Statistische Methoden

Dr. C.J. Luchsinger

4 Testtheorie

Literatur Kapitel 4

* Lindgren: Kapitel 9, 10
* Cartoon Guide: Kapitel 8
* Krengel: § 6, 14
* Stahel: Kapitel 8, 10, 11, 12

Aufbau dieses Kapitels:

4.1 θ_0 vs θ_1 (Neyman-Pearson stetig und diskret) [Training]

4.2 $\theta \leq \theta_0$ vs $\theta > \theta_0$ (im Fall von MLQ; UMP-Tests) [Münzwurf]

4.3 $\theta \in [\theta_1, \theta_2]$ vs $\theta \notin [\theta_1, \theta_2]$ (im Fall von MLQ; UMPU-Tests, Cox-Hinkley) [Blutdruck]

4.4 Invarianz ($Var[X] = \sigma_0^2$?, t-Test, F-Test, 2 Stichproben t-Test)

4.5 Miscellanea (χ^2-Tests, Kolmogorov-Smirnov, einfache Varianzanalyse, asymptotische Tests)

4.1 θ_0 vs θ_1 (Neyman-Pearson stetig und diskret)

StudentInnen, welche hier unsicher sind, konsultieren bitte Kapitel 6 in WTS!

4.1.1 Neyman-Pearson (Stetiger Fall)

Das Lemma von Neyman-Pearson ist höchstwahrscheinlich das einzige Resultat aus der Wahrscheinlichkeitstheorie und Statistik, bei dem der Satz und der Beweis im stetigen Fall einfacher ist als im diskreten Fall.

Die Ausgangslage ist die folgende:

* Wir haben n Datenpunkte $x_1, x_2, \ldots, x_n$.
* Diese Datenpunkte sind eine Stichprobe $X_1(\omega_1), X_2(\omega_1), \ldots, X_n(\omega_1)$, wobei die X_i's, $1 \leq i \leq n$, iid stetige Zufallsgrössen sind.
* Die gemeinsame Dichte ist entweder $f_0(x_1, x_2, \ldots, x_n)$ oder $f_1(x_1, x_2, \ldots, x_n)$. Wir wissen leider nicht, ob f_0 oder f_1 die richtige Dichte ist.
* Wir definieren Entscheidungs-Funktionen $d : \mathbb{R}^n \to \{0, 1\}$. Wir *vereinbaren*:
  * wenn $d(x_1, x_2, \ldots, x_n) = 0$, heisst es, dass wir uns für f_0 ($\mathcal{H}_0$) entscheiden,
  * wenn $d(x_1, x_2, \ldots, x_n) = 1$, heisst es, dass wir uns für f_1 ($\mathcal{H}_1$) entscheiden.
* $\alpha \in [0, 1]$ ist vorgegeben (z.B. 10 %).
* $K := K(\alpha)$ wählen wir so, dass

  $$P_0\left[\frac{f_1(X_1, X_2, \ldots, X_n)}{f_0(X_1, X_2, \ldots, X_n)} > K\right] = \alpha.$$

  Dabei ist P_0 die Wahrscheinlichkeit, wenn f_0 gilt. **Vorsicht:** Wir haben jetzt nicht mehr die Datenpunkte $x_1, x_2, \ldots, x_n$ eingesetzt, sondern die Zufallsgrössen $X_1, X_2, \ldots, X_n$; dies deshalb, weil wir *vor der Realisation* sagen: mit Wahrscheinlichkeit α darf ein Fehler 1. Art gemacht werden.
* Wahrscheinlichkeit für Fehler 1. Art (auch Risiko 1. Art oder Grösse des Tests oder Signifikanzniveau, engl. Size):

  $$P_0[d(X_1, X_2, \ldots, X_n) = 1] \quad (= \alpha).$$

* Wahrscheinlichkeit für Fehler 2. Art (auch Risiko 2. Art; Macht (engl. Power) ist $1-$ Risiko 2. Art):

  $$P_1[d(X_1, X_2, \ldots, X_n) = 0] \quad (= \beta).$$

Dann verhalten wir uns am Besten gemäss folgendem Lemma:

Satz 4.1 [Lemma von Neyman-Pearson, stetig] *Unter allen Entscheidungs-Funktionen d, welche mit Wahrscheinlichkeit von maximal α einen Fehler 1. Art machen, minimiert*

$$d^{NP}(x_1, x_2, \ldots, x_n) = 1 \Longleftrightarrow \frac{f_1(x_1, x_2, \ldots, x_n)}{f_0(x_1, x_2, \ldots, x_n)} > K$$

die Wahrscheinlichkeit für einen Fehler 2. Art.

Jargon und Bemerkungen: 1. Wir suchen einen Test mit maximaler Macht bei vorgegebener Grösse. *Amerikaner:* The hypothesis testing design problem is: Choose a test to maximize power subject to a pre-specified size.

2. Diesen Test nennt man auch "Likelihood-Quotienten-Test" oder engl. "Likelihood-Ratio-Test"; mehr zum Begriff "Likelihood" in Kapitel 5.

3. Wir vereinbaren noch für Satz 4.1 mit $c > 0$ beliebig $c/0 := \infty$; wie wir $0/0$ definieren ist egal, da dort offenbar beide Dichten 0 sind.

Illustration von Satz 4.1: $n = 1, \alpha = 0.1; \mathcal{N}(0,1)$ vs $\mathcal{N}(1,1)$. Wir skizzieren $f_0, f_1, f_1/f_0$ (Tipp: f_1/f_0 ist *hier* streng monoton wachsend! - zentral wichtige Annahme für fast ganz Kapitel 4), K, α, den Bereich, wo $f_1/f_0 > K$ (= Ablehnungsbereich) und die Wahrscheinlichkeit für den Fehler 2. Art (zuerst jedeR für sich auf einem Blatt und dann hier gemeinsam; Tipp: $qnorm(0.9) \doteq 1.281552$):

Beweis von Satz 4.1: d sei eine andere Entscheidungs-Funktion mit einer Wahrscheinlichkeit für einen Fehler 1. Art von höchstens α. Wir definieren $\mathbf{x} := (x_1, \ldots, x_n)$ und machen den Beweis für allgemeines n (am Anfang denke man aber einfachheitshalber an den Fall $n = 1$). Es gilt immer:

$$\int_{\mathbb{R}^n} \big(d^{NP}(\mathbf{x}) - d(\mathbf{x})\big)\big(f_1(\mathbf{x}) - f_0(\mathbf{x})K\big) d\mathbf{x} \geq 0.$$

Durch Ausmultiplizieren erhalten wir

$$\begin{aligned}
\int_{\mathbb{R}^n} d^{NP}(\mathbf{x}) f_1(\mathbf{x}) d\mathbf{x} - \int_{\mathbb{R}^n} d(\mathbf{x}) f_1(\mathbf{x}) d\mathbf{x} & \\
&\geq K\left(\int_{\mathbb{R}^n} d^{NP}(\mathbf{x}) f_0(\mathbf{x}) d\mathbf{x} - \int_{\mathbb{R}^n} d(\mathbf{x}) f_0(\mathbf{x}) d\mathbf{x}\right) \\
&= K\left(\alpha - \int_{\mathbb{R}^n} d(\mathbf{x}) f_0(\mathbf{x}) d\mathbf{x}\right) \\
&\geq 0,
\end{aligned}$$

da $\int_{\mathbb{R}^n} d(\mathbf{x}) f_0(\mathbf{x}) d\mathbf{x}$ das Risiko 1. Art der Entscheidungs-Funktion d ist. Dieses ist aber maximal α.

Auf der linken Seite haben wir andererseits genau die Macht der beiden Entscheidungsfunktionen. d^{NP} hat immer eine grössere Macht - anders ausgedrückt: die Wahrscheinlichkeit für einen Fehler 2. Art ist bei d^{NP} kleiner.

$\square$

Beispiele zum Lemma von Neyman-Pearson im stetigen Fall

Erstes Beispiel (aus WTS, revisited) Bei der Medikamentenentwicklung und -Zulassung treten Probleme der folgenden Art auf (stark vereinfachte Darstellung mit nicht realistischen Zahlen): ein Pharmakonzern behauptet, dass ihr Medikament den Blutdruck um 4 Einheiten senkt. Pharmakonzern und Zulassungsbehörde einigen sich darauf, den Blutdruck mit einer Normalverteilung zu modellieren. Aus epidemiologischen Messungen weiss man, dass dies mit $\mathcal{N}(\mu_0, 64)$ sehr gut stimmt (man kann dies auch in Frage stellen). μ_0 ist demnach eine bekannte Zahl, vielleicht 80. Wenn der Pharmakonzern recht hat, sollte das neue μ_1 vier Einheiten kleiner sein: $\mu_1 := \mu_0 - 4$. Wir haben 10 Personen zufällig ausgewählt, an denen das Medikament ausprobiert wurde (man kann sich fragen, ob es noch zufällig ist, wenn man nur unter Freiwilligen auswählen darf - eine Alternative existiert nicht wirklich!). Wie können wir mit dem Lemma von Neyman-Pearson dieses Problem angehen?

* $\mathcal{H}_0 : \mu = \mu_0, \mathcal{H}_1 : \mu = \mu_1$
* $\alpha = 0.05$ (darüber kann man streiten, je nach Gebiet anders)
* $\sigma^2 = 64$ wird angenommen (darüber könnte man streiten, vgl. auch 4.4)
* Die neuen Blutdruck-Messungen nach (korrekter) Medikamenteneinnahme seien

$$x_1, x_2, \ldots, x_{10}.$$

Testen mit dem Lemma von Neyman-Pearson: die gemeinsame Dichte ist (iid, $n = 10, \sigma^2 = 64, j \in \{0, 1\}$):

$$\begin{aligned}
f_j(x_1, \ldots, x_n) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i-\mu_j)^2}{2\sigma^2}} \\
&= \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu_j)^2} \\
&= \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left[-\frac{1}{2\sigma^2}\Big(\sum_{i=1}^{n} x_i^2 - 2\sum_{i=1}^{n}\mu_j x_i + n\mu_j^2\Big)\right] \\
&= \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left[-\frac{1}{2\sigma^2}\Big(\sum_{i=1}^{n} x_i^2 - 2\mu_j n\overline{x} + n\mu_j^2\Big)\right]
\end{aligned}$$

Wir müssen jetzt wegen Satz 4.1. den Quotienten f_1/f_0 untersuchen:

$$\begin{aligned}
\frac{f_1(x_1,\ldots,x_n)}{f_0(x_1,\ldots,x_n)} &= \frac{\left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left[-\frac{1}{2\sigma^2}\left(\sum_{i=1}^n x_i^2 - 2\mu_1 n\overline{x} + n\mu_1^2\right)\right]}{\left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left[-\frac{1}{2\sigma^2}\left(\sum_{i=1}^n x_i^2 - 2\mu_0 n\overline{x} + n\mu_0^2\right)\right]} \\
&= \frac{\exp\left[-\frac{1}{2\sigma^2}\left(-2\mu_1 n\overline{x} + n\mu_1^2\right)\right]}{\exp\left[-\frac{1}{2\sigma^2}\left(-2\mu_0 n\overline{x} + n\mu_0^2\right)\right]} \\
&= \exp\left[-\frac{1}{2\sigma^2}\left(-2\mu_1 n\overline{x} + 2\mu_0 n\overline{x} + n\mu_1^2 - n\mu_0^2\right)\right] \\
&= \exp\left[-\frac{1}{2\sigma^2}\left(-2n(\mu_1-\mu_0)\overline{x} + n(\mu_1^2-\mu_0^2)\right)\right]
\end{aligned}$$

Wir müssen jetzt ein $K = K(\alpha)$ finden, sodass das Risiko 1. Art gleich α ist. Dazu machen wir weitere Umformungen:

$$\exp\left[-\frac{1}{2\sigma^2}\left(-2n(\mu_1-\mu_0)\overline{x} + n(\mu_1^2-\mu_0^2)\right)\right] > K$$

$$-\frac{1}{2\sigma^2}\left(-2n(\mu_1-\mu_0)\overline{x} + n(\mu_1^2-\mu_0^2)\right) > \log(K)$$

Nach vielen weiteren Umformungsschritten (Achtung: Vorzeichen und $\mu_1 < \mu_0$) landet man bei:

$$\overline{x} < \frac{2\sigma^2\log(K) + n(\mu_1^2-\mu_0^2)}{2n(\mu_1-\mu_0)} =: K'.$$

Man beachte, dass $\overline{x}$ eine min suff Stat für dieses Problem ist! Wir werden also die Hypothese $\mathcal{H}_0$ ablehnen, sobald $\overline{x} < K'$. Aber wie ist K'? Das ist auf den ersten Blick ein schrecklich komplizierter Ausdruck. Aber: K' muss lediglich erfüllen, dass

$$P_0[\overline{X} < K'] = \alpha.$$

Unter der $\mathcal{H}_0$-Hypothese hat $\overline{X}$ eine $\mathcal{N}(80, 64/n)$-Verteilung (warum?). Damit können wir fortfahren ($n = 10$):

$$P_0[\overline{X} < K'] = P[\mathcal{N}(0, 6.4) < K' - 80] = P[\mathcal{N}(0,1) < (K'-80)/\sqrt{6.4}] = 0.05$$

$qnorm(0.05) = -1.644854$, also $(K' - 80)/\sqrt{6.4} = -1.644854 \Rightarrow K' \doteq 75.84$. Wie lautet nun das Resultat?

Bei einem Test mit Signifikanzniveau 0.05 wird man die $\mathcal{H}_0$-Hypothese ablehnen ($\mathcal{H}_1$-Hypothese annehmen), genau dann wenn

$$\overline{x} < 75.84. \tag{4.1}$$

Die Zahl $\mu_1 = 76$ kommt am Schluss *gar nicht mehr vor*! Dies ist eine Konsequenz davon, dass f_1/f_0 bei der Normalverteilung mit gleicher Varianz streng monoton ist. Die Konsequenzen dieses Resultats sind gewaltig (4.2 & 4.3 bauen darauf auf). Soviel vorausgeschickt: Der Test sieht offenbar immer genau gleich aus, sodass dies auch ein Test ist für die Hypothesen: $\mu = \mu_0$ gegen $\mu < \mu_0$.

$\overline{x}$ nennt man in diesem Beispiel übrigens auch die **Teststatistik**; das Ganze ist ein **statistischer Test**.

Ein weiterer Begriff wird hier noch kurz eingeführt, der **P-Wert**: Wenn Sie als WissenschaftlerIn langen Streitereien um "das richtige α" aus dem Weg gehen wollen, können Sie einfach mit dem Resultat $a := \overline{x}$ berechnen, wie gross die Wahrscheinlichkeit ist, dass unter der $\mathcal{H}_0$-Hypothese ein so extremer oder noch extremerer Wert herauskommt:

$$P_0[\overline{X} < a].$$

Das ist der P-Wert (von Versuchsausgang zu Versuchsausgang (leicht) verschieden). Wenn z.B. oben $a = 77$, so hat man einen P-Wert von 0.1178:

$$\begin{aligned} P_0[\overline{X} < 77] &= P[\mathcal{N}(80, 6.4) < 77] = P[\mathcal{N}(0, 6.4) < -3] = P[\mathcal{N}(0, 1) < -3/\sqrt{6.4}] \\ &= P[\mathcal{N}(0, 1) < -1.186] = pnorm(-1.186) \doteq 0.1178. \end{aligned}$$

Das heisst, dass man bei $\alpha = 0.15$ die $\mathcal{H}_0$-Hypothese ablehnt; bei "normalen" α's wird $\mathcal{H}_0$ hier aber nicht abgelehnt (sobald $\alpha < 0.1178$).

Zweites Beispiel In einer Aufgabe ist ein analoges Beispiel mit der Exponentialverteilung zu rechnen.

4.1.2 Neyman-Pearson (Diskreter Fall)

Wie bereits angekündigt, gibt es im diskreten Fall ein kleines Problem zu überwinden. Wir illustrieren dies am Besten anhand eines Beispiels:

$\mathcal{H}_0 : X$ ist eine Be(p)-Zufallsgrösse mit $p = 0.5$.

$\mathcal{H}_1 : X$ ist eine Be(p)-Zufallsgrösse mit $p = 0.9$.

$\alpha = 0.1, n = 1$. Wie sollen wir uns entscheiden, $x_1 \in \{0, 1\}$?

Die Ausgangslage ist die gleiche wie beim stetigen Fall mit kleinen Änderungen:

* Wir haben n Datenpunkte $x_1, x_2, \ldots, x_n$.
* Diese Datenpunkte sind eine Stichprobe $X_1(\omega_1), X_2(\omega_1), \ldots, X_n(\omega_1)$, wobei die X_i's, $1 \leq i \leq n$, iid diskrete Zufallsgrössen sind.
* Die gemeinsame Wahrscheinlichkeitsfunktion ist entweder $p_0(x_1, x_2, \ldots, x_n)$ oder $p_1(x_1, x_2, \ldots, x_n)$. Wir wissen leider nicht, ob p_0 oder p_1 die richtige Wahrscheinlichkeitsfunktion ist.
* Wir definieren Entscheidungs-Funktionen d^r : Datenraum $\rightarrow \{0, \gamma, 1\}, \gamma \in [0, 1)$. "r" steht für randomisiert. Wir *vereinbaren*: wenn $d^r(x_1, x_2, \ldots, x_n) = 0$, heisst es, dass wir uns für p_0 ($\mathcal{H}_0$) entscheiden, wenn $d^r(x_1, x_2, \ldots, x_n) = 1$, heisst es, dass wir uns für p_1 ($\mathcal{H}_1$) entscheiden, wenn $d^r(x_1, x_2, \ldots, x_n) = \gamma$, heisst es, dass wir uns mit Wahrscheinlichkeit γ für p_1 ($\mathcal{H}_1$) entscheiden (losgelöst vom Versuchsausgang, "Münze"). Man kann dies eigentlich auch zusammenfassen: $d^r(x_1, x_2, \ldots, x_n)$ gibt *in allen 3 Fällen* die Wahrscheinlichkeit an, mit der wir uns für p_1 ($\mathcal{H}_1$) entscheiden.
* $\alpha \in [0, 1]$ ist vorgegeben (z.B. 10 %).
* $(K, \gamma) := (K, \gamma)(\alpha)$ wählen wir so, dass (vgl. Gleichung (4.2) weiter hinten)

$$P_0\left[\frac{p_1(X_1, X_2, \ldots, X_n)}{p_0(X_1, X_2, \ldots, X_n)} > K\right] + \gamma P_0\left[\frac{p_1(X_1, X_2, \ldots, X_n)}{p_0(X_1, X_2, \ldots, X_n)} = K\right] = \alpha.$$

  Dabei ist P_0 die Wahrscheinlichkeit, wenn p_0 gilt.
* Wahrscheinlichkeit für Fehler 1. Art (auch Risiko 1. Art oder Grösse des Tests oder Signifikanzniveau, engl. Size):

$$P_0[d^r(X_1, X_2, \ldots, X_n) = 1] + \gamma P_0[d^r(X_1, X_2, \ldots, X_n) = \gamma] \quad (= \alpha).$$

* Wahrscheinlichkeit für Fehler 2. Art (auch Risiko 2. Art; Macht (engl. Power) ist $1-$ Risiko 2. Art):

$$P_1[d^r(X_1, X_2, \ldots, X_n) = 0] + (1 - \gamma)P_1[d^r(X_1, X_2, \ldots, X_n) = \gamma].$$

Dann verhalten wir uns am Besten gemäss folgendem Lemma:

Satz 4.2 [Lemma von Neyman-Pearson, diskret] *Unter allen Entscheidungs-Funktionen d^r, welche mit Wahrscheinlichkeit von maximal α einen Fehler 1. Art machen, minimiert*

$$d^{NP,r}(x_1, x_2, \ldots, x_n) = \begin{pmatrix} 1 \\ \gamma \\ 0 \end{pmatrix} \Longleftrightarrow \frac{p_1(x_1, x_2, \ldots, x_n)}{p_0(x_1, x_2, \ldots, x_n)} \begin{pmatrix} > \\ = \\ < \end{pmatrix} K$$

die Wahrscheinlichkeit für einen Fehler 2. Art.

Bemerkung: Auch hier vereinbaren wir mit $c > 0$ beliebig $c/0 := \infty$; wie wir $0/0$ definieren ist egal, da dort offenbar beide Wahrscheinlichkeitsfunktionen 0 sind.

Beweis von Satz 4.2: Man macht genau die gleichen Schritte wie beim Beweis zu Satz 4.1. Integrale werden zu Summen. Wir lassen dies als kleine, freiwillige Hausaufgabe.

$\square$

<u>Beispiele zum Lemma von Neyman-Pearson im diskreten Fall</u>

Vergleichen Sie die folgenden Umformungen jeweils mit den Schritten in 4.1.2; alles geht genau gleich bis auf die Bestimmung von K, γ.

* Stichprobe vom Umfang n aus Be(θ)-Zufallsgrössen, $x_i \in \{0, 1\}$:

$$x_1, x_2, \ldots, x_n$$

* $\mathcal{H}_0 : \theta = \theta_0, \mathcal{H}_1 : \theta = \theta_1; \theta_0 < \theta_1$
* $\alpha = 0.05$

Testen mit dem Lemma von Neyman-Pearson: die gemeinsame Wahrscheinlichkeitsfunktion ist (iid, $j \in \{0, 1\}$):

$$\begin{aligned} p_j(x_1, \ldots, x_n) &= \prod_{i=1}^{n} \theta_j^{x_i} (1 - \theta_j)^{1-x_i} \\ &= \theta_j^{\sum_{i=1}^{n} x_i} (1 - \theta_j)^{\sum_{i=1}^{n}(1-x_i)} \\ &= \theta_j^{\sum_{i=1}^{n} x_i} (1 - \theta_j)^{n-\sum_{i=1}^{n} x_i} \\ &= (1 - \theta_j)^n \left(\frac{\theta_j}{1 - \theta_j} \right)^{\sum_{i=1}^{n} x_i}. \end{aligned}$$

Wir müssen jetzt wegen Satz 4.2. den Quotienten p_1/p_0 untersuchen:

$$\begin{aligned}
\frac{p_1(x_1,\ldots,x_n)}{p_0(x_1,\ldots,x_n)} &= \frac{(1-\theta_1)^n\left(\frac{\theta_1}{1-\theta_1}\right)^{\sum_{i=1}^n x_i}}{(1-\theta_0)^n\left(\frac{\theta_0}{1-\theta_0}\right)^{\sum_{i=1}^n x_i}} \\
&= \frac{(1-\theta_1)^n}{(1-\theta_0)^n}\left(\frac{\theta_1(1-\theta_0)}{\theta_0(1-\theta_1)}\right)^{\sum_{i=1}^n x_i}.
\end{aligned}$$

Wir müssen jetzt $(K,\gamma) = (K,\gamma)(\alpha)$ finden, sodass das Risiko 1. Art gleich α ist. Dazu machen wir weitere Umformungen:

$$\frac{(1-\theta_1)^n}{(1-\theta_0)^n}\left(\frac{\theta_1(1-\theta_0)}{\theta_0(1-\theta_1)}\right)^{\sum_{i=1}^n x_i} * K,$$

wobei "*" hier für > oder = steht.

$$\sum_{i=1}^n x_i * \frac{\log\left(\frac{K(1-\theta_0)^n}{(1-\theta_1)^n}\right)}{\log\left(\frac{\theta_1(1-\theta_0)}{\theta_0(1-\theta_1)}\right)} =: K'.$$

Wir werden also die Hypothese $\mathcal{H}_0$ ablehnen, sobald $\sum_{i=1}^n x_i * K'$, mit Spezialfall wo $*$ das Gleichheitszeichen ist. Aber wie sind K',γ? Das ist auf den ersten Blick (wieder) ein schrecklich komplizierter Ausdruck. Aber: K',γ müssen lediglich erfüllen, dass

$$P_0\left[\sum_{i=1}^n X_i > K'\right] + \gamma P_0\left[\sum_{i=1}^n X_i = K'\right] = \alpha. \tag{4.2}$$

Unter der $\mathcal{H}_0$-Hypothese hat $\sum_{i=1}^n X_i$ eine Bin(n,θ_0)-Verteilung. Damit können wir fortfahren: Suche bei konkretem n das Paar $(K'-1,K')$, sodass $pbinom(K'-1,n,\theta_0) < 0.95$ und $pbinom(K',n,\theta_0) \geq 0.95$. Dann passt man noch das γ an, um das Risiko 1. Art voll auszuschöpfen.

In den Übungen ist dazu eine Aufgabe zu lösen, bei der man einfach mit Gleichung (4.2) arbeiten kann.

4.1.3 Mathematische Formalisierung des bisherigen Vorgehens: Verlustfunktion, Risiko, etc.

Wir wollen die bisherigen Resultate aus Kapitel 4 im Licht von "3.3 Mathematische Formalisierung" nochmals anschauen:

Der **Aktionsraum** ist $\mathcal{A} := \{\theta_0, \theta_1\}$. Die **Entscheidungsfunktion** geht vom Datenraum (z.B. $\mathbb{R}^n$) in den Aktionsraum. Wenn wir im diskreten Fall noch das γ einsetzen, müssen wir formal den Aktionsraum leicht erweitern ("unabhängiger Münzwurf") - wir gehen nicht weiter darauf ein. Als **Verlustfunktion** nehmen wir $L(\theta_0, \theta_0) := 0, L(\theta_0, \theta_1) := 1, L(\theta_1, \theta_0) := 1, L(\theta_1, \theta_1) := 0$. Mit obigen Formalisierungen stimmt unsere Definition von Risiko (1. und 2. Art) mit der Definition von Risiko aus 3.3 überein (Warum?).

Wir wollen die Risiko-Situation noch graphisch veranschaulichen. Dazu tragen wir im Einheitsquadrat (x-Achse ist Risiko 1. Art, y-Achse ist Risiko 2. Art) alle möglichen Tests ein:

* "Immer θ_0" sagen bzw. "immer θ_1" sagen.
* Voll-Randomisierte Tests: schaue Daten nicht an, sage mit Wahrscheinlichkeit u: "θ_0 richtig" und mit Wahrscheinlichkeit $1-u$ sage ich "θ_1 richtig".
* Test d und d^{ve} (d^{ve} macht immer genau Gegenteil von d, vgl. Aufgabe in den Übungen; **V**ulgär-**E**manzipation)
* Wo ist "Traumpunkt", mögliche Tests, konvexe Menge, zulässige Tests

4.2 $\theta \leq \theta_0$ vs $\theta > \theta_0$ (im Fall von MLQ; UMP-Tests)

Wir haben in den bisherigen Beispielen festgestellt, dass θ_1 in den konkreten Rechnungen keine Rolle spielt und der Quotient der Dichten eine streng monoton wachsende Funktion ist. Die Tests waren am Schluss von einer sehr einfachen und einleuchtenden Form (einseitiger Ablehnungsbereich). Dies muss nicht immer so sein. In einer Aufgabe ist mit Hilfe von Satz 4.1 ein optimaler Test in einem Umfeld zu konstruieren, wo all diese schönen Eigenschaften nicht vorhanden sind. Wir werden aber in 4.2.3 sehen (exponentielle Familie), dass sehr viele Verteilungen genau diese Eigenschaften haben. Dies rechtfertigt eine genauere Betrachtung in 4.2.1 und 4.2.2.

4.2.1 MLQ, Zusammengesetzte Hypothesen, Machtfunktion $\pi(\theta)$

Definition 4.4 [MLQ (engl. MLR)] *Das Testproblem für Parameter $\theta \in \mathbb{R}^1$ hat die MLQ (Monoton-Likelihood-Quotient)-Eigenschaft, wenn es eine minimal suffiziente Statistik $t(\mathbf{x})$: Datenraum $\rightarrow \mathbb{R}^1$ gibt, sodass mit $\theta_0 < \theta_1$ gilt:*

$$g_{\theta_0\theta_1}(t(\mathbf{x})) := \frac{p(\mathbf{x}, \theta_1)}{p(\mathbf{x}, \theta_0)}$$

ist eine streng monoton wachsende Funktion von t. p ist entweder eine Dichte oder eine Wahrscheinlichkeitsfunktion.

In allen bisherigen Beispielen war diese MLQ-Eigenschaft gegeben (Normalverteilung, Binomialverteilung, Exponentialverteilung).

Wir werden die gesamte Theorie hier nur für $\theta_0 < \theta_1$ und streng monoton *wachsend* formulieren und illustrieren; entsprechende Sätze gelten auch für die anderen Fälle (und sind schon vorgekommen!).

Wir haben in 4.1 sogenannte *einfache Hypothesen* behandelt: θ_0 und θ_1 sind Punkte. Wir werden jetzt aber sogenannte *zusammengesetzte Hypothesen* (engl. Compound) behandeln: $\theta \leq \theta_0$ vs $\theta > \theta_0$ hat ganze Intervalle als Hypothesen, in denen das "richtige" θ sein könnte.

Dies ist nicht weiter dramatisch; man muss jedoch das Risiko erster Art verallgemeinern - zudem führen wir gleich die Machtfunktion ein.

Definition 4.5 [Grösse des Tests bei zusammengesetzter Nullhypothese] *Bei allgemeiner Nullhypothese (einfach oder zusammengesetzt) definieren wir die Grösse des Tests d als:*

$$\sup_{\theta \in \mathcal{H}_0} E_\theta[d(\mathbf{X})] = \sup_{\theta \in \mathcal{H}_0} P_\theta[\text{sage "}\mathcal{H}_1\text{" mit Test } d].$$

Wir werden fordern, dass diese Grösse $\leq \alpha$ sein wird.

Wir wollen das Risiko 1. Art auf α beschränken. Wir wissen aber nicht, welches $\theta \in \mathcal{H}_0$ das richtige ist. Die Idee ist, dass man den schlimmsten Fall annimmt ("sup"). Wegen MLQ liegt der übrigens genau an der Grenze, wie in 4.2.2 gezeigt wird.

Definition 4.6 [Machtfunktion $\pi(\theta)$ (engl. Powerfunction)] *Die Machtfunktion des Tests d definieren wir als*

$$\pi(\theta) := E_\theta[d(\mathbf{X})].$$

Die Machtfunktion gibt in Abhängigkeit von θ an, mit welcher Wahrscheinlichkeit wir uns für $\mathcal{H}_1$ entscheiden.

Im Fall von $\mathcal{N}(\theta, 1)$, wo $\mathcal{H}_0 : \theta \leq 0$ vs $\mathcal{H}_1 : \theta > 0$, haben wir folgende Situation (Vorgriff auf 4.2.2):

4.2.2 UMP-Tests

Der Hauptsatz von 4.2 ist Satz 4.9. Er wird schrittweise in 3 Sätzen entwickelt:

Satz 4.7 [$\theta = \theta_0$ **vs** $\theta > \theta_0$] *Wenn MLQ gilt, existiert ein Test der Grösse α, welcher im Testproblem $\theta = \theta_0$ vs $\theta > \theta_0$ auf ganz $\theta > \theta_0$ (eben* ***u****niform) die Macht $\pi(\theta)$ maximiert.*

Beweis von Satz 4.7 (stetig, diskret analog): Wir nehmen ein $\theta_1 > \theta_0$ beliebig und konstruieren einen Test d^{NP} nach Satz 4.1 (Lemma von Neyman-Pearson). Wir haben damit

$$d^{NP}(\mathbf{x}) := \begin{pmatrix} 1 \\ 0 \end{pmatrix} \Leftrightarrow \frac{f_{\theta_1}(\mathbf{x})}{f_{\theta_0}(\mathbf{x})} \begin{pmatrix} > \\ \leq \end{pmatrix} K.$$

Wegen der MLQ-Eigenschaft ist dies gleich

$$d^{NP}(\mathbf{x}) := \begin{pmatrix} 1 \\ 0 \end{pmatrix} \Leftrightarrow g_{\theta_0 \theta_1}(t(\mathbf{x})) \begin{pmatrix} > \\ \leq \end{pmatrix} K$$

und wegen *streng* monoton wachsend

$$d^{NP}(\mathbf{x}) := \begin{pmatrix} 1 \\ 0 \end{pmatrix} \Leftrightarrow t(\mathbf{x}) \begin{pmatrix} > \\ \leq \end{pmatrix} g^{-1}_{\theta_0 \theta_1}(K) =: K'.$$

Dieses K' wird jetzt derart gewählt, dass das Risiko 1. Art gleich α ist:

$$E_{\theta_0}[d^{NP}(\mathbf{X})] = P_{\theta_0}[t(\mathbf{X}) > K'] = \alpha. \tag{4.3}$$

In (4.3) kommt θ_1 nicht mehr vor (auch im diskreten Fall nicht). Dieser Test sieht also immer genau gleich aus, egal wie wir $\theta_1 > \theta_0$ wählen. Aber für jedes θ_1 maximiert Satz 4.1 die Macht - und diesen Test haben wir gebraucht.

$\square$

Wir sehen hier auch, dass der Test am Schluss *einseitig* ist: die $\mathcal{H}_0$-Hypothese wird abgelehnt, genau dann wenn

$$t(\mathbf{x}) > K'.$$

An dieser Stelle schreiben wir von jetzt an gleich sofort K statt K', weil dies die entscheidende Gleichung ist.

Folgende Darstellung illustriert (im Fall der Normalverteilung) Satz 4.8:

Satz 4.8 [$\pi(\theta) \nearrow$ **in** θ] *Wenn wir MLQ haben, so ist $\pi(\theta)$ eine wachsende Funktion von θ, wo der Test d^{NP} und somit K, γ fix.*

Beweis von Satz 4.8 (stetig, diskret analog): Sei $a < b$ beliebig. Die Daten seien aus dem Datenraum W, z.B. $W = \mathbb{R}^n$. Wir haben einerseits

$$\begin{aligned}
\pi(b) = P_b[t(\mathbf{X}) > K] = E_b[I[t(\mathbf{X}) > K]] &= \int_W I[t(\mathbf{x}) > K] f_b(\mathbf{x}) d\mathbf{x} \\
&= \int_W I[t(\mathbf{x}) > K] f_a(\mathbf{x}) g_{ab}(t(\mathbf{x})) d\mathbf{x} \\
&\geq \int_W I[t(\mathbf{x}) > K] f_a(\mathbf{x}) g_{ab}(K) d\mathbf{x} \\
&= g_{ab}(K) \int_W I[t(\mathbf{x}) > K] f_a(\mathbf{x}) d\mathbf{x} \\
&= g_{ab}(K) \pi(a).
\end{aligned}$$

Andererseits gilt analog

$$\begin{aligned}
1 - \pi(b) = P_b[t(\mathbf{X}) \leq K] = E_b[I[t(\mathbf{X}) \leq K]] &= \int_W I[t(\mathbf{x}) \leq K] f_b(\mathbf{x}) d\mathbf{x} \\
&= \int_W I[t(\mathbf{x}) \leq K] f_a(\mathbf{x}) g_{ab}(t(\mathbf{x})) d\mathbf{x} \\
&\leq \int_W I[t(\mathbf{x}) \leq K] f_a(\mathbf{x}) g_{ab}(K) d\mathbf{x} \\
&= g_{ab}(K) \int_W I[t(\mathbf{x}) \leq K] f_a(\mathbf{x}) d\mathbf{x} \\
&= g_{ab}(K)(1 - \pi(a)).
\end{aligned}$$

Damit gilt aber

$$\frac{\pi(b)}{1 - \pi(b)} \geq \frac{\pi(a)}{1 - \pi(a)};$$

d.h. $\pi(b) \geq \pi(a)$.

$\square$

Wir können jetzt zum Hauptsatz (von 4.2) schreiten (vgl. mit Satz 4.7):

Satz 4.9 [UMP-Test bei MLQ im Fall $\theta \leq \theta_0$ vs $\theta > \theta_0$] *Wenn MLQ gilt, existiert ein Test der Grösse α, welcher im Testproblem $\theta \leq \theta_0$ vs $\theta > \theta_0$ auf ganz $\theta > \theta_0$ die Macht $\pi(\theta)$ maximiert - eben:* **U***niformly* **M***ost* **P***owerful!*

Beweis von Satz 4.9 (in der Stunde - 2 Sachen sind zu beweisen):

$\square$

Zusammenfassend: Wir werden bei einseitigen Testproblemen immer den Test nach Satz 4.1 an der Grenze (θ_0) benutzen. Wir sind damit optimal (Satz 4.9). Dies gilt jedoch nur bei MLQ. Wir werden jetzt in 4.2.3 eine Familie von Verteilungen kennenlernen, wo die MLQ-Eigenschaft und vieles mehr immer kanonisch gegeben ist.

4.2.3 Die exponentielle Familie

Die folgenden Verteilungen:

* Bernoulli
* Binomial
* Geometrisch
* Negativ Binomial
* Poisson
* Exponential
* Gamma
* Normal

gehören alle zur exponentiellen Familie von Verteilungen (Cauchy= t_1 und Uniform nicht!), mit folgender

Definition 4.10 [exponentielle Familie von Verteilungen] *Eine Verteilung gehört zur (d-dimensionalen) exponentiellen Familie, wenn sich die Wahrscheinlichkeitsfunktion oder Dichte $p(x, \underline{\theta})$ darstellen lässt als*

$$p(x, \underline{\theta}) = c(\underline{\theta})h(x) \exp[\sum_{j=1}^{d} \theta_j t_j(x)].$$

Im Fall einer gemeinsamen Wahrscheinlichkeitsfunktion oder gemeinsamen Dichte erhalten wir bei Unabhängigkeit analog (z.B. bei einer n-Stichprobe)

$$p(\mathbf{x}, \underline{\theta}) = \left[c(\underline{\theta})\right]^n [\prod_{i=1}^{n} h(x_i)] \exp[\sum_{j=1}^{d} \theta_j \sum_{i=1}^{n} t_j(x_i)].$$

Wir nennen $(\theta_j)_{j=1}^{d}$ die natürlichen Parameter der Verteilung (nicht immer gleich den gewöhnlichen Parametern). Wir werden manchmal fordern müssen, dass der Parameterraum eine offene d-Kugel enthält.

Stellen Sie die Dichte einer n-Stichprobe aus einer Normalverteilung $\mathcal{N}(\mu, \sigma^2)$ in dieser Form dar (weitere Beispiele in den Übungen). Wie sind die natürlichen Parameter? Beobachten Sie das Auftreten der suffizienten Statistik für (μ, σ^2).

Untersuchen Sie die exponentielle Familie zuerst mit Hilfe des Faktorisierungskriteriums auf Suffizienz und danach noch auf minimale Suffizienz.

Untersuchen Sie die 1-dimensionale exponentielle Familie auf MLQ.

4.3 $\theta \in [\theta_1, \theta_2]$ vs $\theta \notin [\theta_1, \theta_2]$ (im Fall von MLQ; UMPU-Tests, Cox-Hinkley)

Das Hauptresultat aus diesem Teil (Satz 4.12) ist leider sehr umständlich: viele Voraussetzungen müssen erfüllt und überprüft werden und meist ist es nur mit Rechnern möglich, optimale Lösungen exakt zu berechnen. An Stelle von Satz 4.12 werden wir in der Praxis mit dem Verfahren von Cox-Hinkley zwar nicht optimale, jedoch meist ziemlich gute Tests einfach (und nachvollziehbar) entwickeln.

Wir setzen im Folgenden wieder MLQ voraus. Die Hypothesen lauten: $\mathcal{H}_0 : \theta \in [\theta_1, \theta_2]$ vs $\mathcal{H}_1 : \theta \notin [\theta_1, \theta_2]$. Wie ist die Wunschsituation, die Traumlösung?

Wir werden die Definition von Grösse aus 4.2 weiterhin brauchen können:

$$\sup_{\theta \in [\theta_1, \theta_2]} E_\theta[d(\mathbf{X})] = \sup_{\theta \in \mathcal{H}_0} E_\theta[d(\mathbf{X})] = \sup_{\theta \in \mathcal{H}_0} P_\theta[\text{sage "}\mathcal{H}_1\text{" mit Test } d].$$

Wir werden fordern, dass diese Grösse $\leq \alpha$ sein wird.

Die Traumlösung weiter oben ist nur in Situationen realistisch, bei denen es keineN StatistikerIN braucht. Können Sie wenigstens einen Test angeben, der immer existiert und Grösse $\leq \alpha$ hat?

Weitere Vorschläge für unser Problem?

Wir haben gesehen, dass wir sinnvollerweise die Menge der Tests einschränken sollten. Dazu definieren wir

Definition 4.11 [Unbiased Test (dt. unverfälscht)] *Ein Test d der Grösse α ist unbiased, wenn*

$$E_\theta[d(\mathbf{X})] \geq \alpha$$

für alle $\theta \in \mathcal{H}_1$.

Drei Bemerkungen: 1. Der Voll-Randomisierte Test ist unbiased. Somit existiert mindestens ein solcher Test.

2. In der Schätztheorie (Kapitel 5) heisst unbiased etwas ganz anderes. Es wird keine Verwechslungen geben.

3. **Das neue Ziel ist jetzt** also, unter den Unbiased-Tests einen most powerful zu finden - eben einen **U**niformly **M**ost **P**owerful **U**nbiased oder UMPU-Test!

Wir könnten unseren ersten Versuch leicht modifizieren in der Hoffnung, dadurch einen UMPU-Test zu erhalten und zwar mit dem sogenannten Ansatz von Cox-Hinkley:

Ist dieser Test von Grösse $\leq \alpha$?

Ist dieser Test unbiased?

Heuristischer Grund, weshalb im Allgemeinen für 4.3 kein UMP-Test existiert (ohne Forderung nach Unbiased)?

Wir listen hier Bedingungen auf, welche garantieren, dass wir einen UMPU nach Satz 4.12 konstruieren dürfen (stetig; diskret analog):

1. d messbar von der Form (t suffiziente Statistik) $d(t) = 1 \Leftrightarrow t \notin [t_1, t_2], t_1 < t_2 \in \mathbb{R}$, 0 sonst.
2. MLQ
3. für alle θ_0 gilt: $\sup_{\theta \in N(\theta_0)} \{g_{\theta_0\theta}(t_1) + g_{\theta_0\theta}(t_2)\} < \infty$, dabei bezeichne $N(\theta_0)$ eine beliebig kleine, offene Menge um θ_0
4. $p(x, \theta)$ (Dichte oder Wahrscheinlichkeitsfunktion) ist stetig in θ
5. für alle θ gilt: $g_{\theta\theta'}(t)$ ist stetig in θ' für alle $t \in [t_1, t_2]$
6. im Fall einer stetigen Verteilung muss die Dichte der suffizienten Statistik auf dem gesamten interessierenden Intervall positiv sein (keine Lücken)
7. $\frac{dg_{\theta\theta'}(t)}{dt}$ ist streng monoton wachsend für alle $\theta \neq \theta'$ (so etwas ist durchaus möglich; vgl. e^x und e^{-x})

Wir verwenden hier ohne System manchmal $d(t)$ und $d(\mathbf{X})$; obige Bedingungen müssen damit zum Beispiel sowohl für die Dichten der suffizienten Statistik wie auch für $\mathbf{X}$ gelten; entsprechend $g_{\theta\theta'}(t)$ und weitere!

Traurige Bemerkung: Eine Verteilung aus der exponentiellen Familie genügt nicht zwingend all diesen Bedingungen.

Damit kommen wir zur Formulierung des Hauptsatzes von 4.3:

Satz 4.12 [UMPU-Test] *Unter obigen Bedingungen existiert ein UMPU-Test der Hypothesen $\mathcal{H}_0 : \theta \in [\theta_1, \theta_2]$ vs $\mathcal{H}_1 : \theta \notin [\theta_1, \theta_2]$. Der Test ist zweiseitig von der Form (t die minimal suffiziente Statistik)*

$$d(t) := \begin{pmatrix} 1 \\ \gamma(t) \\ 0 \end{pmatrix} \Leftrightarrow t(\mathbf{x}) \begin{pmatrix} \notin [t_1, t_2] \\ \in \{t_1, t_2\} \\ \in (t_1, t_2) \end{pmatrix}.$$

Dabei werden $t_1, t_2, \gamma(t_1), \gamma(t_2)$ so gewählt, dass $E_{\theta_i}[d(\mathbf{X})] = \alpha$; $i \in \{1, 2\}$.

Bemerkung zu Satz 4.12 Wie wir gleich sehen werden, kann man mit Hilfe der beiden Gleichungen $E_{\theta_i}[d(\mathbf{X})] = \alpha$, $i \in \{1, 2\}$, die Werte $t_1, t_2, \gamma(t_1), \gamma(t_2)$ bestimmen.

Beispiel zu Satz 4.12 Sei $x_1, \ldots, x_n$ eine Stichprobe aus einer $\mathcal{N}(\theta, 1)$-Verteilung. $\mathcal{H}_0 : \theta \in [\theta_1, \theta_2]$ vs $\mathcal{H}_1 : \theta \notin [\theta_1, \theta_2]$. Die Bedingungen von Satz 4.12 sind erfüllt (freiwillige Hausaufgabe). Gemäss Satz 4.12 existiert ein UMPU-Test; er ist zweiseitig und wir müssen $t_1 < t_2$ finden, sodass $E_{\theta_i}[d(\mathbf{X})] = \alpha$; $i \in \{1, 2\}$. Die minimal suffiziente Statistik ist hier $\sum_{i=1}^n x_i$ und $\sum_{i=1}^n X_i$ ist $\mathcal{N}(n\theta, n)$-verteilt. Wir suchen also t_1, t_2, sodass

$$\frac{1}{\sqrt{2\pi n}} \int_{t_1}^{t_2} \exp\big[-\frac{1}{2n}(t - n\theta_i)^2\big]\, dt = 1 - \alpha;\ i \in \{1, 2\}.$$

Folgende Skizze

zeigt, dass wir aus Symmetriegründen eine Lösung der Art $P[\mathcal{N}(n\delta, n) \in (-\tau, +\tau)] = 1 - \alpha$ haben werden mit

1. $\overline{\theta} := (\theta_1 + \theta_2)/2$,

2. $t_{1,2} =: n\overline{\theta} \pm \tau$,

3. $\delta := \overline{\theta} - \theta_1$.

Die konkreten Werte für t_1, t_2 sind in einem konkreten Zahlenbeispiel in einer Aufgabe zu berechnen.

Wir untersuchen jetzt den **Fall, wo** $\theta_1 = \theta_2 =: \theta_0$. Wir verlieren in Satz 4.12 offenbar damit *eine Gleichung*, müssen aber trotzdem *zwei Werte* t_1, t_2 berechnen. Folgende Skizze (ohne Beweis; mehr in Ferguson (Math. Stats, a decision theoretic approach, Academic Press)) motiviert folgende neue Gleichung:

Wir haben dann also neben der einen Gleichung

$$E_{\theta_0}[d(\mathbf{X})] = \alpha, \tag{I}$$

welche uns von Satz 4.12 übrigbleibt, noch folgende neue Gleichung erhalten:

$$\frac{\partial}{\partial\theta} E_\theta[d(\mathbf{X})]\big|_{\theta=\theta_0} = 0. \tag{II}$$

Wir wollen dies gleich an einem einfachen Beispiel üben (Hausaufgabe: Voraussetzungen für Satz 4.12 sind erfüllt): sei x eine Einerstichprobe aus einer Exponentialverteilung mit Parameter θ. $\mathcal{H}_0 : \theta = 1$ vs $\mathcal{H}_1 : \theta \neq 1$. Wir müssen obige beiden Gleichungen (I) und (II) benutzen; wir rechnen in der Stunde (zuerst auf Blatt!):

Die konkreten Werte sind in einer geeigneten Rechenumgebung (weder R noch S-PLUS!) in einer Aufgabe zu bestimmen. Dort ist auch in diesem Beispiel der Vorschlag von Cox-Hinkley zu benutzen (je $\alpha/2$ links und rechts), um dieses Problem zu lösen. Die Lösung nach Cox-Hinkley wird nicht genau gleich sein wie diejenige unter Benutzung von (I) und (II). Damit ist der Vorschlag nach Cox-Hinkley entweder mit Bias oder nicht bester unter den unbiased Tests.

Wir haben in 4.3 wie angekündigt mit Satz 4.12 nur eine unbefriedigende Lösung unseres Ausgangsproblems erhalten. Die Lösungen sind meist nur mit Rechnern zu bestimmen und den KundInnen (BiologInnen, MedizinerInnen, ÖkonomInnen und SozialwissenschaftlerInnen) schwer zu verkaufen. Cox-Hinkley leuchtet eher ein und wird deshalb in der Praxis fast ausschliesslich benutzt. Im kommenden Teil 4.4 werden wir auch oft Cox-Hinkley einsetzen. Es sei noch ohne Beweis erwähnt, dass im Fall einer einfachen Nullhypothese und bei symmetrischer Dichte (zum Beispiel Normalverteilung), Cox-Hinkley auch UMPU ist!

Wir kommen jetzt noch zum längeren Beweis von Satz 4.12.

Beweis von Satz 4.12:

Wir bezeichnen die Bedingungen vor Satz 4.12 mit "(Bed)" und werden sie im Folgenden immer voraussetzen. Der Beweis wird nur für den schwierigeren Fall, nämlich stetig vorgeführt; diskret als freiwillige längere Schreibübung.

Schritt 1 **Wenn (Bed) erfüllt ist, so ist die Machtfunktion $\pi(\theta)$ stetig in θ:**

Sei T die suffiziente Statistik; Realisationen bezeichnen wir mit t. $f_\theta(t)$ sei die Dichte von T, wenn θ richtig ist. Wir prüfen die Stetigkeit in einem Punkt θ_0. Es gilt (wir untersuchen $(1-\pi(\theta))$ statt $\pi(\theta)$ aus theoretischen Gründen):

$$1-\pi(\theta) = 1 - E_\theta[d(T)] = \int_{[t_1,t_2]} f_\theta(t)dt = \int_{[t_1,t_2]} f_{\theta_0}(t) g_{\theta_0\theta}(t)dt \rightarrow$$

falls $\theta \rightarrow \theta_0$

$$\rightarrow \int_{[t_1,t_2]} f_{\theta_0}(t) g_{\theta_0\theta_0}(t)dt = \int_{[t_1,t_2]} f_{\theta_0}(t)dt = 1-\pi(\theta_0),$$

da $g_{\theta_0,\theta_0}(t) = 1$. Weshalb dürfen wir den Limesübergang machen:

$\square$

Wir untersuchen die 1-parametrige exp-Familie auf diejenigen Bedingungen in (Bed), welche wir bis jetzt benutzt haben.

Schritt 2 (kleines technisches Lemma KTL) **Seien f und g wachsende Funktionen und X eine Zufallsgrösse, sodass $E[f(X)] < \infty, E[g(X)] < \infty$ und $E[f(X)g(X)] < \infty$. Dann gilt:**

$$E[f(X)g(X)] \geq E[f(X)]E[g(X)].$$

(Damit hat man also insbesondere auch eine nichtnegative Korrelation!):

$$\begin{aligned} E[f(X)g(X)] - E[f(X)]E[g(X)] &= E[f(X)(g(X) - E[g(X)])] \\ &= E[(f(X) - c)(g(X) - E[g(X)])] \end{aligned}$$

für alle $c \in \mathbb{R}$. Wir wählen c wie in der Skizze unten angegeben; damit folgt die Behauptung.

$\square$

Schritt 3 (Existenz von t_1, t_2) **Sei** $0 < \alpha < 0.5$**. Wenn (Bed) erfüllt ist, dann existieren** (t_1, t_2) **(im diskreten Fall noch** (γ_1, γ_2)**), sodass**

$$P_{\theta_i}[T < t_1] + \gamma_1 P_{\theta_i}[T = t_1] + \gamma_2 P_{\theta_i}[T = t_2] + P_{\theta_i}[T > t_2] = \alpha$$

für $i \in \{1, 2\}$**.**:

Der Beweis ist sehr konstruktiv (und wird dann konkret auch zur Konstruktion der Lösung herangezogen, wenn man es numerisch lösen muss). Man konzentriert sich zuerst auf die Verteilung von T wenn θ_1 richtig ist und bildet ein Intervall $(-\infty, t_2^*]$ so, dass T (unter P_{θ_1}) mit Wahrscheinlichkeit $(1-\alpha)$ in $(-\infty, t_2^*]$ liegt. Nicht überraschend wird jedoch T kaum gleichzeitig auch unter P_{θ_2} mit Wahrscheinlichkeit $(1-\alpha)$ in $(-\infty, t_2^*]$ liegen.

Dann verschiebt man das Intervall langsam nach rechts, bis auch die 2. Gleichung stimmt. Es ist natürlich zu zeigen, dass das alles geht... - es geht!

Wir setzen erstmal ganz links und ganz rechts an und definieren damit unsere Intervalle

$$(-\infty, t_2^*], [t_1^*, \infty)$$

so, dass jeweils unter P_{θ_1} gilt:

$$P_{\theta_1}[T < t_2^*] = P_{\theta_1}[T > t_1^*] = 1 - \alpha.$$

Da $0 < \alpha < 0.5$, ist $t_1^* < t_2^*$ und anschaulich ist t_1^* der Punkt am weitesten rechts, der potentiell als linke Intervallgrenze des Annahmebereiches von $\mathcal{H}_0$ in Frage kommt. Für $t \in (-\infty, t_1^*]$ definieren wir eine Funktion $\Psi(t)$ so, dass

$$\int_t^{\Psi(t)} f_{\theta_1}(u)du = 1 - \alpha :$$

Unter P_{θ_1} haben wir unser Ziel schon mal erreicht. Wir wollen analog den Ausdruck unter P_{θ_2} analysieren und machen dazu einen Dichtewechsel:

$$\int_t^{\Psi(t)} f_{\theta_2}(u)du = \int_t^{\Psi(t)} g_{\theta_1\theta_2}(u) f_{\theta_1}(u)du.$$

Als nächstes zeigen wir, dass

$$\int_t^{\Psi(t)} f_{\theta_2}(u)du$$

wachsend ist in t:

Sei dazu $t < t'$ (und damit auch $t' < \Psi(t)$, siehe unten $(0 < \alpha < 0.5)$).

$$\int_{t'}^{\Psi(t')} f_{\theta_2}(u)du - \int_t^{\Psi(t)} f_{\theta_2}(u)du = \int_{\Psi(t)}^{\Psi(t')} f_{\theta_2}(u)du - \int_t^{t'} f_{\theta_2}(u)du,$$

wegen Dichtewechsel und MLQ gilt nun weiter

$$\begin{aligned}\int_{\Psi(t)}^{\Psi(t')} f_{\theta_2}(u)du - \int_t^{t'} f_{\theta_2}(u)du &\geq g_{\theta_1\theta_2}(\Psi(t)) \int_{\Psi(t)}^{\Psi(t')} f_{\theta_1}(u)du - g_{\theta_1\theta_2}(t') \int_t^{t'} f_{\theta_1}(u)du \\ &= R[g_{\theta_1\theta_2}(\Psi(t)) - g_{\theta_1\theta_2}(t')] \geq 0,\end{aligned}$$

wobei

$$R := \int_{\Psi(t)}^{\Psi(t')} f_{\theta_1}(u)du = \int_t^{t'} f_{\theta_1}(u)du = 1 - \alpha - \int_{t'}^{\Psi(t)} f_{\theta_1}(u)du.$$

Damit haben wir auch bewiesen:

$$\int_{-\infty}^{t_2^*} f_{\theta_2}(u)du \leq \int_t^{\Psi(t)} f_{\theta_2}(u)du \leq \int_{t_1^*}^{\infty} f_{\theta_2}(u)du$$

Die weitere Beweis-Strategie ist nun die folgende: Wir zeigen, dass (1) die linke Seite $\leq (1-\alpha)$ ist, (2) die rechte Seite $\geq (1-\alpha)$ und (3) dass der mittlere Ausdruck

$$h(t) := \int_t^{\Psi(t)} f_{\theta_2}(u)du$$

stetig in t ist. Dann folgt (Zwischenwertsatz für stetige Funktionen aus Analysis I), dass es auch ein t_1 geben muss, sodass $h(t_1) = (1-\alpha)$. Dann gilt auch

$$\int_{t_1}^{\Psi(t_1)} f_{\theta_1}(u)du = \int_{t_1}^{\Psi(t_1)} f_{\theta_2}(u)du = (1-\alpha),$$

womit wir die $(t_1, t_2) := (t_1, \Psi(t_1))$ gefunden haben - mit naheliegender Rechenstrategie für die Praxis.

Wir beweisen (2) und lassen (1) als analoge Übung: Mit Schritt 2 (KTL) folgt:

$$\begin{aligned}
h(t_1^*) &:= \int_{t_1^*}^{\infty} f_{\theta_2}(u)du \\
&= \int_{t_1^*}^{\infty} g_{\theta_1\theta_2}(u) f_{\theta_1}(u)du \\
&= E_{\theta_1}[g_{\theta_1\theta_2}(T) I[T \geq t_1^*]] \\
&\geq E_{\theta_1}[g_{\theta_1\theta_2}(T)] E_{\theta_1}[I[T \geq t_1^*]] \\
&= \int_{-\infty}^{\infty} g_{\theta_1\theta_2}(u) f_{\theta_1}(u)du P_{\theta_1}[T \geq t_1^*] \\
&= \int_{-\infty}^{\infty} f_{\theta_2}(u)du P_{\theta_1}[T \geq t_1^*] \\
&= 1(1-\alpha) = 1-\alpha
\end{aligned}$$

Es bleibt noch die Stetigkeit von $h(t)$: Wir zeigen Rechtsstetigkeit - Linksstetigkeit analog. Wir rechnen analog letzte Seite: Sei $t < t'$.

$$0 \leq \int_{t'}^{\Psi(t')} f_{\theta_2}(u)du - \int_{t}^{\Psi(t)} f_{\theta_2}(u)du = \int_{\Psi(t)}^{\Psi(t')} f_{\theta_2}(u)du - \int_{t}^{t'} f_{\theta_2}(u)du,$$

wegen Dichtewechsel und MLQ gilt nun weiter

$$\begin{aligned}
\int_{\Psi(t)}^{\Psi(t')} f_{\theta_2}(u)du - \int_{t}^{t'} f_{\theta_2}(u)du &\leq g_{\theta_1\theta_2}(\Psi(t')) \int_{\Psi(t)}^{\Psi(t')} f_{\theta_1}(u)du - g_{\theta_1\theta_2}(t) \int_{t}^{t'} f_{\theta_1}(u)du \\
&= R[g_{\theta_1\theta_2}(\Psi(t')) - g_{\theta_1\theta_2}(t)],
\end{aligned}$$

wobei

$$R := \int_{\Psi(t)}^{\Psi(t')} f_{\theta_1}(u)du = \int_{t}^{t'} f_{\theta_1}(u)du = 1 - \alpha - \int_{t'}^{\Psi(t)} f_{\theta_1}(u)du.$$

Wegen des Satzes der majorisierten Konvergenz von Lebesgue konvergiert aber $R = R(t')$ gegen 0, wenn $t' \to t$.

$\square$

Schritt 4 (kleines technisches Lemma KTL II) **Sei (Bed) erfüllt (v.a. Bed 7), sei** $t_1 < t_2$ **und** $\theta_a < \theta_b < \theta_c$**, dann gibt es** $k_1, k_2 > 0$**, sodass**

$$f_{\theta_b}(t) \begin{pmatrix} > \\ = \\ < \end{pmatrix} k_1 f_{\theta_a}(t) + k_2 f_{\theta_c}(t) \Longleftrightarrow t \in \begin{pmatrix} (t_1, t_2) \\ \{t_1, t_2\} \\ [t_1, t_2]^c \end{pmatrix}. \tag{KTL II}$$

Wir formulieren (KTL II) um als

$$1 \begin{pmatrix} > \\ = \\ < \end{pmatrix} k_1 g_{\theta_b \theta_a}(t) + k_2 g_{\theta_b \theta_c}(t) \Longleftrightarrow t \in \begin{pmatrix} (t_1, t_2) \\ \{t_1, t_2\} \\ [t_1, t_2]^c \end{pmatrix}.$$

Wenn $k_1, k_2 > 0$, so ist $k_1 g_{\theta_b \theta_a}(t) + k_2 g_{\theta_b \theta_c}(t)$ streng konvex. Damit hat

$$1 = k_1 g_{\theta_b \theta_a}(t) + k_2 g_{\theta_b \theta_c}(t)$$

aber entweder 0, 1, oder 2 Lösungen. Wir müssen jetzt nur k_1, k_2 so finden, dass diese Gleichung 2 Lösungen t_1, t_2 hat. Diese kleine Aufgabe (2 Gleichungen, 2 Unbekannte) bitte als Hausaufgabe fertig lösen; warum sind in der Lösung automatisch $k_1, k_2 > 0$?

$\square$

Schritt 5 (Verallgemeinertes Neyman-Pearson-Lemma) **Sei (Bed) erfüllt. Sei $d^*(t)$ der Form**

$$d^*(t) = \begin{pmatrix} 1 \\ \gamma(t) \\ 0 \end{pmatrix} \Leftrightarrow f_\theta(t) \begin{pmatrix} > \\ = \\ < \end{pmatrix} k_1 f_1(t) + k_2 f_2(t),$$

wobei $k_1, k_2 \in \mathbb{R}$. Dann maximiert $d^*(t)$ die Macht $E_\theta[d(t)]$ unter allen Testfunktionen d, die die Gleichungen (steht jeweils für Risiken 1. Art):

$$\int d(t) f_1(t) dt = \int d^*(t) f_1(t) dt, \int d(t) f_2(t) dt = \int d^*(t) f_2(t) dt \tag{A}$$

erfüllen. Wenn $k_1, k_2 \geq 0$, maximiert d^* sogar unter allen d, sodass

$$\int d(t) f_1(t) dt \leq \int d^*(t) f_1(t) dt, \int d(t) f_2(t) dt \leq \int d^*(t) f_2(t) dt. \tag{B}$$

Wie im Beweis des Lemmas von Neyman Pearson (Satz 4.1) haben wir analog:

$$\int \big((d^*(t) - d(t)\big)(f_\theta(t) - k_1 f_1(t) - k_2 f_2(t)) dt \geq 0.$$

Ausmultipliziert ergibt sich auf der linken Seite (T sei die min suff Stat mit Dichte $f_\theta(t)$):

$$\begin{aligned} E_\theta[d^*(T)] - E_\theta[d(T)] - k_1 &\left[\int d^*(t) f_1(t) dt - \int d(t) f_1(t) dt\right] \\ - k_2 &\left[\int d^*(t) f_2(t) dt - \int d(t) f_2(t) dt\right] \end{aligned}$$

Wenn d (A) erfüllt, haben wir (egal wie k_1, k_2) $E_\theta[d^*(T)] \geq E_\theta[d(T)]$. Wenn (B) gilt, haben wir $E_\theta[d^*(T)] \geq E_\theta[d(T)]$.

□

Man beachte schon mal vorbereitend, dass in Schritt 4 und Schritt 5 kombiniert die Alternativ-Hypothese dann angenommen wird, wenn die suffiziente Statistik $t \in [t_1, t_2]$ ist und nicht umgekehrt; es gibt also vorübergehend eine "Umstülpung" der Hypothesen - und Tests!

Schritt 6 ("Umgestülpte" Fragestellung) **Sei (Bed) erfüllt. Sei $\mathcal{H}_0 : \theta \in [\theta_1, \theta_2]^c$ und $\mathcal{H}_1 : \theta \in [\theta_1, \theta_2]$. Dann gibt es einen UMP-Test der Grösse α und zwar der Form:**

$$d(t) = \begin{pmatrix} 1 \\ \gamma(t) \\ 0 \end{pmatrix} \Leftrightarrow t \begin{pmatrix} \in (t_1, t_2) \\ \{t_1, t_2\} \\ [t_1, t_2]^c \end{pmatrix}, \tag{+}$$

wobei t_1, t_2 (und allenfalls $\gamma(t_1), \gamma(t_2)$) so gewählt werden, dass

$$E_{\theta_1}[d(T)] = E_{\theta_2}[d(T)] = \alpha. \tag{$*$}$$

Wir prüfen erstmals, ob der Test maximale Macht auf $\mathcal{H}_1$ hat. Dazu stellen wir erstmals fest (kleine Hausaufgabe), dass wir in Schritt 3 gar nicht voraussetzen mussten, dass $\alpha < 0.5$ - es hat nur die Anschauung erleichtert. Bitte als kleine Hausaufgabe überprüfen.

Nach Schritt 3 gibt es (t_1, t_2) so, dass für Test (+) die Bedingungen ($*$) erfüllt sind. Von Schritt 4 (KTL II) sehen wir, dass wir $k_1, k_2 > 0$ finden, sodass (+) äquivalent ist zu

$$f_\theta(t) \begin{pmatrix} > \\ = \\ < \end{pmatrix} k_1 f_{\theta_1}(t) + k_2 f_{\theta_2}(t),$$

mit $\theta_a = \theta_1, \theta_b = \theta, \theta_c = \theta_2; \theta_a < \theta_b < \theta_c$. Von Schritt 5 haben wir damit

$$E_\theta[d(T)] \geq E_\theta[d'(T)]$$

für alle alternativen Testfunktionen d', sodass $E_{\theta_i}[d'(T)] \leq \alpha$ wo $i \in \{1, 2\}$.

Wir prüfen noch, ob unser in (+) **definiertes d Grösse α hat.** Wir zeigen nur den Fall $\theta < \theta_1 (< \theta_2)$; der andere Fall analog als Hausaufgabe. Sei also $\theta < \theta_1$ - es bleibt zu zeigen: $E_\theta[d(T)] \leq \alpha$: Von Schritt 4 (KTL II) mit $\theta_a = \theta, \theta_b = \theta_1, \theta_c = \theta_2$; es existieren $k_1', k_2' > 0$, sodass

$$f_{\theta_1}(t) \begin{pmatrix} > \\ = \\ < \end{pmatrix} k_1' f_\theta(t) + k_2' f_{\theta_2}(t)$$

oder anders

$$f_\theta(t) \begin{pmatrix} < \\ = \\ > \end{pmatrix} \frac{1}{k_1'} f_{\theta_1}(t) - \frac{k_2'}{k_1'} f_{\theta_2}(t).$$

Mit Schritt 5, Fall (A) und Umkehrung (Minimieren statt Maximieren) minimiert unser in (+) definiertes d die Macht $E_\theta[d(T)]$ unter allen Tests, welche Bedingung (A) in Schritt 5 erfüllen. Aber der Voll-Randomisierte Test erfüllt alle diese Bedingungen. Also gibt es mindestens einen solchen Test und zwar der Grösse α. Also hat unser in (+) definiertes d auch maximal eine Grösse α.

Nebenbemerkung I: Der Test ist übrigens auch unverfälscht; wegen Vergleich mit Voll-Randomisiertem Test.

Nebenbemerkung II: Man kann den Beweis von Satz 4.12 auch so führen, dass man auf Schritt 6 verzichtet - es ist aber auch ein schönes Resultat für sich.

□

Schritt 7 (Abschliessende Beweisschritte für Satz 4.12) Wegen **Schritt 1** wissen wir, dass ein UMPU erfüllen muss, dass $E_{\theta_i}[d(T)] = \alpha$, $i \in \{1, 2\}$.

Wegen **Schritt 3** wissen wir, dass wir (t_1, t_2) finden können, sodass

$$d(t) := \begin{pmatrix} 1 \\ \gamma(t) \\ 0 \end{pmatrix} \Leftrightarrow t \begin{pmatrix} \notin [t_1, t_2] \\ \in \{t_1, t_2\} \\ \in (t_1, t_2) \end{pmatrix}.$$

wobei $E_{\theta_i}[d(T)] = \alpha$; $i \in \{1, 2\}$ erfüllt wird.

Wegen **Schritt 4** können wir von den t_1, t_2 eine äquivalente Form der Art

$$f_{\theta_b}(t) \begin{pmatrix} > \\ = \\ < \end{pmatrix} k_1 f_{\theta_a}(t) + k_2 f_{\theta_c}(t) \Longleftrightarrow t \in \begin{pmatrix} (t_1, t_2) \\ \{t_1, t_2\} \\ [t_1, t_2]^c \end{pmatrix} \tag{D1}$$

finden ($\theta_a < \theta_b < \theta_c$, $k_1, k_2 > 0$); dies kann analog als

$$f_{\theta_a}(t) \begin{pmatrix} < \\ = \\ > \end{pmatrix} \frac{1}{k_1} f_{\theta_b}(t) - \frac{k_2}{k_1} f_{\theta_c}(t) \Longleftrightarrow t \in \begin{pmatrix} (t_1, t_2) \\ \{t_1, t_2\} \\ [t_1, t_2]^c \end{pmatrix} \tag{D2}$$

dargestellt werden.

Wir **wollen schauen, ob $d(t)$, der ursprüngliche Test, den wir von Schritt 3 haben, auf $\theta < \theta_1 (< \theta_2)$ die Macht maximiert** unter allen Tests, die $E_{\theta_i}[d(T)] = \alpha$,

$i \in \{1, 2\}$, erfüllen; $\theta > \theta_2$ analog. Wir definieren jetzt $\phi^*(t) := 1 - d(t)$. Wegen Schritt 5 und (D2) minimiert ϕ^* die Macht $E_\theta[\phi(T)]$; das heisst aber, dass $d(t)$ die Macht $E_\theta[\phi(T)]$ maximiert.

Wir **müssen noch prüfen, ob** $d(t)$ **Grösse** α **hat.** Sei dazu $\theta_1 < \theta < \theta_2$; wir benutzen jetzt Schritt 5 und (D1) und können folgern, dass ϕ^* die Macht maximiert, also $d(t)$ die Macht minimiert. Durch Vergleich mit dem Voll-Randomisierten Test folgt, dass der Test somit auf $\mathcal{H}_0$ Macht $\leq \alpha$ haben muss; mithin also Grösse α.

□

4.4 Invarianz ($Var[X] = \sigma_0^2$?, t-Test, F-Test, 2 Stichproben t-Test)

Erfahrungsgemäss haben StudentInnen beim ersten Durchgehen von 4.4 Mühe mit 4.4.1. Wenn man dann die einleuchtenden Anwendungen in 4.4.2, 4.4.3, 4.4.4 und 4.4.5 durchgearbeitet hat, ist man für die Grundlagen aus 4.4.1 eher empfänglich und sollte diese nochmals durcharbeiten. 4.4.1 ist einfach mathematisch genau begründet, was PraktikerInnen (IngenieurInnen, SozialwissenschaftlerInnen, NaturwissenschaftlerInnen) mit gesundem Menschenverstand in 4.4.2 und folgende automatisch richtig machen.

4.4.1 Grundidee Invarianz

Die folgende Theorie wird am Beispiel der Normalverteilung illustriert, weil dort die wichtigsten Anwendungen sind und die Transformationen der Daten problemlos sind. Wenn wir eine Stichprobe $x_1, x_2, \ldots, x_n$ aus einer $\mathcal{N}(\mu, \sigma^2)$-Verteilung haben

* und uns ausschliesslich für die Varianz interessieren, ist es sicher erlaubt, *vor jeder weiteren Untersuchung* die gleiche konstante Zahl von jeder einzelnen Zahl abzuziehen und mit den transformierten Daten fortzufahren, oder:
* μ schätzen wollen, so ist es sicher erlaubt, *vor jeder weiteren Untersuchung* die gleiche konstante Zahl von jeder einzelnen Zahl abzuziehen und mit den transformierten Daten das arithmetische Mittel zu bestimmen *und am Schluss diese eine konstante Zahl wieder zu addieren*, oder:
* allgemein: Manipulationen an den Daten durchführen, wenn wir dies bei der Auswertung *richtig* berücksichtigen.

Die letzte Formulierung ("richtig" berücksichtigen) werden wir jetzt formalisieren:

Wir werden 2 Transformationen der Daten genauer anschauen: Verschiebung der Daten um b (Lage) und Streckung mit Faktor $c > 0$ (Skala). Es fällt auf, dass wenn X eine $\mathcal{N}(\mu, \sigma^2)$-Zufallsgrösse ist, dann hat mit

$$g_{bc}(x_i) := b + cx_i, \ i \in \{1, \ldots, n\}, \tag{4.4}$$

die Zufallsgrösse $g(X)$ eine $\mathcal{N}(b+c\mu, c^2\sigma^2)$-Verteilung. Wir haben also insbesondere wieder eine Normalverteilung! Damit haben wir folgende Definition genügend motiviert:

Definition 4.13 [Invarianz einer Familie von Verteilungen bzgl. Gruppe von Transformationen] *Die Familie $(P_\theta)_{\theta \in \mathcal{H}}$ ist invariant bezüglich einer Gruppe von Transformationen G, wenn für alle $g \in G$ und für alle $\theta \in \mathcal{H}$ gilt: wenn X Verteilung P_θ besitzt, dann existiert ein $\theta' \in \mathcal{H}$, sodass $g(X)$ Verteilung $P_{\theta'}$ besitzt. Die dadurch induzierte Abbildung nennen wir ϕ_g: $\phi_g(\theta) := \theta'$.*

Wie ist bei (4.4) die Abbildung ϕ?

Welche Gruppe von Transformationen (innerhalb von "g_{bc}") ist noch zugelassen, wenn die Familie von Verteilungen aus $\mathcal{N}(\mu, 1)$, $\mu \in \mathbb{R}$, besteht, damit wir Invarianz haben?

Wenn wir eine Familie $(P_\theta)_{\theta \in \mathcal{H}}$ haben, welche invariant bezüglich einer Gruppe von Transformationen G ist, so wollen wir auch noch fordern, dass die Verlustfunktion invariant ist:

Definition 4.14 [Invarianz der Verlustfunktion] *Wenn die Familie $(P_\theta)_{\theta \in \mathcal{H}}$ invariant bezüglich einer Gruppe von Transformationen G ist, dann ist eine Verlustfunktion $L(\theta, a)$ (a die gewählte Aktion) invariant bezüglich G, wenn bei gegebenen $g \in G$, $a \in \mathcal{A}$ beliebig, existiert ein $a' \in \mathcal{A}$, sodass*

$$L(\theta, a) = L(\phi_g(\theta), a')$$

für alle $\theta \in \mathcal{H}$. Die dadurch induzierte Abbildung nennen wir ψ_g: $\psi_g(a) := a'$.

In dieser Definition spiegelt sich die Überlegung, dass wenn wir bei einer Schätzung die Daten mit g transformieren und einen neuen Parameter θ' erhalten, sollten wir vielleicht auch eine andere Aktion a' wählen (vgl. Beispiel 4.16).

Wir kommen zur entscheidenden Definition

Definition 4.15 [Invarianz der Entscheidungsfunktion] *Die Familie* $(P_\theta)_{\theta \in \mathcal{H}}$ *sei invariant bezüglich einer Gruppe von Transformationen* G*; dann ist eine Entscheidungsfunktion* $d : \mathbb{R}^n \to \mathcal{A}$ *invariant bezüglich* G*, wenn*

$$L(\theta, d(\mathbf{x})) = L(\phi_g(\theta), d(g(\mathbf{x})))$$

für alle $g, \mathbf{x}, \theta$ *oder kurz:*

$$d(g(\mathbf{x})) = \psi_g(d(\mathbf{x})).$$

Wir können dies graphisch veranschaulichen:

Diese Serie von Definitionen sollte anhand eines **Schätzproblems** illustriert werden:

Beispiel 4.16

Wir kehren zurück zu **Testproblemen**. Wir wählen als Verlustfunktion $L(\mathcal{H}_0, \mathcal{H}_0) := 0, L(\mathcal{H}_0, \mathcal{H}_1) := 1, L(\mathcal{H}_1, \mathcal{H}_0) := 1, L(\mathcal{H}_1, \mathcal{H}_1) := 0$. Wir fordern von unseren Transformationen (das dürfen wir - wir schränken uns einfach in unserer Freiheit ein)

$$\phi_g(\mathcal{H}_0) \subseteq \mathcal{H}_0,\ \phi_g(\mathcal{H}_1) \subseteq \mathcal{H}_1. \tag{4.5}$$

Sei $\mathbf{x}$ so, dass $d(\mathbf{x}) = a_0$ (nehme $\mathcal{H}_0$ an mit Daten $\mathbf{x}$) und $\theta \in \mathcal{H}_0$. Damit d invariant ist, muss gelten:

$$L(\theta, d(\mathbf{x})) = 0 = L(\phi_g(\theta), d(g(\mathbf{x})))\ \forall g \in G.$$

Wegen (4.5) geht dies aber nur, wenn auch $d(g(\mathbf{x})) = a_0$ (analog mit $L(\theta, a_1)$, wo $\theta \in \mathcal{H}_1$). Damit ist aber $\psi = \mathrm{id}$ oder anders

$$d(g(x)) = d(x). \tag{4.6}$$

Unser Vorgehen wird also folgendermassen aussehen:

1. Daten mit Suffizienz reduzieren
2. Daten mit Invarianz eventuell weiter reduzieren
3. Innerhalb dieser Möglichkeiten gute (optimale) Lösung suchen

Mathematische Statistik, u.a. Suche nach Komplexitätsreduktion: Suffizienz, Invarianz; bzw Optimalität: Neyman Pearson, UMP(U), Cramer-Rao-Schranke in Kapitel 5.

Wir wollen dies jetzt in 4.4.2 und folgende anwenden (**wieder einsteigen falls ausgestiegen**).

4.4.2 Ist die Varianz bei einer Normalverteilung gleich einem vorgegebenen σ_0^2?

Angenommen, wir haben eine Stichprobe $x_1, \ldots, x_n$, von der wir wissen, dass sie aus einer Normalverteilung stammt. Wir kennen leider weder Erwartungswert noch Varianz. Jemand hat nun behauptet, die Varianz sei 1. Wie könnten wir das testen ($\mathcal{H}_0 : (\mu, \sigma^2) : \sigma^2 = 1, \mathcal{H}_1 : (\mu, \sigma^2) : \sigma^2 \neq 1$)? Wir wissen, dass mit der Transformation von (4.4):

$$g_{bc}(x_i) := b + cx_i,\ i \in \{1, \ldots, n\}$$

die Klasse der Normalverteilungen nicht verlassen wird. Diese Familie ist invariant unter dieser Gruppe. Wie steht es mit der Abbildung ϕ? Wir haben diese auch bereits untersucht:

$$\phi(\mu, \sigma^2) = (b + c\mu, c^2\sigma^2).$$

Damit wir (4.5) genügen, muss $c = 1$ sein. Wir können also Lageveränderungen durchführen, was uns bei dieser Fragestellung nicht überrascht. Die Suffizienz sagt uns, dass sinnvollerweise d eine Funktion von

$$(\sum_{i=1}^{n} x_i, \sum_{i=1}^{n} x_i^2)$$

ist, also ein

$$d(\sum_{i=1}^{n} x_i, \sum_{i=1}^{n} x_i^2).$$

Wir haben noch keine Ahnung, wie d konkret aussehen wird! Wegen (4.6) können wir jetzt aber gerade so gut für noch frei zu wählendes b eine Funktion von

$$(\sum_{i=1}^{n}(x_i + b), \sum_{i=1}^{n}(x_i + b)^2)$$

untersuchen. Dies machen wir unter geschickter Wahl von b; nämlich: wähle

$$b := -\frac{1}{n}\sum_{i=1}^{n} x_i \quad (= -\overline{x}).$$

Damit haben wir also unser Problem reduziert auf den Ausdruck

$$d(0, \sum_{i=1}^{n}(x_i - \overline{x})^2). \tag{4.7}$$

Was wir jetzt mit Ausdruck (4.7) anstellen, ist noch nicht bestimmt. Unter $\mathcal{H}_0$ hat $\sum_{i=1}^{n}(X_i - \overline{X})^2$ eine χ^2_{n-1}-Verteilung (vgl. 1.4.2.6). Man nimmt dann üblicherweise mit dem Ansatz von Cox- und Hinkley je links und rechts ein $\alpha/2$ weg und lehnt dort aussen $\mathcal{H}_0$ ab:

Wir haben hier den Fall "$\sigma^2 = 1$?" untersucht. Was erhalten wir im allgemeinen Fall?

Dieser Test ist nicht mit dem χ^2-Test aus 4.5 zu verwechseln - unter χ^2-Test versteht man normalerweise den χ^2-Anpassungstest aus 4.5!

Further Readings: Stahel: 8.8 k (Test nicht robust!)

4.4.3 t-Test (Ist der Mittelwert gleich einem vorgegebenen μ_0 (bei unbekannter Varianz, Normalverteilung)?)

Angenommen, wir haben eine Stichprobe $x_1, \ldots, x_n$, von der wir wissen, dass sie aus einer Normalverteilung stammt. Wir kennen leider weder den Erwartungswert noch die Varianz (das ist übrigens der Normalfall). Jemand hat behauptet, der Erwartungswert sei 0 ($\mathcal{H}_0 : (\mu, \sigma^2) : \mu = 0, \mathcal{H}_1 : (\mu, \sigma^2) : \mu \neq 0$, zweiseitig). Wie könnten wir das testen? Wieder gilt, dass mit der Transformation von (4.4)

$$g_{bc}(x_i) := b + cx_i,\ i \in \{1, \ldots, n\} :$$

die Familie der Normalverteilungen invariant ist. ϕ ist unverändert

$$\phi(\mu, \sigma^2) = (b + c\mu, c^2\sigma^2).$$

Damit wir (4.5) genügen, muss $b = 0$ sein. Wir können also Skalenveränderungen durchführen, was uns bei dieser Fragestellung nicht überrascht. Die Suffizienz sagt uns, dass sinnvollerweise d eine Funktion von

$$(\sum_{i=1}^{n} x_i, \sum_{i=1}^{n} x_i^2)$$

ist, also ein

$$d(\sum_{i=1}^{n} x_i, \sum_{i=1}^{n} x_i^2).$$

Wir haben noch keine Ahnung, wie d konkret aussehen wird! Wegen (4.6) können wir jetzt aber gerade so gut für noch frei zu wählendes c eine Funktion von

$$(c\sum_{i=1}^{n} x_i, c^2\sum_{i=1}^{n} x_i^2)$$

untersuchen. Dies machen wir unter geschickter Wahl von c; nämlich: wähle

$$c := \frac{1}{\sqrt{\sum_{i=1}^{n} x_i^2}}.$$

Damit haben wir also unser Problem reduziert auf den Ausdruck

$$d\left(\frac{\sum_{i=1}^{n} x_i}{\sqrt{\sum_{i=1}^{n} x_i^2}}, 1\right). \tag{4.8}$$

Was wir jetzt mit Ausdruck (4.8) anstellen, ist noch nicht bestimmt. Die "1" ist nicht informativ und der Ausdruck

$$\frac{\sum_{i=1}^n x_i}{\sqrt{\sum_{i=1}^n x_i^2}}$$

kann folgendermassen umgeformt werden:

$$\frac{\sum_{i=1}^n x_i}{\sqrt{\sum_{i=1}^n x_i^2}} = \frac{\sum_{i=1}^n x_i}{\sqrt{\sum_{i=1}^n (x_i - \overline{x})^2 + n\overline{x}^2}} = \frac{\text{sign}(\sum_{i=1}^n x_i)}{\sqrt{\frac{1}{[t_{n-1}(\mathbf{x})]^2} + 1/n}},$$

wo

$$t_{n-1}(\mathbf{x}) := \frac{\sum_{i=1}^n x_i}{\sqrt{\sum_{i=1}^n (x_i - \overline{x})^2}}. \tag{4.9}$$

Also können wir gerade so gut $|t_{n-1}(\mathbf{x})|$ untersuchen. Unter $\mathcal{H}_0$ hat das um 0 symmetrische

$$\frac{\frac{1}{\sqrt{n}} \sum_{i=1}^n X_i}{\sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \overline{X})^2}} = \frac{\overline{X}}{\sqrt{\frac{1}{n(n-1)} \sum_{i=1}^n (X_i - \overline{X})^2}}$$

eine t_{n-1}-Verteilung (vgl. 1.4.2.8) - diese ist tabelliert und damit werden wir also

$$\frac{\frac{1}{\sqrt{n}} \sum_{i=1}^n x_i}{\sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \overline{x})^2}} = \frac{\overline{x}}{\sqrt{\frac{1}{n(n-1)} \sum_{i=1}^n (x_i - \overline{x})^2}}$$

untersuchen/berechnen. Man nimmt dann üblicherweise mit dem Ansatz von Cox- und Hinkley je links und rechts ein $\alpha/2$ weg und lehnt dort aussen $\mathcal{H}_0$ ab:

Wir haben hier den Fall "$\mu = 0$?" untersucht. Was erhalten wir im allgemeinen Fall? Interpretation der Formel?

Im Beispiel von 4.1.2 (Mittel zur Blutdrucksenkung) haben wir einfach angenommen, die Varianz sei 64 *und dies sei erst noch bekannt!*. In der Praxis hat man aber normalerweise die Situation von 4.4.3, wo wir die Standardabweichung schätzen müssen und deshalb im Nenner noch mit diesem Ausdruck normieren müssen. Weil hier eben auch der Zufall drin ist, ist die t-Verteilung langschwänziger als die Normalverteilung (konvergiert aber mit df $\rightarrow \infty$ gegen die Normalverteilung).

Further Readings: Stahel: 8.5 / Lam: 7.4.1 / Dalgaard: 4.1

4.4.4 F-Test (Sind die Varianzen von zwei unabhängigen Stichproben gleich (Normalverteilung)?)

Angenommen, wir haben zwei unabhängige Stichproben $x_1, \ldots, x_m$ und $y_1, \ldots, y_n$, von denen wir wissen, dass sie je aus einer Normalverteilung stammen. Wir kennen leider weder die beiden Erwartungswerte noch die Varianzen. Jemand hat behauptet, die Varianzen von X und Y seien gleich ($\mathcal{H}_0 : \sigma_X^2 = \sigma_Y^2, \mathcal{H}_1 : \sigma_X^2 \neq \sigma_Y^2$). Wie könnten wir das testen? Die Familie von Verteilungen, welche wir untersuchen, besteht aus Zufallsvektoren der Länge $m+n$; die ersten m Elemente haben eine $\mathcal{N}(\mu_1, \sigma_X^2)$-Verteilung, die letzten n Elemente haben eine $\mathcal{N}(\mu_2, \sigma_Y^2)$-Verteilung. Mit Transformationen der Art

$$g_{b_1 b_2 c_1 c_2}(\mathbf{x}, \mathbf{y}) := (b_1 + c_1\mathbf{x}, b_2 + c_2\mathbf{y})$$

wird diese Familie nicht verlassen. Wie steht es mit der Abbildung ϕ?

$$\phi(\mu_1, \mu_2, \sigma_X^2, \sigma_Y^2) = (b_1 + c_1\mu_1, b_2 + c_2\mu_2, c_1^2\sigma_X^2, c_2^2\sigma_Y^2).$$

Damit wir (4.5) genügen, muss $c_1 = c_2$ sein. Wir können also in X und Y unterschiedliche Lageveränderungen durchführen und gleiche Skalenänderungen, was uns bei dieser Fragestellung nicht überrascht. Die Suffizienz sagt uns, dass sinnvollerweise d eine Funktion von

$$(\sum_{i=1}^m x_i, \sum_{i=1}^m x_i^2, \sum_{i=1}^n y_i, \sum_{i=1}^n y_i^2)$$

ist, also ein

$$d(\sum_{i=1}^m x_i, \sum_{i=1}^m x_i^2, \sum_{i=1}^n y_i, \sum_{i=1}^n y_i^2).$$

Wir haben noch keine Ahnung, wie d konkret aussehen wird! Wegen (4.6) können wir jetzt aber gerade so gut für noch frei zu wählende $b_1, b_2, c := c_1 = c_2$ eine Funktion von

$$(\sum_{i=1}^m (cx_i + b_1), \sum_{i=1}^m (cx_i + b_1)^2, \sum_{i=1}^n (cy_i + b_2), \sum_{i=1}^n (cy_i + b_2)^2)$$

oder geschickter

$$(mb_1 + c\sum_{i=1}^m x_i, \sum_{i=1}^m (cx_i + b_1)^2, nb_2 + c\sum_{i=1}^n y_i, \sum_{i=1}^n (cy_i + b_2)^2)$$

untersuchen. Dies machen wir unter geschickter Wahl von b_1, b_2, c; nämlich: wähle

$$b_1 := -c\overline{x} \quad b_2 := -c\overline{y}.$$

Damit haben wir noch eine Funktion

$$d(0, c^2 \sum_{i=1}^{m} (x_i - \overline{x})^2, 0, c^2 \sum_{i=1}^{n} (y_i - \overline{y})^2).$$

Jetzt nehmen wir da $\sum (y_i - \overline{y})^2 > 0$ (gilt mit W'keit 1!)

$$c^2 := \frac{1}{\sum_{i=1}^{n} (y_i - \overline{y})^2}$$

und erhalten

$$d(0, \frac{\sum_{i=1}^{m} (x_i - \overline{x})^2}{\sum_{i=1}^{n} (y_i - \overline{y})^2}, 0, 1). \tag{4.10}$$

Was wir jetzt mit Ausdruck (4.10) anstellen, ist noch nicht bestimmt. Tabelliert (bzw. in Rechenlibraries abgelegt) ist

$$\frac{\sum_{i=1}^{m} (x_i - \overline{x})^2/(m-1)}{\sum_{i=1}^{n} (y_i - \overline{y})^2/(n-1)}. \tag{4.11}$$

(4.11) hat (als Zufallsgrösse) unter $\mathcal{H}_0$ eine $F_{m-1,n-1}$-Verteilung (vgl. 1.4.2.7). Man nimmt dann üblicherweise mit dem Ansatz von Cox- und Hinkley je links und rechts ein $\alpha/2$ weg und lehnt dort aussen $\mathcal{H}_0$ ab:

Solche Probleme treten dort auf, wo ein Test bei 2 Stichproben die gleiche Varianz *voraussetzt* (sehr häufig). Wir können dann zuerst mit 4.4.4 testen, ob die Varianzen gleich sind und danach mit dem eigentlichen Test beginnen. Es gibt dann aber das gewaltige (und unbefriedigend gelöste) Problem, dass man dann zwei mal mit einem gewissen Signifikanzniveau einen Test macht. Wie vertrauenswürdig insgesamt die Entscheidung ist (für oder gegen $\mathcal{H}_0$), ist schwer einzuschätzen.

Further Readings: Stahel: 8.8 l (Test nicht robust!) / Dalgaard: 4.4

4.4.5 Zwei-Stichproben t-Test

Angenommen, wir haben zwei unabhängige Stichproben $x_1, \ldots, x_m$ und $y_1, \ldots, y_n$, von denen wir wissen, dass sie je aus einer Normalverteilung stammen. Wir kennen leider weder die beiden Erwartungswerte noch die Varianzen, wissen aber, dass die Varianzen gleich sind. Jemand hat behauptet, die Erwartungswerte von X und Y seien gleich ($\mathcal{H}_0 : \mu_1 = \mu_2, \mathcal{H}_1 : \mu_1 \neq \mu_2$). Wie könnten wir das testen? Die Familie von Verteilungen, welche wir untersuchen, besteht aus Zufallsvektoren der Länge $m+n$; die ersten m Elemente haben eine $\mathcal{N}(\mu_1, \sigma^2)$-Verteilung, die letzten n Elemente haben eine $\mathcal{N}(\mu_2, \sigma^2)$-Verteilung. Mit Transformationen der Art

$$g_{b_1 b_2 c}(\mathbf{x}, \mathbf{y}) := (b_1 + c\mathbf{x}, b_2 + c\mathbf{y})$$

wird diese Familie nicht verlassen. Wie steht es mit der Abbildung ϕ?

$$\phi(\mu_1, \mu_2, \sigma^2) = (b_1 + c\mu_1, b_2 + c\mu_2, c^2\sigma^2).$$

Damit wir (4.5) genügen, muss $b_1 = b_2$ sein. Wir können also in X und Y gleiche Lageveränderungen durchführen und gleiche Skalenänderungen, was uns bei dieser Fragestellung nicht überrascht. Die Suffizienz sagt uns, dass sinnvollerweise d eine Funktion von

$$(\sum_{i=1}^{m} x_i, \sum_{i=1}^{n} y_i, \sum_{i=1}^{m} x_i^2 + \sum_{i=1}^{n} y_i^2)$$

ist (kleine, freiwillige Hausaufgabe), also ein

$$d(\sum_{i=1}^{m} x_i, \sum_{i=1}^{n} y_i, \sum_{i=1}^{m} x_i^2 + \sum_{i=1}^{n} y_i^2).$$

Wir haben noch keine Ahnung, wie d konkret aussehen wird! Wegen (4.6) können wir jetzt aber gerade so gut für noch frei zu wählende $b := b_1 = b_2, c$ eine Funktion von

$$(\sum_{i=1}^{m}(cx_i + b), \sum_{i=1}^{n}(cy_i + b), \sum_{i=1}^{m}(cx_i + b)^2 + \sum_{i=1}^{n}(cy_i + b)^2)$$

oder geschickter

$$(mb + c\sum_{i=1}^{m} x_i, nb + c\sum_{i=1}^{n} y_i, \sum_{i=1}^{m}(cx_i + b)^2 + \sum_{i=1}^{n}(cy_i + b)^2)$$

untersuchen. Dies machen wir unter geschickter Wahl von b, c; nämlich: wähle b, c so, dass wir am Schluss einen Ausdruck der Art

$$(m, -n, \text{etwas})$$

haben. Über ein einfaches lineares 2×2-Gleichungssystem erhalten wir für b, c:

$$b = \frac{\overline{x} + \overline{y}}{\overline{y} - \overline{x}}, \quad c = \frac{2}{\overline{x} - \overline{y}}.$$

Das "etwas" sollten wir vielleicht noch etwas genauer bestimmen:

$$(m, -n, \frac{1}{(\overline{x} - \overline{y})^2}(\sum_{i=1}^{m}(-(\overline{x} + \overline{y}) + 2x_i)^2 + \sum_{i=1}^{n}(-(\overline{x} + \overline{y}) + 2y_i)^2). \tag{4.12}$$

Was wir jetzt mit Ausdruck (4.12) anstellen, ist noch nicht bestimmt. Wir werden den relevanten Ausdruck

$$\begin{aligned} Z &:= \frac{1}{(\overline{x} - \overline{y})^2}(\sum_{i=1}^{m}(-(\overline{x} + \overline{y}) + 2x_i)^2 + \sum_{i=1}^{n}(-(\overline{x} + \overline{y}) + 2y_i)^2) \\ &= \frac{1}{(\overline{x} - \overline{y})^2}(\sum_{i=1}^{m}(2(x_i - \overline{x}) - (\overline{y} - \overline{x}))^2 + \sum_{i=1}^{n}(2(y_i - \overline{y}) + (\overline{y} - \overline{x}))^2) \end{aligned}$$

jedoch noch ein wenig umformen:

$$\begin{aligned} Z(\overline{x} - \overline{y})^2 &= 4\sum_i (x_i - \overline{x})^2 - 4(\overline{y} - \overline{x})\sum_i (x_i - \overline{x}) + m(\overline{y} - \overline{x})^2 \\ &\quad + 4\sum_i (y_i - \overline{y})^2 + 4(\overline{y} - \overline{x})\sum_i (y_i - \overline{y}) + n(\overline{y} - \overline{x})^2 \\ &= 4\sum_i (x_i - \overline{x})^2 + m(\overline{y} - \overline{x})^2 + 4\sum_i (y_i - \overline{y})^2 + n(\overline{y} - \overline{x})^2. \end{aligned}$$

Damit haben wir

$$Z = 4\frac{\sum_i (x_i - \overline{x})^2 + \sum_i (y_i - \overline{y})^2}{(\overline{x} - \overline{y})^2} + m + n.$$

Die "4" und $m + n$ sind nicht informativ und wir nehmen gleich den Kehrwert

$$\frac{(\overline{x} - \overline{y})^2}{\sum_i (x_i - \overline{x})^2 + \sum_i (y_i - \overline{y})^2}.$$

Sieht schon sehr nach t-Verteilung aus. Wir können sicher die Wurzel nehmen und erhalten (Vorzeichen wegen Symmetrie ohne Probleme)

$$\frac{\overline{x}-\overline{y}}{\sqrt{\sum_i (x_i-\overline{x})^2+\sum_i (y_i-\overline{y})^2}}.$$

Tabelliert (bzw. in Rechenlibraries abgelegt) ist jedoch der Ausdruck

$$\frac{\overline{x}-\overline{y}}{\sqrt{\frac{1}{m}+\frac{1}{n}}}\frac{1}{\sqrt{\frac{\sum_i (x_i-\overline{x})^2+\sum_i (y_i-\overline{y})^2}{m+n-2}}}. \tag{4.13}$$

(4.13) hat (als Zufallsgrösse) unter $\mathcal{H}_0$ eine t_{m+n-2}-Verteilung (vgl. 1.4.2.8). Warum?

Man nimmt dann üblicherweise mit dem Ansatz von Cox- und Hinkley je links und rechts ein $\alpha/2$ weg und lehnt dort aussen $\mathcal{H}_0$ ab.

Further Readings: Stahel: 8.8i / Lam: 7.4.2 / Dalgaard: 4.3 (var.equal=T)

Historisch darf man keine falschen Schlüsse ziehen: viele Verteilungen der Statistik (v.a. χ^2, t, F) *sind aus statistischen Problemstellungen heraus entstanden und tabelliert worden* (z.B. oberste und unterste 5 %-Quantile). Man hat diese Verteilungen nicht einfach so mal studiert und dann überraschend festgestellt, dass sie in der Statistik bei konkreten Problemstellungen (wieder) auftreten.

In 4.4.3 und 4.4.5 haben wir am Schluss je einen Quotienten erhalten, von dem wir gesagt haben, dass er eine t-Verteilung hat (unter $\mathcal{H}_0$). Wenn wir die Definition der t-Verteilung in 1.4.2.8 anschauen, so steht dort, dass Zähler und Nenner unabhängig sein müssen. Dies ist jeweils der Fall und wird in Kapitel 7 in allgemeinerem Zusammenhang ohne Zirkelschluss gezeigt werden.

4.5 Miscellanea (χ^2-Tests, Kolmogorov-Smirnov, einfache Varianzanalyse, asymptotische Tests)

Der bisherige Aufbau von Kapitel 4 ist streng logisch gewesen:

in 4.1 bis 4.3	haben wir die klassische Testtheorie entwickelt (von einfachen Situationen zu komplexeren Tests, bei denen wir nur noch mit Einschränkungen Optimalität erhalten (unbiased))
in 4.4	haben wir eine allgemeine Theorie entwickelt: Invarianzüberlegungen halfen uns nach Suffizienzüberlegungen Probleme weiter zu reduzieren, sodass die allseits bekannten Teststatistiken gewonnen werden konnten
in 4.5	folgt jetzt eine lose und sehr unvollständige Sammlung von wichtigen Tests aus der Praxis.

4.5.1 χ^2-Tests

Die Beweise zu 4.5.1 finden sich in Statistik-Büchern für fortgeschrittene Semester. Wir verzichten hier aus Zeitgründen im Bachelor-Studium darauf und beschränken uns auf die Motivation und Präsentation der Resultate. Mehr dazu (empfohlen, sind nur 12 Seiten) in Stahel Kapitel 10 - die Beispiele sind von dort.

χ^2-Anpassungstest

Wir motivieren die Grundfrage mit einem historischen Beispiel. Aufgrund der Mendelschen Erbgesetze erwartet man in bestimmten Kreuzungsversuchen, dass Nachkommen in drei verschiedenen Typen mit den Wahrscheinlichkeiten 0.25 (aa), 0.5 (aA) und 0.25 (AA) auftreten.

In einer Untersuchung (Daten aus Stahel) fand man: $s_1 = 29, s_2 = 77$ und $s_3 = 44$ bei 150 Nachkommen. Wäre es exakt "fair" gewesen, so "müsste" es sein: $150/4 = 37.5$, $150/2 = 75$ und wieder $150/4 = 37.5$. Offenbar gibt es im Versuch eine Abweichung gegenüber diesen "fairen" Werten und jedeR WissenschaflerIn, welche mit genau diesen "fairen" Werten aufwarten würde, muss als Statistik-Banause bezeichnet werden (selbst wenn er/sie bei den halben Fällen *je einmal* auf- und abrundet). *Eine gewisse Variation in den Daten ist zu erwarten! Aber wieviel ist noch mit dem Modell von Mendel vereinbar? Ab wann müssen wir von einer signifikanten Abweichung sprechen?*

Wir müssen dazu erstmal eine neue Zufallsgrösse einführen, welche zu Recht an die Binomialverteilung von (1.4.1.2) erinnert: die **Multinomialverteilung** (die Binomialverteilung ist ein Spezialfall hiervon). Wenn wir n unabhängige Versuche haben mit m möglichen Ereignissen (in obigem Beispiel: $n = 150, m = 3$ - bei der Binomialverteilung $m = 2$ (Erfolg/Misserfolg)), so ist ein Zufallsvektor $(S_1, \ldots, S_m)$ per Definitionem multinomialverteilt, wenn

$$P[S_1 = s_1, \ldots, S_m = s_m] = \frac{n!}{s_1! \ldots s_m!} \pi_1^{s_1} \ldots \pi_m^{s_m}.$$

Dabei muss dann gelten: $s_1 + \ldots + s_m = n$ und $\pi_1 + \ldots + \pi_m = 1$.

Die Hypothesen sind derart, dass der Zufallsvektor $(S_1, \ldots, S_m)$ unter $\mathcal{H}_0$ eine Multinomialverteilung mit Wahrscheinlichkeiten $\pi_1, \ldots, \pi_m$ hat gegen $\mathcal{H}_1$ einer Multinomialverteilung mit Wahrscheinlichkeiten $\theta_1, \ldots, \theta_m$, wo mindestens bei zwei (!) Wahrscheinlichkeitspaaren (π_i, θ_i), $1 \leq i \leq m$, gelten muss $\pi_i \neq \theta_i$.

Wie bei der Binomialverteilung wird man auch hier die Parameter sinnvollerweise mit

$$\hat{\pi}_i := s_i/n$$

schätzen (das haben Sie höchstwahrscheinlich im Leben schon häufig getan, ohne sich bewusst zu sein, dass Sie es hier mit einer Multinomialverteilung zu tun haben - ist nicht weiter schlimm). Damit erhalten wir m Abweichungen

$$\hat{\pi}_i - \pi_i, \ 1 \leq i \leq m.$$

Positive und negative Abweichungen könnten sich aufheben. Somit ist es sinnvoll, z.B. die absoluten Werte zu betrachten:

$$|\hat{\pi}_i - \pi_i| = |s_i - n\pi_i|/n, \ 1 \le i \le m.$$

Wir wollen diese m Werte alle zusammen in einer einzigen Testgrösse betrachten und es hat sich herausgestellt, dass

$$u := \sum_{i=1}^{m} \frac{(s_i - n\pi_i)^2}{n\pi_i} \qquad (\chi^2 - \text{Anpassung})$$

eine brauchbare Testgrösse ist. Mit $n \to \infty$ gilt, dass diese Grösse als Zufallsgrösse U unter $\mathcal{H}_0$ eine χ^2_{m-1}-Verteilung besitzt. In Stahel sind Praktikerregeln aufgeführt, ab wann wir n als genügend gross bezeichnen dürfen. Es ist noch beachtenswert, dass in der Referenzverteilung nur m und nicht die einzelnen π_i's vorkommen. Wann werden wir die $\mathcal{H}_0$-Hypothese ablehnen, das heisst, von welcher Form ist der Ablehnungsbereich?

Wir untersuchen jetzt mit diesem Test das einführende Beispiel:

$$\begin{aligned} u &= \frac{(29 - 150 * 0.25)^2}{150 * 0.25} + \frac{(77 - 150 * 0.5)^2}{150 * 0.5} + \frac{(44 - 150 * 0.25)^2}{150 * 0.25} \\ &= 1.927 + 0.053 + 1.127 \\ &= 3.11. \end{aligned}$$

Es gilt: $P[\chi^2_2 > 5.99] = 0.05$; somit werden wir auf dem $\alpha = 5\%$-Niveau die $\mathcal{H}_0$-Hypothese nicht ablehnen. Die Versuchsresultate sind demnach auf diesem Niveau durchaus mit den Mendel'schen Gesetzen vereinbar.

χ^2-Test für Unabhängigkeit in Kontingenztafeln

Wir motivieren die Grundfrage wieder mit einem Beispiel aus Stahel (ursprünglich aus einer EMNID-Erhebung). Kurz zusammengefasst: "Hat die Schulbildung etwas mit der Wahrnehmung von Umweltbelastungen zu tun?" Dazu hat man 2004 Personen befragt. Die Schulbildung geht von 1 (ungelernt) bis 5 (Hochschulabschluss) und die Antworten auf die Frage, ob man sich durch Umweltschadstoffe beeinträchtigt fühle, gingen von 1 (überhaupt nicht beeinträchtigt) über 2 (etwas beeinträchtigt), 3 (ziemlich beeinträchtigt), 4 (sehr beeinträchtigt).

Schulbildung	1	2	3	4	Summe
1	212	85	38	20	355
2	434	245	85	35	799
3	169	146	74	30	419
4	79	93	56	21	249
5	45	69	48	20	182
Summe	939	638	301	126	2004

Tabelle 4.1: Schulbildung und Wahrnehmung von Umweltbelastung; Lese-Beispiel: 79 Personen haben Bildungsniveau 4 und gaben an, dass sie sich überhaupt nicht beeinträchtigt fühlen.

Allgemein hat man ein Schema der folgenden Art:

Merkmale $X\backslash Y$	1	2	...	s	Rand-Summe
1	n_{11}	n_{12}	...	n_{1s}	$n_{1.}$
2	n_{21}	n_{22}	...	n_{2s}	$n_{2.}$
...	...	...	n_{ij}	...	$n_{i.}$
r	n_{r1}	n_{r2}	...	n_{rs}	$n_{r.}$
Rand-Summe	$n_{.1}$	$n_{.2}$	... $n_{.j}$ ...	$n_{.s}$	n

Tabelle 4.2: Allgemeiner Fall von Kontingenztafeln oder Kreuztabellen; n ist die Gesamtanzahl Versuche; die Randsummen sind die $n_{.j}$ und $n_{i.}$. Merkmal X hat r mögliche Ausprägungen und Merkmal Y deren s. Lese-Beispiel zum Vergleich: die 79 Personen von oben sind hier n_{41}.

Wir haben den Merkmalen gleich Bezeichungen gegeben, welche an Zufallsgrössen erinnern sollten: X und Y. Wir stellen jetzt ein Wahrscheinlichkeitsmodell auf:

$$P[X = i, Y = j] =: \pi_{ij}.$$

Die obigen Anzahlen n_{ij} in den einzelnen Kästchen fassen wir auch als Realisationen von Zufallsgrössen N_{ij} auf. Dabei sind *die einzelnen Versuche unabhängig voneinander - nicht aber die einzelnen N_{ij}!* Der Zufallsvektor $(N_{11}, \ldots, N_{rs})$ ist multinomialverteilt:

$$P[N_{11} = n_{11}, \ldots, N_{rs} = n_{rs}] = \frac{n!}{n_{11}! \ldots n_{rs}!} \pi_{11}^{n_{11}} \ldots \pi_{rs}^{n_{rs}}.$$

Dabei muss dann gelten: $n_{11} + \ldots + n_{rs} = n$ und $\pi_{11} + \ldots + \pi_{rs} = 1$. Damit sind wir also in der gleichen Situation wie beim χ^2-Anpassungstest.

Was uns hingegen jetzt interessiert, ist ob die beiden Ausprägungen unabhängig voneinander auftreten (im Beispiel: Es besteht kein Zusammenhang zwischen Schulbildung und Wahrnehmung von Umweltbelastungen). Formalisiert wird das in der Forderung (Nullhypothese)

$$P[X = i, Y = j] = P[X = i]P[Y = j] =: \pi_{i.}\pi_{.j} \;\; \forall \, (i, j).$$

Von (χ^2-Anpassung) erhalten wir jetzt folgende Teststatistik

$$u^{(\text{prov})} := \sum_{i,j} \frac{(N_{ij} - n\pi_{i.}\pi_{.j})^2}{n\pi_{i.}\pi_{.j}}. \qquad (\chi^2 - \text{Test auf Unabhaengigkeit, provisorisch})$$

Die $\pi_{i.}$ und $\pi_{.j}$ sind unbekannt und wir müssen sie aus den Daten mit $\hat{\pi}_{i.} := N_{i.}/n, \hat{\pi}_{.j} := N_{.j}/n$ schätzen. Dadurch wird obiger Ausdruck zur definitiven Variante:

$$u := \sum_{i,j} \frac{(N_{ij} - N_{i.}N_{.j}/n)^2}{N_{i.}N_{.j}/n}. \qquad (\chi^2 - \text{Test auf Unabhaengigkeit})$$

Man kann zeigen, dass diese Grösse als Zufallsgrösse U und bei Unabhängigkeit der Merkmale (Nullhypothese) mit $n \to \infty$ eine χ^2-Verteilung mit $(r-1)(s-1)$ Freiheitsgraden besitzt.

Im einleitenden Beispiel (Schulbildung / Umweltbelastung) erhalten wir einen Wert $u = 125$. Der P-Wert von 125 bei einer χ^2-Verteilung mit $(5-1)(4-1) = 12$ Freiheitsgraden ist $1 - \text{pchisq}(125, 12)) \doteq 0$; bei $\alpha = 5\%$ erhalten wir wegen $1 - \text{pchisq}(21.03, 12) = 0.04994287$ einen Ablehnungsbereich der Nullhypothese ab Werten $u \geq 21.03$. Was schliessen wir daraus? Was sollten wir nicht schliessen?

2 Nachträge: χ^2-Verteilung mit kleinen df und Begründung der Formel

I Vorsicht: χ^2-Verteilung mit kleinen df

Wir haben die Dichtefunktion der χ^2-Verteilung bis jetzt immer nur für grosse df $= n$ gezeichnet (wegen der Verteilung des Schätzers der Varianz). In 4.5.1 sind kleine Freiheitsgrade aber häufig. Wir sollten deshalb die χ^2-Verteilung ein bisschen besser kennen lernen und in R/S-PLUS folgende Befehle als kleine Hausaufgabe ausführen:

x <- seq(0,5,0.1)
plot(x,dchisq(x,1),type="l") (∞ bei 0; vgl. 1.4.2.6)
plot(x,dchisq(x,2),type="l") (ist Exponentialverteilung; vgl. 1.4.2.2, 1.4.2.3 und 1.4.2.6)
plot(x,dchisq(x,3),type="l")
plot(x,dchisq(x,4),type="l")
plot(x,dchisq(x,5),type="l")
plot(x,dchisq(x,6),type="l")

Ein Beispiel mit $r = s = 2$ (sogenannte Vierfeldertafel) ist in den Übungen zu rechnen (Todesstrafe in den USA; das Beispiel zeigt auch, wie eine unvollständige Datenanalyse zu falschen Schlüssen führen kann).

II Begründung der Formel

Bevor wir zu χ^2-Anpassung eine Begründung der Formel angeben, wollen wir im Fall $m = 2$ den Ausdruck

$$u := \sum_{i=1}^{m} \frac{(s_i - n\pi_i)^2}{n\pi_i}$$

genauer untersuchen.

$$\begin{aligned}
\frac{(s_1 - n\pi_1)^2}{n\pi_1} + \frac{(s_2 - n\pi_2)^2}{n\pi_2} &= \frac{(s_1 - n\pi_1)^2}{n\pi_1} + \frac{(n - s_1 - n(1 - \pi_1))^2}{n(1 - \pi_1)} \\
&= \frac{(s_1 - n\pi_1)^2}{n\pi_1} + \frac{(s_1 - n\pi_1)^2}{n(1 - \pi_1)} \\
&= \frac{(s_1 - n\pi_1)^2(\pi_1 + 1 - \pi_1)}{n\pi_1(1 - \pi_1)} \\
&= \frac{(s_1 - n\pi_1)^2}{n\pi_1(1 - \pi_1)} \\
&= \left(\frac{s_1 - n\pi_1}{\sqrt{n\pi_1(1 - \pi_1)}}\right)^2
\end{aligned}$$

S_1 ist Binomialverteilt mit Parametern n, π_1. Der Ausdruck

$$\frac{S_1 - n\pi_1}{\sqrt{n\pi_1(1 - \pi_1)}}$$

konvergiert wegen Theorem 1.25 (CLT) gegen eine $\mathcal{N}(0, 1)$-Zufallsgrösse. Das Quadrat einer $\mathcal{N}(0, 1)$-Zufallsgrösse hat eine χ^2-Verteilung mit Parameter 1. Damit ist der χ^2-Anpassungstest für df $= (m - 1) = 1$ direkt bestätigt worden.

Wenn wir bei $m = 2$ (Binomialverteilung mit $p = \pi$ hier) einen Test ob $\pi = \pi_0$ ist durchführen, würden wir natürlich direkt mit einer Binomialverteilung und der Theorie am Anfang dieses Kapitels arbeiten (genaueres Verfahren). Wir haben jetzt aber über den zentralen Grenzwertsatz eine Approximation für grosse n erhalten. Ob man dann direkt mit der Normalverteilung arbeitet oder zuerst quadriert und dann mit der χ^2_1-Verteilung arbeitet ist egal - die Entscheidung ($\mathcal{H}_0$ ablehnen oder annehmen) sollte bei korrekten Rechnungen nicht ändern.

Wie sieht jetzt die Begründung der Formel für den allgemeinen Fall aus (ohne Berücksichtigung der fehlenden Unabhängigkeit und um eins kleineren Freiheitsgrad)?

Mehr dazu in Dalgaard 7.4 und wie erwähnt in Stahel, Kapitel 10.

4.5.2 Kolmogorov-Smirnov

Im Rahmen von "Überprüfen von Voraussetzungen" wird man mit mehr oder weniger mathematischen und exakten Methoden die gemachten Annahmen einer Prüfung unterziehen wollen (sind die Daten normalverteilt, unabhängig voneinander generiert worden, etc.). In Stahel wird in Kapitel 11 eine neunseitige Übersicht gegeben, welche lesenswert ist. Wir behandeln hier im Rahmen der mathematischen Statistik lediglich den Kolmogorov-Smirnov-Test (Goodness of Fit-Test). Dies obschon er in wichtigen Abweichungen von der Nullhypothese wenig Macht besitzt.

Die mathematischen Resultate sind jedoch so überraschend und einfach herzuleiten, dass wir diesen Test unbedingt anschauen sollten. Wir beginnen mit einem Resultat, welches bei bisherigen Beispielen und in den Übungen vielleicht da und dort schon vermutet wurde: bei stetigen Testgrössen gilt, dass **P-Werte unter der Nullhypothese $U[0,1]$-verteilt sind**. Wir zeigen vorerst mal

Lemma 4.17 [Verteilung von $F_X(X)$] *Hat die Zufallsgrösse X die stetige Verteilungsfunktion F_X, so hat $F_X(X)$ eine $U[0,1]$-Verteilung.*

Beweis von Lemma 4.17: Wir definieren $x_u := \max\{x | F_X(x) = u\}$. Sei jetzt $U := F_X(X)$, dann gilt (vgl. nachfolgende Skizze) für $0 \leq u \leq 1$:

$$F_U(u) = P[U \leq u] = P[F_X(X) \leq u] = P[X \leq x_u] = F_X(x_u) = u.$$

Für $u < 0$ gilt offensichtlich $F_U(u) = 0$ und für $u > 1$ gilt $F_U(u) = 1$. Damit ist das Lemma bewiesen.

$\square$

Bemerkung zu Lemma 4.17: Wir haben angekündigt, dass wegen Lemma 4.17 gilt: bei stetigen Testgrössen sind P-Werte unter der Nullhypothese $U[0,1]$-verteilt. Weshalb ist dies so? Dazu wieder zuerst eine Skizze:

Sei jetzt T die stetige Testgrösse (eine Zufallsgrösse mit Verteilungsfunktion $F_T(u)$); sie nimmt in obiger Skizze Wert t an. Der P-Wert als Zufallsgrösse aufgefasst ist die Grösse $1 - F_T(T)$ (manchmal auch als $F_T(T)$ definiert; obiges bleibt richtig). Unter Verwendung von Lemma 4.17 folgt jetzt:

$$P[1 - F_T(T) \leq u] = P[F_T(T) \geq 1 - u] = 1 - P[F_T(T) < 1 - u] = 1 - (1 - u) = u,$$

d.h. in der Tat ist der P-Wert uniform verteilt.

Wir kommen jetzt zum eigentlichen Kolmogorov-Smirnov-Test. Von 3.2 repetierend: Wenn wir eine Stichprobe $(x_1, x_2, \ldots, x_n)$ vom Umfang n haben, so geben wir jedem Punkt x_i die Wahrscheinlichkeit $1/n$ (uniforme Verteilung). Wir nennen dies die empirische Verteilung. Die empirische Verteilungsfunktion $F_n(x)$ definiert man dann durch $F_n(x) := \frac{1}{n}|\{i|x_i \leq x\}|$. Stammt dann die Stichprobe aus einer Verteilung mit Verteilungsfunktion F_X, so gilt (wenn wir jetzt $F_n(x) := \frac{1}{n}|\{i|X_i \leq x\}|$ als Zufallsgrösse *vor* der Realisierung auffassen) wegen Satz 3.1 [Satz von Glivenko-Cantelli] fs:

$$\sup_x |F_n(x) - F(x)| \to 0$$

für $n \to \infty$. Dies ist jedoch nur ein asymptotisches Resultat. Wir werden in der Datenanalyse immer nur endliches n haben. Man wird sich dann die Frage stellen, wie weit zum Beispiel

$$D_n := \sup_x |F_n(x) - F(x)| \qquad \text{(KS – GoF – Test)}$$

von Null verschieden sein darf, **wenn F die richtige Verteilung ist ($\mathcal{H}_0$-Hypothese)**, wenn also zum Beispiel die Stichprobe aus einer Normalverteilung stammt und $F(x)$ auch die Verteilungsfunktion dieser Normalverteilung ist (Wir könnten auch Daten aus einer Cauchy-Verteilung generieren und mit der Verteilungsfunktion der Standardnormalverteilung vergleichen - dann sollte die Abweichung von 0 tendenziell grösser sein als unter $\mathcal{H}_0$). Wir präzisieren: der Zufall ist im F_n; F ist eine fixe Funktion. Man erhält dann Darstellungen der folgenden Art (Beiblatt, generiert mit folgenden Befehlen):

```
a<-rnorm(50)
x<-seq(-4,4,0.1)
plot(x,pnorm(x),type=”l”)
lines(sort(a),(1:50)/50)
ks.test(a,”pnorm”)
One-sample Kolmogorov-Smirnov test
data: a
D = 0.1659, p-value = 0.1276 [Wir können die (zutreffende) Nullhypothese annehmen]
alternative hypothesis: two.sided
```

Das praktische und nach Lemma 4.17 nicht mehr so überraschende ist, dass wir nicht für jede Verteilung F eine eigene Tabelle für Ablehnungsbereiche anlegen müssen:

Satz 4.18 [**D_n hängt nur von n ab!**] *Bei stetigen Verteilungsfunktionen gilt: Unter $\mathcal{H}_0$ hat*

$$D_n := \sup_x |F_n(x) - F(x)| \qquad \text{(KS – GoF – Test)}$$

eine Verteilung, welche nur von n abhängt (und insbesondere nicht von F).

Bemerkung zu Satz 4.18: Die Verteilung von D_n ist zum Beispiel in R/S-PLUS abgelegt und in allen grösseren Statistik-Werken.

Beweis von Satz 4.18: Im Licht von Lemma 4.17 ist das Resultat eigentlich nicht so überraschend; Beweis im Lindgren; kurz aber trickreich.

□

4.5.3 Einfache Varianzanalyse

Die jetzt zu besprechende Situation ist der Prototyp von Statistics at it's best. Englisch nennt man das Gebiet ANOVA (=**An**alysis **o**f **Va**riance). Der einfachste Fall ist die einfache Varianzanalyse oder Einweg-Varianzanalyse (Single-Factor Experiment) - komplizierter Zweiweg-Varianzanalysen und mehr. In Kapitel 7 "Regression" werden wir die Resultate wieder als Spezialfall erhalten.

Wir entwickeln die Theorie zur ANOVA anhand eines Beispiels: In einem AGRO-Konzern wird das Wachstum von Pflanzen unter 4 klar kontrollierten Bedingungen untersucht. Wir nennen die 4 verschiedenen Bedingungen I-IV. Dies kann zum Beispiel unterschiedliche Mengen von Düngemittel sein. Wir erhalten dann zum Beispiel folgende Tabelle des Wachstums der Pflanzen über einen gegebenen Zeitraum:

I	II	III	IV
33.3	35.5	29.6	38.5
47.8	35.4	33.4	42.4
44.4	47.6	32.8	45.5
42.9	38.8	38.8	38.9
40.9		42.8	38.9
35.5			44.5

Tabelle 4.3: Wachstum von Pflanzen unter 4 verschiedenen Bedingungen (1 Faktor); Lese-Beispiel: die vierte Pflanze unter Bedingungen I wuchs im beobachteten Zeitraum 42.9 cm hoch. In der zweiten Kategorie sind 2 Pflanzen leider durch ein Feuer zerstört worden und in der dritten Kategorie hat man bei einer Pflanze die Samen verwechselt - such is life - wir werden die Analyse trotzdem durchführen können.

Wir werden jetzt ein mathematisches Modell mit k Gruppen aufstellen (oben $k = 4$):

$$Y_{ij} = \mu_j + E_{ij} \quad (1 \leq j \leq k; 1 \leq i \leq n_j). \tag{4.14}$$

Dabei bezeichne n_j die Anzahl Messungen in Gruppe j (oben (6,4,5,6)). Die totale Anzahl Beobachtungen ist

$$n = \sum_{j=1}^{k} n_j \ .$$

Die Zahl 42.9 in der Beschreibung ist dann die Realisation von Y_{41}. Wir fassen die Messgrössen also als Realisationen von Zufallsgrössen auf. Die E_{ij} sind iid $\mathcal{N}(0, \sigma^2)$-verteilt. **Achtung: Gleiche Varianz in allen Gruppen!** Damit gilt

$$Y_{ij} \sim \mathcal{N}(\mu_j, \sigma^2)$$

und die n_j Realisationen in Gruppe j sind unabhängig voneinander. Wir interessieren uns jetzt für die Frage, ob die k Mittelwerte gleich sind oder nicht. Wir wollen in obigem Beispiel ein hohes Wachstum erzielen und den Dünger auswählen, der das höchste Wachstum hervorbringt. Die Nullhypothese lautet

$$\mu_1 = \mu_2 = \ldots = \mu_k. \qquad (\mathcal{H}_0 - \text{Hypothese})$$

Sicher werden wir aus den Daten die unbekannten Mittelwerte schätzen; in Kapitel 5 werden wir sehen, dass bei Normalverteilungen das arithmetische Mittel nach fast jedem denkbaren Kriterium die beste Wahl für einen solchen Schätzer ist. Wir werden also den Mittelwert μ_1 in Gruppe 1 durch

$$\hat{\mu}_1 := \bar{Y}_{.1} := \frac{\sum_{i=1}^{n_1} Y_{i1}}{n_1}$$

schätzen und alle weiteren nach der allgemeinen Formel

$$\hat{\mu}_j := \bar{Y}_{.j} := \frac{\sum_{i=1}^{n_j} Y_{ij}}{n_j}$$

wo $1 \leq j \leq k$. Damit haben wir jetzt k geschätzte Mittelwerte - sie werden mit Wahrscheinlichkeit 1 alle verschieden sein. Sollen wir jetzt einfach den grössten nehmen und die entsprechende Gruppe zum Sieger erklären? Nein! Aus unserer bisherigen Erfahrung in Kapitel 4 wissen wir, dass auch unter der Nullhypothese immer einer am grössten sein wird *und* dass dieser eine signifikant grösser sein muss, damit wir die Nullhypothese verwerfen. Aber was heisst signifikant? Je nachdem, wie gross σ^2 ist, ist auch mit grösseren Abweichungen selbst bei Gültigkeit der Nullhypothese zu rechnen. Zudem kennen wir σ^2 gar nicht. Es scheint hoffnungslos zu sein - aber:

Wir können ja auch σ^2 schätzen. Dies geschieht in 2 Schritten und wird uns (eher unerwartet) durch geschickte Umformung gleich die Lösung des Problems liefern.

1. Unter der Nullhypothese haben wir eine iid-Stichprobe und wir können alle n Datenpunkte gleichberechtigt zur Berechnung eines **G**rand **M**ean GM einsetzen: Mit

$$GM := \frac{\sum_{j=1}^{k} n_j \bar{Y}_{.j}}{n} = \frac{\sum_{j=1}^{k} n_j \frac{\sum_{i=1}^{n_j} Y_{ij}}{n_j}}{n} = \frac{\sum_{j=1}^{k} \sum_{i=1}^{n_j} Y_{ij}}{n}$$

können wir in einem

2. Schritt einen Schätzer für die Varianz (genauer das n-fache davon) angeben mit:

$$\begin{aligned}
\sum_{j=1}^{k}\sum_{i=1}^{n_j}(Y_{ij}-GM)^2 &= \sum_{j=1}^{k}\sum_{i=1}^{n_j}\Big((\bar{Y}_{.j}-GM)+(Y_{ij}-\bar{Y}_{.j})\Big)^2 \\
&= \sum_{j=1}^{k}\sum_{i=1}^{n_j}(\bar{Y}_{.j}-GM)^2 + \sum_{j=1}^{k}\sum_{i=1}^{n_j}(Y_{ij}-\bar{Y}_{.j})^2 \\
&\quad + 2\sum_{j=1}^{k}\sum_{i=1}^{n_j}(\bar{Y}_{.j}-GM)(Y_{ij}-\bar{Y}_{.j}) \\
&= \sum_{j=1}^{k}\sum_{i=1}^{n_j}(\bar{Y}_{.j}-GM)^2 + \sum_{j=1}^{k}\sum_{i=1}^{n_j}(Y_{ij}-\bar{Y}_{.j})^2 \\
&\quad + 2\sum_{j=1}^{k}(\bar{Y}_{.j}-GM)\sum_{i=1}^{n_j}(Y_{ij}-\bar{Y}_{.j}) \\
&= \sum_{j=1}^{k}\sum_{i=1}^{n_j}(\bar{Y}_{.j}-GM)^2 + \sum_{j=1}^{k}\sum_{i=1}^{n_j}(Y_{ij}-\bar{Y}_{.j})^2 \\
&\quad + 2\sum_{j=1}^{k}(\bar{Y}_{.j}-GM)*0 \\
&= \sum_{j=1}^{k}\sum_{i=1}^{n_j}(\bar{Y}_{.j}-GM)^2 + \sum_{j=1}^{k}\sum_{i=1}^{n_j}(Y_{ij}-\bar{Y}_{.j})^2.
\end{aligned}$$

Wir haben also:

$$\sum_{j=1}^{k}\sum_{i=1}^{n_j}(Y_{ij}-GM)^2 = \sum_{j=1}^{k}\sum_{i=1}^{n_j}(\bar{Y}_{.j}-GM)^2 + \sum_{j=1}^{k}\sum_{i=1}^{n_j}(Y_{ij}-\bar{Y}_{.j})^2. \qquad \text{(Fundi ANOVA)}$$

Diese Gleichung wird auch als Fundamentalgleichung der Varianzanalyse bezeichnet (deshalb (Fundi ANOVA)). Was sagt sie aus? Die gesamte Summe der quadratischen Abweichungen der einzelnen Beobachtungen vom Grand Mean (linke Seite) lässt sich aufspalten (”+”) in Summe der quadrierten Abweichungen der Behandlungsmittelwerte vom Grand Mean (”zwischen den Behandlungen”, erster Summand) und der Summe der quadrierten Abweichungen der einzelnen Beobachtungen vom jeweiligen Behandlungsmittelwert (”innerhalb der Behandlung”, zweiter Summand).

Zwischenfrage an’s Publikum: Angenommen, die Annahme gleicher Mittelwerte ist verletzt. Welcher der beiden Summanden auf der rechten Seite von (Fundi ANOVA) wird tendenziell grösser im Verhältnis zum anderen? Gehen Sie bei solchen Überlegungen in die Extreme - kein Tipp für alle Lebenslagen!

Wir werden in Kapitel 7 ”Regression” die erstaunliche Tatsache beweisen, dass die Summanden auf der rechten Seite stochastisch unabhängig sind. Dies wird in einem Zug bewiesen, wenn wir auch zeigen, dass in 4.4.3 und 4.4.5 Zähler und Nenner der T-Teststatistiken unabhängig voneinander sind. Ein Hinweis, dass das alles in einem geht ist, dass wir mit $(t_n)^2$ eine F_{1n}-Zufallsgrösse haben.

Damit bietet sich als Test-Statistik folgende Grösse an:

$$V := \frac{\sum_{j=1}^{k} n_j(\bar{Y}_{.j} - GM)^2/(k-1)}{\sum_{j=1}^{k}\sum_{i=1}^{n_j}(Y_{ij} - \bar{Y}_{.j})^2/(n-k)}.$$

Diese Test-Statistik hat unter $\mathcal{H}_0$ eine $F_{k-1,n-k}$-Verteilung. Weshalb?

Wir werden den Ablehnungsbereich der Nullhypothese einseitig dort wählen, wo diese Statistik gross ist. Auf www.luchsinger-mathematics.ch/onewayanova.txt ist eine R-Session zur Analyse des obigen Beispiels angegeben. Mehr in Dalgaard Kapitel 6, Lam p. 98, Stahel 12.1.

4.5.4 Asymptotische Tests

Wir haben in einer Aufgabe und bei den χ^2-Tests bereits asymptotische Resultate benutzt. Wir werden jetzt noch 4 Situationen anfügen, wo wir wegen des CLT (Theorem 1.25) mit der Standard-Normalverteilung einen asymptotischen Test machen können. Dies ist dann praktisch, wenn man gerade weder Statistik-Paket noch Statistik-Literatur zur Hand hat und "Handgelenk mal π" abschätzen will, wie signifikant irgendwelche Daten sind. **Wir setzen immer iid Messgrössen mit endlicher Varianz voraus (dann geht der CLT!)**. Ein Beispiel dazu findet sich in den Übungen. Solche Aufgaben sind mit dieser Methode ohne Rechner lösbar. Kritische Werte wie 1.64 und 1.96 sollte man als StatistikerIn auswendig kennen!

<u>Beliebige stetige Zufallsgrösse: ist $\mu = \mu_0$?</u> Etwa ab $n > 30$ (Praktikerregel) benutzt man die Testfunktion

$$\frac{\bar{x} - \mu_0}{\sigma}\sqrt{n}$$

und vergleicht mit $\mathcal{N}(0,1)$. Ist σ nicht bekannt, schätzt man es mit

$$\hat{\sigma} := \sqrt{\sum (x_i - \mu_0)^2/n}.$$

<u>Bernoulli-Versuch: ist $p = p_0$?</u> Etwa ab $np_0 \geq 5$ und $n(1-p_0) \geq 5$ (Praktikerregel) benutzt man die Testfunktion

$$\frac{\bar{x} - p_0}{\sqrt{p_0(1-p_0)}}\sqrt{n}$$

und vergleicht mit $\mathcal{N}(0,1)$.

<u>2 beliebige stetige Zufallsgrössen mit gleicher Varianz: ist $\mu_1 = \mu_2$?</u> Etwa ab $m, n > 30$ (Praktikerregel) benutzt man die Testfunktion

$$\frac{\bar{x} - \bar{y}}{\sigma\sqrt{1/m + 1/n}}$$

und vergleicht mit $\mathcal{N}(0,1)$. Ist σ nicht bekannt, schätzt man es mit

$$\hat{\sigma} := \sqrt{(\sum (x_i - \bar{x})^2 + \sum (y_i - \bar{y})^2)/(m+n-2)},$$

vgl. mit (4.13).

2 Stichproben, je aus Bernoulli-Verteilung: ist $p_1 = p_2$? Etwa ab $mp \geq 5$ und $m(1-p) \geq 5$ und $np \geq 5$ und $n(1-p) \geq 5$ (Praktikerregel, p schätzen) benutzt man die Testfunktion

$$\frac{\bar{x} - \bar{y}}{\sqrt{p(1-p)(1/m + 1/n)}}$$

und vergleicht mit $\mathcal{N}(0,1)$; p aus den Daten schätzen mit $\hat{p} := (\sum x_i + \sum y_i)/(m+n)$.

Damit haben wir die Testtheorie vorläufig abgeschlossen. Wir werden in Kapitel 7 "Regression" wieder testen - in einem Modell (dem linearen Modell). Testen und Schätzen von Parametern (Kapitel 5) sind wichtigste Aufgaben der Statistik. Beides wird in der Regression angewendet. Aber auch in den meisten anderen Modellen der Datenanalyse, welche wir aus Zeitgründen nicht behandeln, geht es immer wieder um Schätzen und Testen.

Falls Ihnen die vielen Tests, welche wir im langen Kapitel 4 behandelt haben, mittlerweile langweilig werden *und Sie MathematikerIn sind oder werden wollen*, dann liegt das hoffentlich weder am Skript noch am Dozenten selber sondern einfach daran, dass Sie das Grundprinzip auf einer höheren Ebene mittlerweile begriffen haben.

Du bist nicht allein:

What was the muddiest point so far - FRAGEN?

Auf diese Frage gab einE StudentIn folgende (schriftliche, anonyme) Antwort: "Ein ganzer Zoo von Tests, von denen jeder für sich betrachtet einsichtig ist, aber wie findet man den richtigen; oder wann ist welcher Test anwendbar?"

Antwort: In der praktischen Beratung wird man mit viel Erfahrung diese Tests immer besser kennenlernen (Pros und Contras; Dos and Donts). Dazu eignen sich auch Workshops/Konferenzen/Kurse wie auf www.math-jobs.com/conf.php angekündigt.