. Zeigen wir, daß eine beliebige Lösung β' der Normalengleichung die Form (3.3.3) besitzt. Sei β die Lösung (3.3.3). Wir bilden die Differenz der Gleichungen

$$\begin{array}{rl} (X^\top X)\beta' & = X^\top Y \\ - \quad (X^\top X)\beta & = X^\top Y \\ \hline (X^\top X)(\beta' - \beta) & = 0 \end{array}$$

$$\begin{aligned} \beta' &= (\beta' - \beta) + \beta \\ &= \beta' - \beta + (X^\top X)^- X^\top Y + \left(\mathcal{I}_m - (X^\top X)^- X^\top X\right) z \\ &= (X^\top X)^- X^\top Y + \left(\mathcal{I}_m - (X^\top X)^- X^\top X\right) z + (\beta' - \beta) - \underbrace{(X^\top X)^- X^\top X(\beta' - \beta)}_{=0} \\ &= (X^\top X)^- X^\top Y + \left(\mathcal{I}_m - (X^\top X)^- X^\top X\right) \Big(\underbrace{z + \beta' - \beta}_{=z_0}\Big) \end{aligned}$$

$\Longrightarrow \beta'$ besitzt die Darstellung (3.3.3).

□

Bemerkung 3.3.1. Der Satz 3.3.1 liefert die Menge aller Extremalpunkte der MKQ-Minimierungsaufgabe

$$e(\beta) = \frac{1}{n} |Y - X\beta|^2 \longrightarrow \min_\beta .$$

Deshalb soll die Menge aller MKQ-Schätzer von β in (3.3.3) zusätzliche Anforderungen erfüllen.

Satz 3.3.2. 1. Alle MKQ-Schätzer von β haben die Form

$$\overline{\beta} = \left(X^\top X\right)^- X^\top Y, \quad \text{wobei}$$

$(X^\top X)^-$ eine beliebige verallgemeinerte Inverse von $X^\top X$ ist.

2. $\overline{\beta}$ ist nicht erwartungstreu, denn

$$\mathbb{E}\,\overline{\beta} = \left(X^\top X\right)^- X^\top X\beta.$$

3. Es gilt:

$$\mathrm{Cov}\overline{\beta} = \sigma^2 \left(X^\top X\right)^- \left(X^\top X\right) \left((X^\top X)^- \right)^\top .$$

Beweis. 1. Zeigen wir, daß $e(\beta) \geq e(\overline{\beta}) \quad \forall \beta \in \mathbb{R}^m$.

$$\begin{aligned}
n \cdot e(\beta) &= |Y - X\beta|^2 = (Y - X\overline{\beta} + X(\overline{\beta} - \beta))^\top (Y - X\overline{\beta} + X(\overline{\beta} - \beta)) \\
&= (Y - X\overline{\beta})^\top (Y - X\overline{\beta}) + \left(X(\overline{\beta} - \beta)\right)^\top \left(X(\overline{\beta} - \beta)\right) \\
&\quad + 2(\overline{\beta} - \beta)^\top X^\top (Y - X\overline{\beta}) \\
&= n \cdot e(\overline{\beta}) + \underbrace{2 \cdot (\overline{\beta} - \beta)^\top (X^\top Y - (X^\top X\overline{\beta}))}_{=0} + \left|X(\overline{\beta} - \beta)\right|^2 \\
&\geq n \cdot e(\overline{\beta}) + 0 = n \cdot e(\overline{\beta}), \qquad \text{denn}
\end{aligned}$$

$$X^\top Y - X^\top X\overline{\beta} = X^\top Y - \underbrace{X^\top X (X^\top X)^- X^\top}_{=X^\top (\text{Lemma } 3.3.2)} Y = X^\top Y - X^\top Y = 0.$$

2. Es gilt:

$$\begin{aligned}
\mathbb{E}\,\overline{\beta} &= \mathbb{E}\left((X^\top X)^- X^\top Y\right) = \left(X^\top X\right)^- X^\top \mathbb{E}\,Y \\
&= (X^\top X)^- X^\top X\beta, \qquad \text{weil aus} \\
Y &= X\beta + \varepsilon, \quad \mathbb{E}\,\varepsilon = 0 \qquad \text{die Relation } \mathbb{E}\,Y = X\beta \text{ folgt.}
\end{aligned}$$

Warum ist $\overline{\beta}$ nicht erwartungstreu? Also warum ist $(X^\top X)^- X^\top X\beta \neq \beta$, $\beta \in \mathbb{R}^m$? Da $\mathrm{Rang}(X) = r < m$, ist $\mathrm{Rang}(X^\top X) < m$ und damit $\mathrm{Rang}((X^\top X)^- X^\top X) < m$. Darum existiert ein $\beta \neq 0$, für das gilt:

$$\left(X^\top X\right)^- X^\top X\beta = 0 \neq \beta,$$

also ist $\overline{\beta}$ nicht erwartungstreu. Es gilt sogar, daß alle Lösungen von (3.3.3) keine erwartungstreuen Schätzer sind. Wenn wir den Erwartungswert an (3.3.3) anwenden, so erhielten wir im Falle der Erwartungstreue:

$$\begin{aligned}
&\forall \beta \in \mathbb{R}^m : \quad \beta = (X^\top X)^- X^\top X\beta + \left(\mathcal{I}_m - (X^\top X)^-(X^\top X)\right) z, \quad z \in \mathbb{R}^m. \\
&\Longrightarrow \left(\mathcal{I}_m - (X^\top X)^-(X^\top X)\right)(z - \beta) = 0 \quad \forall z, \beta \in \mathbb{R}^m \\
&\Longrightarrow (X^\top X)^-(X^\top X)(\beta - z) = \beta - z, \quad \forall z, \beta \in \mathbb{R}^m.
\end{aligned}$$

Da diese Gleichung nicht für alle $\beta \in \mathbb{R}^m$ gelten kann (siehe oben), führt die Annahme der Erwartungstreue zum Widerspruch.

3. Es gilt:

$$\begin{aligned}
\operatorname{Cov}\left(\overline{\beta}_i, \overline{\beta}_j\right) &= \operatorname{Cov}\Big(\Big(\underbrace{(X^\top X)^- X^\top}_{:=A=(a_{kl})} Y\Big)_i, \Big((X^\top X)^- X^\top Y\Big)_j\Big) \\
&= \operatorname{Cov}\left(\sum_{k=1}^n a_{ik} Y_k, \sum_{l=1}^n a_{jl} Y_l\right) \\
&= \sum_{k,l=1}^n a_{ik} a_{jl} \underbrace{\operatorname{Cov}\Big(Y_k, Y_l\Big)}_{=\sigma^2 \cdot \delta_{kl}} = \sigma^2 \sum_{k=1}^n a_{ik} a_{jk} = \Big(\sigma^2 A A^\top\Big)_{i,j} \\
&= \left(\sigma^2 (X^\top X)^- X^\top X \left((X^\top X)^-\right)^\top\right)_{i,j=1,\dots,m}
\end{aligned}$$

□

3.3.3 Erwartungstreu schätzbare Funktionen

Definition 3.3.2. Eine Linearkombination $a^\top \beta$ von $\beta_1, \dots, \beta_m$, $a \in \mathbb{R}^m$ heißt *(erwartungstreu) schätzbar* , falls

$$\exists\, c \in \mathbb{R}^n : \quad \mathbb{E}\left(c^\top Y\right) = a^\top \beta,$$

das heißt, falls es einen linearen, erwartungstreuen Schätzer $c^\top Y$ für $a^\top \beta$ gibt.

Satz 3.3.3. Die Funktion $a^\top \beta$, $a \in \mathbb{R}^m$ ist genau dann erwartungstreu schätzbar, wenn eine der folgenden Bedingungen erfüllt ist:

1. $\exists\, c \in \mathbb{R}^n : \quad a^\top = c^\top X.$

2. a erfüllt die Gleichung

$$a^\top \left(X^\top X\right)^- X^\top X = a^\top. \tag{3.3.4}$$

Beweis. 1. „$\Longrightarrow$“: Falls $a^\top \beta$ schätzbar, dann existiert ein $d \in \mathbb{R}^n$ mit $\mathbb{E}\,(d^\top Y) = a^\top \beta \quad \forall \beta \in \mathbb{R}^m$. Also

$$\begin{aligned}
&a^\top \beta = d^\top \mathbb{E}\, Y = d^\top X \beta \Rightarrow \left(a^\top - d^\top X\right) \beta = 0, \quad \forall \beta \in \mathbb{R}^m \\
&\Longrightarrow a^\top = d^\top X,
\end{aligned}$$

setze $c = d$, damit ist die erste Richtung bewiesen.

„$\Longleftarrow$“: $\mathbb{E}\,(c^\top Y) = c^\top \mathbb{E}\, Y = c^\top X \beta \;= a^\top \beta$, also ist $a^\top \beta$ erwartungstreu schätzbar.

2. „$\Longrightarrow$“: Falls $a^\top\beta$ erwartungstreu schätzbar ist, dann gilt:

$$a^\top(X^\top X)^- X^\top X \overset{\text{Punkt 1}}{=} c^\top \underbrace{X\cdot(X^\top X)^- X^\top X}_{=X \text{ (Lemma 3.3.2)}} = c^\top X \overset{\text{(Punkt 1)}}{=} a^\top.$$

Also ist (3.3.4) erfüllt.

„$\Longleftarrow$“: Falls $a^\top(X^\top X)^- X^\top X = a^\top$, dann gilt mit $c = (a^\top(X^\top X)^- X^\top)^\top$ nach Punkt 1, daß $a^\top\beta$ schätzbar ist.

$\square$

Bemerkung 3.3.2. Im Falle der Regression mit $\mathrm{Rang}(X) = m$ ist die Gleichung (3.3.4) immer erfüllt, denn $(X^\top X)^- = (X^\top X)^{-1}$ und damit ist $a^\top\beta$ schätzbar für alle $a \in \mathbb{R}^m$.

Satz 3.3.4 (*Beispiele schätzbarer Funktionen*)**.** Falls $\mathrm{Rang}(X) = r < m$, dann sind folgende Linearkombinationen von β schätzbar:

1. Die Koordinaten $\sum\limits_{j=1}^m x_{ij}\beta_j$, $i = 1,\dots,n$ des Erwartungswertvektors $\mathbb{E}Y = X\beta$.
2. Beliebige Linearkombinationen schätzbarer Funktionen.

Beweis. 1. Führe die Bezeichnung $\tilde{x}_i = (x_{i1},\dots,x_{im})$, $i = 1,\dots,n$ ein. Dann ist

$$\sum_{j=1}^m x_{ij\beta_j} = \tilde{x}_i^\top\beta \quad \forall i = 1,\dots,n,$$

$$X\beta = (\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_n)^\top \beta,$$

$\tilde{x}_i\beta$ ist schätzbar, falls $\tilde{x}_i$ die Gleichung (3.3.4) erfüllt, die für alle $i = 1,\dots,n$ folgendermaßen in Matrixform dargestellt werden kann:

$$X\left(X^\top X\right)^- X^\top X = X,$$

was nach Lemma 3.3.2 Gültigkeit besitzt.

2. Für alle $a_1,\dots,a_k \in \mathbb{R}^m$ seien $a_1^\top\beta,\dots,a_k^\top\beta$ schätzbare Funktionen. Für alle $\lambda = (\lambda_1,\dots,\lambda_k)^\top \in \mathbb{R}^k$ zeigen wir, daß $\sum\limits_{i=1}^k \lambda_i \cdot a_i^\top\beta = \lambda^\top A\beta$ schätzbar ist, wobei $A = (a_1,\dots,a_k)^\top$. Zu zeigen bleibt: $b = (\lambda^\top A)^\top$ erfüllt (3.3.4), also

$$\lambda^\top A\left(X^\top X\right)^- X^\top X = \lambda^\top A.$$

Diese Gleichung stimmt, weil $a_i^\top(X^\top X)^- X^\top X = a_i^\top$, $i = 1,\dots,k$. Nach Satz 3.3.3, 2.) ist $\lambda^\top A\beta$ schätzbar.

$\square$

Satz 3.3.5 (*Gauß-Markov*)**.** Es sei $a^\top\beta$ eine schätzbare Funktion, $a \in \mathbb{R}^m$ im linearen Regressionsmodell $Y = X\beta + \varepsilon$ mit $\mathrm{Rang}(X) \le m$.

1. Der beste lineare erwartungstreue Schätzer (engl. BLUE - best linear unbiased estimator) von $a^\top \beta$ ist durch $a^\top \overline{\beta}$ gegeben, wobei
$$\overline{\beta} = \left(X^\top X\right)^- X^\top Y$$
ein MKQ-Schätzer für β ist.
2. $\operatorname{Var}(a^\top \overline{\beta}) = \sigma^2 a^\top (X^\top X)^- a$

Beweis. Die Linearität von $a^\top \overline{\beta} = a^\top (X^\top X)^- X^\top Y$ als Funktion von Y ist klar. Zeigen wir die Erwartungstreue:
$$\begin{aligned}\mathbb{E}\,(a^\top \overline{\beta}) &= a^\top \mathbb{E}\,\overline{\beta} = a^\top (X^\top X)^- X^\top X \beta \\ &= c^\top \underbrace{X(X^\top X)^- X^\top X}_{=X \text{ (Lemma 3.3.2)}} \beta = \underbrace{c^\top X}_{=a^\top} \beta = a^\top \beta \quad \forall \beta \in \mathbb{R}^m.\end{aligned}$$

Berechnen wir $\operatorname{Var}(a^\top \overline{\beta})$ (also beweisen wir Punkt 2), und zeigen, daß sie minimal ist.
$$\begin{aligned}\operatorname{Var}(a^\top \overline{\beta}) &= \operatorname{Var}\left(\sum_{i=1}^m a_i \overline{\beta}_i\right) = \sum_{i,j=1}^m a_i a_j \cdot \operatorname{Cov}\left(\overline{\beta}_i, \overline{\beta}_j\right) \\ &= a^\top \operatorname{Cov}\left(\overline{\beta}\right) a \overset{\text{(Satz 3.3.2)}}{=} a^\top \sigma^2 \left((X^\top X)^- X^\top X (X^\top X)^-\right)^\top a \\ &= \sigma^2 \cdot a^\top \underbrace{\left((X^\top X)^-\right)^\top}_{=(X^\top X)^-} X^\top X \underbrace{\left((X^\top X)^-\right)^\top}_{(X^\top X)^-} a \\ &\overset{\text{Lemma 3.3.2, 1.)}}{=} \sigma^2 a^\top (X^\top X)^- X^\top X (X^\top X)^- a \\ &\overset{\text{Satz 3.3.3, 1.)}}{=} \sigma^2 \cdot c^\top \underbrace{X \cdot (X^\top X) X^\top X}_{=X} (X^\top X)^- X^\top c \\ &= \sigma^2 \underbrace{c^\top X}_{=a^\top} (X^\top X)^- \underbrace{X^\top c}_{=a} = \sigma^2 a^\top (X^\top X)^- a.\end{aligned}$$

Jetzt zeigen wir, daß für einen beliebigen linearen, erwartungstreuen Schätzer $b^\top Y$ von $a^\top \beta$ gilt: $\operatorname{Var}(b^\top Y) \geq \operatorname{Var}(a^\top \overline{\beta})$. Weil $b^\top Y$ erwartungstreu ist, gilt: $\mathbb{E}\,(b^\top Y) = a^\top \beta$. Nach Satz 3.3.3 gilt: $a^\top = b^\top X$. Betrachten wir die Varianz von
$$\begin{aligned}0 &\leq \operatorname{Var}\left(b^\top Y - a^\top \overline{\beta}\right) = \operatorname{Var}\left(b^\top Y\right) - 2\operatorname{Cov}\left(b^\top Y, a^\top \overline{\beta}\right) + \operatorname{Var}\left(a^\top \overline{\beta}\right) \\ &= \operatorname{Var}(b^\top Y) - 2\sigma^2 a^\top (X^\top X)^- a + \sigma^2 a^\top (X^\top X)^- a = \operatorname{Var}(b^\top Y) - \operatorname{Var}\left(a^\top \overline{\beta}\right)\end{aligned}$$
mit
$$\begin{aligned}\operatorname{Cov}\left(b^\top Y, a^\top \overline{\beta}\right) &= \operatorname{Cov}\left(b^\top Y, a^\top (X^\top X)^- X^\top Y\right) = \sigma^2 a^\top (X^\top X)^- \underbrace{X^\top b}_{=a} \\ &= \sigma^2 a^\top (X^\top X)^- a.\end{aligned}$$

Damit ist $\operatorname{Var}\left(b^\top Y\right) \geq \operatorname{Var}\left(a^\top \overline{\beta}\right)$ und $a^\top \overline{\beta}$ ist ein bester, linearer, erwartungstreuer Schätzer für $a^\top \beta$. □

Bemerkung 3.3.3. 1. Falls $\text{Rang}(X) = m$, dann ist $a^\top\hat{\beta}$ der beste lineare, erwartungstreue Schätzer für $a^\top\beta$, $a \in \mathbb{R}^m$.

2. Wie im folgenden Satz gezeigt wird, hängt der Schätzer $a^\top\overline{\beta} = a^\top(X^\top X)^- X^\top Y$ nicht von der Wahl der verallgemeinerten Inversen ab.

Satz 3.3.6. Der beste lineare, erwartungstreue Schätzer $a^\top\overline{\beta}$ für $a^\top\beta$ ist eindeutig bestimmt.

Beweis.

$$a^\top\overline{\beta} = a^\top(X^\top X)^- X^\top Y \overset{\text{Satz 3.3.3, 1.)}}{=} c^\top X(X^\top X)^- X^\top Y.$$

Wir zeigen, daß $X(X^\top X)^- X^\top$ nicht von der Wahl von $(X^\top X)^-$ abhängt. Zeigen wir, daß für beliebige verallgemeinerte Inverse A_1 und A_2 von $(X^\top X)$ gilt: $XA_1X^\top = XA_2X^\top$. Nach Lemma 3.3.2, 2.) gilt:

$$XA_1X^\top X = X = XA_2X^\top X.$$

Multiplizieren wir alle Teile der Gleichung mit $A_1X^\top$ von rechts:

$$XA_1 \underbrace{X^\top XA_1X^\top}_{=X^\top} = XA_1X^\top = XA_2 \underbrace{X^\top XA_1X^\top}_{=X^\top}$$

Also ist $XA_1X^\top = XA_2X^\top$. □

3.3.4 Normalverteilte Störgrößen

Sei $Y = X\beta + \varepsilon$ ein lineares Regressionsmodell mit $\text{Rang}(X) = r < m$ und $\varepsilon \sim N(0, \sigma^2\mathcal{I})$. Genauso wie in Abschnitt 3.2.3 können Maximum-Likelihood-Schätzer $\tilde{\beta}$ und $\tilde{\sigma}^2$ für β und σ^2 hergeleitet werden. Und genauso wie im Satz 3.2.5 kann gezeigt werden, daß

$$\tilde{\beta} = \overline{\beta} = (X^\top X)^- X^\top Y \quad \text{und}$$
$$\tilde{\sigma}^2 = \frac{1}{n}\left|Y - X\overline{\beta}\right|^2.$$

Jetzt werden die Verteilungseigenschaften von $\overline{\beta}$ und $\tilde{\sigma}^2$ untersucht. Wir beginnen mit der Erwartungstreue von $\tilde{\sigma}^2$. Wir zeigen, daß $\tilde{\sigma}^2$ nicht erwartungstreu ist, dafür ist aber der korrigierte Schätzer

$$\overline{\sigma}^2 = \frac{1}{n-r}\left|Y - X\beta\right|^2 = \frac{n}{n-r}\tilde{\sigma}^2$$

erwartungstreu.

Satz 3.3.7. Der Schätzer $\overline{\sigma}^2$ ist erwartungstreu für σ^2.

Der Beweis des Satzes 3.3.7 folgt dem Beweis des Satzes 3.2.4, in dem der Schätzer $\hat{\beta} = (X^\top X)^{-1}X^\top Y$ und $\hat{\sigma}^2 = \frac{1}{n-m}|Y - X\beta|^2$ im Fall $\text{Rang}(X) = m$ betrachtet worden. Somit ist die Aussage des Satzes 3.2.4 ein Spezialfall des Satzes 3.3.7. Führen wir die Matrix $D = \mathcal{I} - X(X^\top X)^- X^\top$ ein.

Lemma 3.3.3. Für D gelten folgende Eigenschaften:

1. $D^\top = D$ (Symmetrie),
2. $D^2 = D$ (Idempotenz),
3. $DX = 0$,
4. $\mathrm{Spur}(D) = n - r$.

Beweis. 1. Es gilt:

$$\begin{aligned} D^\top &= \left(\mathcal{I} - X(X^\top X)^- X^\top\right)^\top = \mathcal{I} - X\left((X^\top X)^-\right)^\top X^\top \\ &= \mathcal{I} - X(X^\top X)^- X^\top = D, \end{aligned}$$

weil $\left((X^\top X)^-\right)^\top$ auch eine verallgemeinerte Inverse von $X^\top X$ ist (vergleiche Lemma 3.3.2, 1.)).

2. Es gilt:

$$\begin{aligned} D^2 &= \left(\mathcal{I} - X(X^\top X)^- X^\top\right)^2 = \mathcal{I} - 2X(X^\top X)^- X^\top + \underbrace{X(X^\top X)^- X^\top X}_{=X(\text{Lemma 3.3.2, 2.}))}(X^\top X)^- X^\top \\ &= \mathcal{I} - X(X^\top X)^- X^\top = D. \end{aligned}$$

3. $DX = X - \underbrace{X(X^\top X)^- X^\top X}_{=X\ (\text{Lemma 3.3.2, 2.}))} = X - X = 0.$

4. Es gilt:

$$\mathrm{Spur}(D) = \mathrm{Spur}(I) - \mathrm{Spur}\left(X(X^\top X)^- X^\top\right) = n - \mathrm{Spur}\left(X(X^\top X)^- X^\top\right).$$

Verwenden wir die Eigenschaft der symmetrischen idempotenten Matrizen A aus der linearen Algebra, daß $\mathrm{Spur}(A) = \mathrm{Rang}(A)$. Da $X(X^\top X)^- X^\top$ symmetrisch und idempotent ist, genügt es zu zeigen, daß $\mathrm{Rang}(X(X^\top X)^- X^\top) = r$. Nach Lemma 3.3.2 2.) gilt:

$$\begin{aligned} \mathrm{Rang}(X) = r &= \mathrm{Rang}(X(X^\top X)^- X^\top X) \\ &\le \min\left\{\mathrm{Rang}(X(X^\top X)^- X^\top), \underbrace{\mathrm{Rang}(X)}_{=r}\right\} \\ &\le \mathrm{Rang}\left(X(X^\top X)^- X^\top\right) \le \mathrm{Rang}(X) = r \\ &\Longrightarrow \mathrm{Rang}\left(X(X^\top X)^- X^\top\right) = r \\ &\Longrightarrow \mathrm{Spur}\left(X(X^\top X)^- X^\top\right) = r. \end{aligned}$$

□

Beweis des Satzes 3.3.7. Mit Hilfe des Lemmas 3.3.3 bekommt man

$$\begin{aligned}\overline{\sigma}^2 &= \frac{1}{n-r}\left|Y - X\overline{\beta}\right|^2 = \frac{1}{n-r}\left|Y - X(X^\top X)^- X^\top Y\right|^2 = \frac{1}{n-r}|DY|^2 \\ &= \frac{1}{n-r}\Big|\underbrace{DX}_{=0}\beta + D\varepsilon\Big|^2 = \frac{1}{n-r}|D\varepsilon|^2 = \frac{1}{n-r}\varepsilon^\top \underbrace{D^\top D}_{=D^2=D}\varepsilon = \frac{1}{n-r}\varepsilon^\top D\varepsilon.\end{aligned}$$

Deshalb gilt:

$$\begin{aligned}\mathbb{E}\,\overline{\sigma}^2 &= \frac{1}{n-r}\mathbb{E}\left(\varepsilon^\top D\varepsilon\right) = \frac{1}{n-r}\mathbb{E}\,\mathrm{Spur}\left(\varepsilon^\top D\varepsilon\right) = \frac{1}{n-r}\mathrm{Spur}\Big(D\cdot\mathbb{E}\,\big(\underbrace{\varepsilon\varepsilon^\top}_{\sigma^2\mathcal{I},\ \text{da } \varepsilon\sim N(0,\sigma^2\mathcal{I})}\big)\Big) \\ &= \frac{\sigma^2}{n-r}\cdot\mathrm{Spur}(D) = \sigma^2 \text{ nach Lemma 3.3.3, 4.), weil } \mathbb{E}\,\varepsilon\varepsilon^\top = \sigma^2\mathcal{I} \\ &\quad \text{wegen } \varepsilon\sim N(0,\sigma^2\mathcal{I}).\end{aligned}$$

□

Satz 3.3.8. Es gelten folgende Verteilungseigenschaften:

1. $\overline{\beta}\sim N\left((X^\top X)^- X^\top X\beta,\ \sigma^2(X^\top X)^-(X^\top X)\left((X^\top X)^-\right)^\top\right)$,
2. $\frac{(n-r)\overline{\sigma}^2}{\sigma^2}\sim\chi^2_{n-r}$,
3. $\overline{\beta}$ und $\overline{\sigma}^2$ sind unabhängig.

Beweis. 1. Es gilt:

$$\overline{\beta} = (X^\top X)^- X^\top Y = (X^\top X)^- X^\top(X\beta+\varepsilon) = \underbrace{(X^\top X)^- X^\top X\beta}_{=\mu} + \underbrace{(X^\top X)^- X^\top}_{=A}\varepsilon$$

und mit der Definition von $N(\cdot,\cdot)$ bekommt man

$$\begin{aligned}&\overline{\beta} = N\left(\mu,\ \sigma^2 AA^\top\right) = N\left((X^\top X)^- X^\top X\beta,\ \sigma^2(X^\top X)^- X^\top X((X^\top X)^-)^\top\right) \\ &\text{mit } AA^\top = (X^\top X)^- X^\top X((X^\top X)^-)^\top\end{aligned}$$

2. Es gilt $\overline{\sigma}^2 = \frac{1}{n-r}\varepsilon^\top D\varepsilon$ aus dem Beweis des Satzes 3.3.7. Deshalb

$$\frac{(n-r)\overline{\sigma}^2}{\sigma^2} = \underbrace{\left(\frac{\varepsilon}{\sigma}\right)^\top}_{\sim N(0,\mathcal{I})} D\left(\frac{\varepsilon}{\sigma}\right) \overset{(\text{Satz 3.1.8})}{=} \chi^2_{n-r}.$$

3. Betrachten wir $A\varepsilon$ und $\varepsilon^\top D\varepsilon$. Es genügt zu zeigen, daß sie unabhängig sind, um die Unabhängigkeit von $\overline{\beta}$ und $\overline{\sigma}^2$ zu beweisen, weil $\overline{\beta} = \mu + A\varepsilon$, $\overline{\sigma}^2 = \frac{1}{n-r}\varepsilon^\top D\varepsilon$. Es gilt: $A\cdot\sigma^2\mathcal{I}\cdot D = 0$. Nach Satz 3.1.9 sind dann $A\varepsilon$ und $\varepsilon^\top D\varepsilon$ unabhängig.

□

3.3.5 Hypothesentests

Betrachten wir die Hypothesen $H_0 : H\beta = d$ vs. $H_1 : H\beta \neq d$, wobei H eine $(s \times m)$-Matrix $(s \leq m)$ mit $\text{Rang}(H) = s$ ist, und $d \in \mathbb{R}^s$.

Im Satz 3.2.10 haben wir im Fall $\text{Rang}(X) = r = m$ folgende Testgröße dafür betrachtet:

$$T = \frac{(H\hat{\beta} - d)^\top (H(X^\top X)^{-1} H^\top)^{-1} (H\hat{\beta} - d)}{s\hat{\sigma}^2} \overset{(H_0)}{\sim} F_{s,n-m}.$$

Im allgemeinen Fall betrachten wir

$$T = \frac{(H\overline{\beta} - d)^\top (H(X^\top X)^- H^\top)^{-1} (H\overline{\beta} - d)}{s\overline{\sigma}^2}. \tag{3.3.5}$$

Wir wollen zeigen, daß $T \overset{(H_0)}{\sim} F_{s,n-r}$. Dann wird H_0 verworfen, falls $T > F_{s,n-r,1-\alpha}$. Dies ist ein Test zum Niveau $\alpha \in (0, 1)$.

Definition 3.3.3. Die Hypothese $H_0 : H\beta = d$ heißt *testbar*, falls alle Koordinaten des Vektors $H\beta$ schätzbare Funktionen sind.

Satz 3.3.3 gibt Bedingungen an H an, unter denen $H_0 : H\beta = d$ testbar ist.

Lemma 3.3.4. Die Hypothese $H_0 : H\beta = d$ ist testbar genau dann, wenn

1. $\exists\, (s \times n)$-Matrix $C : H = CX$, oder
2. $H(X^\top X)^- X^\top X = H$.

Beweis. Wir zeigen, daß die Testgröße T in (3.3.5) wohldefiniert ist, das heißt, die $(s \times s)$-Matrix $H(X^\top X)^- H^\top$ positiv definit und damit invertierbar ist. Aus Folgerung 3.3.1 haben wir $X^\top X = P^{-1} \begin{pmatrix} \mathcal{I}_r & 0 \\ 0 & 0 \end{pmatrix} P^{-1}$ für eine $(m \times m)$-Matrix P, die invertierbar und symmetrisch ist. Deshalb gilt

$$(X^\top X)^- = P \cdot \begin{pmatrix} \mathcal{I}_r & 0 \\ 0 & \mathcal{I}_{m-r} \end{pmatrix} P = P \cdot P,$$

das heißt, daß es eine eindeutige verallgemeinerte Inverse von $X^\top X$ mit dieser Darstellung gibt. Daraus folgt, daß die $(s \times s)$-Matrix $HPPH^\top = (PH^\top)^\top \cdot PH^\top$ positiv definit ist, weil $\text{Rang}(PH^\top) = s$. Sei nun $(X^\top X)^-$ eine beliebige verallgemeinerte Inverse von $X^\top X$. Dann ist mit Lemma 3.3.4

$$H(X^\top X)^- H^\top = CX(X^\top X)X^\top C^\top = CXPPX^\top C^\top = HPPH^\top,$$

denn $X(X^\top X)^- X^\top$ ist invariant bezüglich der Wahl von $(X^\top X)^-$, laut Beweis des Satzes 3.3.6. Also ist $H\left(X^\top X\right)^- H^\top$ positiv definit für eine beliebige verallgemeinerte Inverse $\left(X^\top X\right)^-$ und die Testgröße T somit wohldefiniert. $\square$

Satz 3.3.9. Falls $H_0 : H\beta = d$ testbar ist, dann gilt $T \overset{(H_0)}{\sim} F_{s,n-r}$.

Beweis. Ähnlich, wie in Satz 3.2.10 gilt

$$H\overline{\beta} - d = H(X^\top X)^- X^\top (X\beta + \varepsilon) - d = \underbrace{H(X^\top X)^- X^\top X\beta - d}_{=\mu} + \underbrace{H(X^\top X)^- X^\top}_{=B} \varepsilon.$$

Zeigen wir, daß $\mu \overset{(H_0)}{=} 0$.

$$\mu \overset{(\text{Lemma 3.3.4})}{=} C \cdot \underbrace{X(X^\top X)^- X^\top X}_{=X \text{ (Lemma 3.3.2, 2.))}} \cdot \beta - d = CX\beta - d = H\beta - d \overset{(H_0)}{=} 0.$$

Nach Satz 3.3.8 sind $(H\overline{\beta} - d)^\top \left(H(X^\top X)^- H^\top\right)^{-1} \left(H\overline{\beta} - d\right)$ und $s \cdot \overline{\sigma}^2$ unabhängig, $\frac{(n-r)\overline{\sigma}^2}{\sigma^2} \sim \chi^2_{n-r}$. Also bleibt nur noch zu zeigen, daß

$$\Big(\underbrace{H\overline{\beta} - d}_{=\varepsilon^\top B^\top}\Big)^\top \left(H(X^\top X)^- H^\top\right)^{-1} \Big(\underbrace{H\overline{\beta} - d}_{=B\varepsilon}\Big) \overset{(H_0)}{\sim} \chi^2_s.$$

Es gilt

$$\begin{aligned}
&\varepsilon^\top B^\top \left(H(X^\top X)^- H^\top\right)^{-1} B\varepsilon \\
&= \varepsilon^\top \underbrace{X\left((X^\top X)^-\right)^\top H^\top \left(H(X^\top X)^- H^\top\right)^{-1} H(X^\top X)^- X^\top}_{A} \varepsilon
\end{aligned}$$

Man kann leicht zeigen, daß A symmetrisch, idempotent und $\operatorname{Rang}(A) = s$ ist. Zeigen wir zum Beispiel die Idempotenz:

$$\begin{aligned}
A^2 &= X\left((X^\top X)^-\right)^\top H^\top \left(H(X^\top X)^- H^\top\right)^{-1} \underbrace{H(X^\top X)^- X^\top X}_{H \text{ (Lemma 3.3.4, 2.))}} \left((X^\top X)^-\right)^T H^\top \cdot \\
&\quad \cdot \left(H(X^\top X)^- H^\top\right)^{-1} H(X^\top X)^- X^\top \\
&= X\left((X^\top X)^-\right)^\top H^\top \left(H(X^\top X)^- H^\top\right)^{-1} H(X^\top X)^- X^\top = A,
\end{aligned}$$

weil $\left((X^\top X)^-\right)^\top$ auch eine verallgemeinerte Inverse von $X^\top X$ ist (nach Lemma 3.3.2). Somit hängt auch $H(X^\top X)^- H^\top = CX(X^\top X)^- X^\top C^\top$ nicht von der Wahl von $(X^\top X)^-$ ab, vgl. den Beweis des Satzes 3.3.6. Nach Satz 3.1.8 ist $\varepsilon^\top A\varepsilon \sim \chi^2_s$, wegen $\varepsilon \sim N(0, \sigma^2 \mathcal{I})$ und somit $T \overset{H_0}{\sim} F_{s,n-r}$. □

3.3.6 Konfidenzbereiche

Ähnlich wie in Abschnitt 3.2.5 werden wir Konfidenzbereiche für unterschiedliche Funktionen vom Parametervektor β angeben. Aus dem Satz 3.3.9 ergibt sich unmittelbar folgender Konfidenzbereich zum Niveau $1-\alpha \in (0,1)$:

Folgerung 3.3.1. Sei $Y = X\beta + \varepsilon$ ein multivariates Regressionsmodell mit $\text{Rang}(X) = r < m$, H eine $(s \times m)$-Matrix mit $\text{Rang}(H) = s$, $s \in \{1, \dots, m\}$ und $H_0 : H\beta = d$ testbar $\forall d \in \mathbb{R}^s$. Dann ist

$$\left\{ d \in \mathbb{R}^s : \frac{\left(H\overline{\beta} - d\right)^\top \left(H(X^\top X)^- H^\top\right)^{-1} \left(H\overline{\beta} - d\right)}{s \cdot \overline{\sigma}^2} \leq F_{s,\, n-r,\, 1-\alpha} \right\}$$

ein Konfidenzbereich für $H\beta$ zum Niveau $1-\alpha$.

Folgerung 3.3.2. Sei $h^\top \beta$ eine schätzbare lineare Funktion von β, $h \in \mathbb{R}^m$. Dann ist

$$\left(h^\top \overline{\beta} - t_{n-1,\, 1-\alpha/2} \cdot \overline{\sigma} \sqrt{h^\top (X^\top X)^- h},\; h^\top \overline{\beta} + t_{n-1,\, 1-\alpha/2} \cdot \overline{\sigma} \sqrt{h^\top (X^\top X)^- h} \right)$$

ein Konfidenzintervall für $h\beta$ zum Niveau $1-\alpha$.

Beweis. Setzen wir $s = 1$ und $H = h^\top$. Aus Satz 3.3.9 folgt

$$\begin{aligned} T &= \frac{\left(h^\top \overline{\beta} - d\right)^\top \left(h^\top (X^\top X)^- h\right)^{-1} \left(h^\top \overline{\beta} - d\right)}{\overline{\sigma}^2} = \frac{\left(h^\top \overline{\beta} - d\right)^\top \left(h^\top \overline{\beta} - d\right)}{\overline{\sigma}^2 \left(h^\top (X^\top X)^- h\right)} \\ &= \frac{\left(h^\top \overline{\beta} - d\right)^2}{\overline{\sigma}^2 \left(h^\top (X^\top X)^- h\right)} \sim F_{1,\, n-r} \end{aligned}$$

unter der Voraussetzung $h^\top \beta = d$, weil $h^\top X^\top X h$ eindimensional (eine Zahl) ist. Deshalb gilt

$$\sqrt{T} = \frac{h^\top \beta - h^\top \overline{\beta}}{\overline{\sigma} \sqrt{h^\top (X^\top X)^- h}} \sim t_{n-r}$$

und somit

$$\mathbb{P}\left(-t_{n-r, 1-\alpha/2} \leq \sqrt{T} \leq t_{n-r, 1-\alpha/2} \right) = 1 - \alpha.$$

Daraus folgt das obige Konfidenzintervall. $\square$

Man kann sogar eine stärkere Version von 3.3.2 beweisen, die für alle h aus einem linearen Unterraum gilt:

Satz 3.3.10 (Konfidenzband von Scheffé)**.** Sei $H = (h_1, \ldots, h_s)^\top$, $h_1, \ldots, h_2 \in \mathbb{R}^m$, $1 \le s \le m$ und $H_0 : H\beta = d$ testbar $\forall d \in \mathbb{R}^s$. Sei $\mathrm{Rang}(M) = s$ und $\mathcal{L} =< h_1, \ldots, h_s >$ der lineare Unterraum, der von den Vektoren $h_1, \ldots, h_s$ aufgespannt wird. Dann gilt:

$$\mathbb{P}_\beta\left(\max_{h\in\mathcal{L}}\left\{\frac{\left(h^\top\beta - h^\top\overline{\beta}\right)^2}{\overline{\sigma}^2 h^\top(X^\top X)^- h}\right\} \le sF_{s,\,n-r,\,1-\alpha}\right) = 1-\alpha$$

Somit ist

$$\left[h^\top\overline{\beta} - \sqrt{sF_{s,\,n-r,\,1-\alpha}}\cdot\overline{\sigma}\sqrt{h^\top(X^\top X)^- h},\; h^\top\overline{\beta} + \sqrt{sF_{s,\,n-r,\,1-\alpha}}\cdot\overline{\sigma}\sqrt{h^\top(X^\top X)^- h}\right]$$

ein (gleichmäßiges bzgl. $h \in \mathcal{L}$) Konfidenzintervall für $h^\top\beta$.

Beweis. Aus dem Satz 3.3.9 folgt $\forall\alpha \in (0,1)$:

$$\mathbb{P}\Big(\underbrace{\left(H\overline{\beta} - H\beta\right)^\top\left(H(X^\top X)^- H^\top\right)^{-1}\left(H\overline{\beta} - H\beta\right)}_{T_1} \le s\cdot\overline{\sigma}^2 F_{s,\,n-r,\,1-\alpha}\Big) = 1-\alpha.$$

Falls wir zeigen können, daß

$$T_1 = \max_{x\in\mathbb{R}^s,\,x\neq 0}\left\{\frac{\left(x^\top\left(H\overline{\beta} - H\beta\right)\right)^2}{x^\top\left(H(X^\top X)^- H^\top\right)x}\right\}, \tag{3.3.6}$$

dann ist der Satz bewiesen, denn

$$\begin{aligned}
1-\alpha = \mathbb{P}\Big(T_1 \le \underbrace{s\overline{\sigma}^2 F_{s,\,n-r,\,1-\alpha}}_{t}\Big) &= \mathbb{P}\left(\max_{x\in\mathbb{R}^s,\,x\neq 0}\left\{\frac{\left(x^\top\left(H\overline{\beta} - H\beta\right)\right)^2}{x^\top\left(H(X^\top X)^- H^\top\right)x}\right\} \le t\right)\\
&= \mathbb{P}\left(\max_{x\in\mathbb{R}^s,\,x\neq 0}\left\{\frac{\left((H^\top x)^\top\overline{\beta} - (H^\top x)^\top\beta\right)^2}{(H^\top x)^\top(X^\top X)^-(H^\top x)}\right\} \le t\right) \quad \text{und weil } H^\top x = h \in \mathcal{L}\\
&= \mathbb{P}\left(\max_{h\in\mathcal{L}}\left\{\frac{\left(h^\top\overline{\beta} - h^\top\beta\right)^2}{h^\top(X^\top X)^- h}\right\} \le s\overline{\sigma}^2 F_{s,\,n-r,\,1-\alpha}\right).
\end{aligned}$$

Also, zeigen wir die Gültigkeit von (3.3.6). Es genügt zu zeigen, daß T_1 die obere Schranke von

$$\frac{\left(x^\top(H\overline{\beta} - H\beta)\right)^2}{x^\top\left(H(X^\top X)^- H^\top\right)x}$$

darstellt, die auch angenommen wird. Da $H(X^\top X)^- H^\top$ positiv definit ist und invertierbar, existiert eine invertierbare $(s \times s)$-Matrix B mit der Eigenschaft $BB^\top = H(X^\top X)^- H^\top$. Dann gilt

$$\begin{aligned}
\left(x^\top (H\overline{\beta} - H\beta)\right)^2 &= \Big(\underbrace{x^\top B}_{(B^\top x)^\top} \cdot B^{-1}(H\overline{\beta} - H\beta)\Big)^2 \\
&\le |B^\top x|^2 - |B^{-1}(H\overline{\beta} - H\beta)|^2 \quad \text{(wegen der Ungleichung von Cauchy-Schwarz)} \\
&= x^\top B B^\top x \left(H\overline{\beta} - H\beta\right)^\top \cdot \underbrace{(B^{-1})^\top B^{-1}}_{=(B^\top)^{-1}B^{-1} = (BB^\top)^{-1}}(H\overline{\beta} - H\beta) \\
&= x^\top H(X^\top X)^- H^\top x \cdot \left(H\overline{\beta} - H\beta\right)^\top \left(H(X^\top X)^- H^\top\right)^{-1} (H\overline{\beta} - H\beta).
\end{aligned}$$

Somit gilt

$$\frac{\left(x^\top (H\overline{\beta} - H\beta)\right)^2}{x^\top \left(H(X^\top X)^- H^\top\right) x} \le \left(H\overline{\beta} - H\beta\right)^\top \left(H(X^\top X)^- H^\top\right)^{-1} \left(H\overline{\beta} - H\beta\right) = T_1.$$

Man kann leicht prüfen, daß diese Schranke für $x = \left(H(X^\top X)^- H^\top\right)^{-1} \left(H\overline{\beta} - H\beta\right)$ angenommen wird. □

3.3.7 Einführung in die Varianzanalyse

In diesem Abschnitt geben wir ein Beispiel für die Verwendung linearer Modelle mit Design-Matrix, die keinen vollen Rang besitzt. Dabei handelt es sich um die Aussage der *Variabilität der Erwartungswerte* in der Stichprobe $Y = (Y_1, \ldots, Y_n)^\top$, die auf englisch *analysis of variance*, kurz *ANOVA*, heißt. Später werden wir auch denselben Begriff *Varianzanalyse* dafür verwenden.

Betrachten wir zunächst die *einfaktorielle Varianzanalyse*, bei der man davon ausgeht, daß die Stichprobe $(Y_1, \ldots, Y_n)$ in k homogene Teilklassen $(Y_{ij},\ j = 1, \ldots, n_i)$, $i = 1, \ldots, k$ zerlegbar ist, mit den Eigenschaften:

1. $\mathbb{E}\,(Y_{ij}) = \mu_i = \mu + \alpha_i, \quad j = 1, \ldots, n_i, \quad i = 1, \ldots, k.$
2. $n_i > 1, \quad i = 1, \ldots, k, \quad \sum\limits_{i=1}^{k} n_i = n, \quad \sum\limits_{i=1}^{k} n_i \alpha_i = 0.$

Dabei ist μ ein Faktor, der allen Klassen gemeinsam ist, und α_i verkörpert die *klassenspezifischen Differenzen* zwischen den Erwartungswerten $\mu_1, \ldots, \mu_k$. Die Nummer $i = 1, \ldots, k$ der Klassen wird als *Stufe eines Einflussfaktors* (zum Beispiel die Dosis eines Medikaments in einer klinischen Studie) und α_i, $i = 1, \ldots, k$ als *Effekt* der i-ten Stufe gedeutet. Die Nebenbedingung $\sum\limits_{i=1}^{k} n_i \alpha_i = 0$ bewirkt, daß die Umrechnung

$(\mu_1, \ldots, \mu_k) \longleftrightarrow (\mu, \alpha_1, \ldots, \alpha_k)$ eindeutig wird und daß $\mu = \frac{1}{n} \sum_{i=1}^{k} \sum_{j=1}^{n_i} \mathbb{E}\, Y_{ij}$. Es wird vorausgesetzt, daß μ_i mit unkorrelierten Meßfehlern ε_{ij} gemessen werden kann, das heißt

$$Y_{ij} = \mu_i + \varepsilon_{ij} = \mu + \alpha_i + \varepsilon_{ij}, \quad i = 1, \ldots, k,\ j = 1, \ldots, n_i \tag{3.3.7}$$

$$\mathbb{E}\, \varepsilon_{ij} = 0, \quad \operatorname{Var} \varepsilon_{ij} = \sigma^2, \varepsilon_{ij} \text{ unkorreliert}, \ i = 1, \ldots, k,\ j = 1, \ldots, n_i. \tag{3.3.8}$$

Es soll die *klassische ANOVA-Hypothese* getestet werden, daß *keine* Variabilität in den Erwartungswerten μ_i auffindbar ist:

$$H_0: \quad \mu_1 = \mu_2 = \ldots = \mu_k,$$

was bedeutet, daß

$$H_0: \quad \alpha_1 = \alpha_2 = \ldots = \alpha_k.$$

Die Problemstellung (3.3.7) kann in der Form der multivariaten linearen Regression folgendermaßen umgeschrieben werden:

$$\begin{aligned}
&Y = X\beta + \varepsilon, \text{ wobei } Y = (Y_{11}, \ldots, Y_{1n_1}, Y_{21}, \ldots, Y_{2n_2}, \ldots, Y_{k1}, \ldots, Y_{kn_k})^\top, \\
&\beta = (\mu, \alpha_1, \ldots, \alpha_k)^\top, \\
&\varepsilon = (\varepsilon_{11}, \ldots, \varepsilon_{1n_1}, \ldots, \varepsilon_{k1}, \ldots, \varepsilon_{kn_k})^\top,
\end{aligned}$$

$$X = \begin{pmatrix}
1 & 1 & 0 & \ldots & \ldots & 0 \\
1 & 1 & 0 & \ldots & \ldots & 0 \\
\vdots & & & & & \\
1 & 1 & 0 & \ldots & \ldots & 0 \\
1 & 0 & 1 & 0 & \ldots & 0 \\
\vdots & & & & & \\
1 & 0 & 1 & 0 & \ldots & 0 \\
\vdots & & & & & \\
1 & 0 & \ldots & \ldots & 0 & 1 \\
\vdots & & & & & \\
1 & 0 & \ldots & \ldots & 0 & 1
\end{pmatrix}
\begin{array}{l}
\left.\begin{array}{l} \\ \\ \\ \\ \end{array}\right\} n_1 \\
\left.\begin{array}{l} \\ \\ \\ \end{array}\right\} n_2 \\
\ \vdots \\
\left.\begin{array}{l} \\ \\ \\ \end{array}\right\} n_k
\end{array}$$

Die $(n \times ((k+1))$-Matrix X hat den Rang $k < m = k+1$, somit ist die Theorie von Abschnitt 3.3 auf diesem Modell komplett anwendbar.

Bei der *zweifaktoriellen Varianzanalyse* wird die Stichprobe $(Y_1, \ldots, Y_n)$ in Abhängigkeit von 2 Faktoren in $k_1 \cdot k_2$ homogene Gruppen aufgeteilt:

$$Y_{i_1 i_2 j}, \quad j = 1, \ldots, n_{i_1 i_2}$$

für $i_1 = 1, \ldots, k_1$, $i_2 = 1, \ldots, k_2$, sodaß

$$\sum_{i_1=1}^{k_1} \sum_{i_2=1}^{k_2} n_{i_1 i_2} = n.$$

Hier wird angenommen, daß

$$\mathbb{E}\,Y_{i_1 i_2 j} = \mu_{i_1 i_2} = \mu + \alpha_{i_1} + \beta_{i_2} + \gamma_{i_1 i_2}, \quad i_1 = 1, \ldots, k_1,\ i_2 = 1, \ldots, k_2,$$

somit stellt man folgendes lineares Modell auf:

$$\begin{aligned} Y_{i_1 i_2 j} = \mu_{i_1 i_2} + \varepsilon_{i_1 i_2 j} = \mu + \alpha_{i_1} + \beta_{i_2} + \gamma_{i_1 i_2} + \varepsilon_{i_1 i_2 j}, \\ j = 1, \ldots, n_{i_1 i_2},\ i_1 = 1, \ldots, k_1,\ i_2 = 1, \ldots, k_2. \end{aligned}$$

Übungsaufgabe 3.3.1. Schreiben Sie die Design-Matrix X für diesen Fall explizit auf! Zeigen Sie, daß sie wieder keinen vollen Rang besitzt.

Literaturverzeichnis

[1] Bickel, P. ; Doksum, K.: *Mathematical Statistics: Basic Ideas and Selected Topics.* 2nd edition, volume 1. London : Prentice Hall, 2001

[2] Casella, G. ; Berger, R. L.: *Statistical Inference.* 2nd edition. Duxbury : Pacific Grove (CA), 2002

[3] Fahrmeir, L. ; Künstler, R. ; I. Pigeot, G. T.: *Statistik. Der Weg zur Datenanalyse.* 3. Auflage. Berlin : Springer, 2001

[4] Georgii, H. O.: *Stochastik: Einführung in die Wahrscheinlichkeitstheorie und Statistik.* Berlin : de Gruyter, 2002

[5] Hartung, J. ; Elpert, B. ; Klösener, K. H.: *Statistik.* München : R. Oldenbourg Verlag, 1993. – 9. Auflage

[6] Irle, A.: *Wahrscheinlichkeitstheorie und Statistik, Grundlagen - Resultate - Anwendungen.* Teubner, 2001

[7] Koch, K. R.: *Parameter Estimation and Hypothesis Testing in Linear Models.* Berlin : Springer, 1999

[8] Krengel, U.: *Einführung in die Wahrscheinlichkeitstheorie und Statistik.* Braunschweig : Vieweg, 2002. – 6. Auflage

[9] L. Fahrmeir, T. K. ; Lang, S.: *Regression. Modelle, Methoden und Anwendungen.* Berlin : Springer, 2007

[10] Lehmann, E. L.: *Testing Statistical Hypothesis.* New York : Springer, 1999

[11] Maindonald, J. ; Braun, J.: *Data Analysis and Graphics Using R.* Cambridge University Press, 2003

[12] Pruscha, H.: *Angewandte Methoden der Mathematischen Statistik.* Stuttgart : Teubner, 2000

[13] Pruscha, H.: *Vorlesungen über Mathematische Statistik.* Stuttgart : Teubner, 2000

[14] Sachs, L.: *Angewandte Statistik.* Springer, 1992

[15] Sachs, L. ; Hedderich, J.: *Angewandte Statistik, Methodensammlung mit R.* 12. Auflage. Berlin : Springer, 2006

[16] Spiegel, M. R. ; Stephens, L. J.: *Statistik*. 3. Auflage. McGraw-Hill, 1999

[17] Stahel, W. A.: *Statistische Datenanalyse*. Vieweg, 1999

[18] Venables, W. ; Ripley, D.: *Modern applied statistics with S-PLUS*. 3rd edition. Springer, 1999

[19] Wasserman, L.: *All of Statistics. A Concise Course in Statistical Inference*. Springer, 2004

Index