Beispiel 6.26 (zwei-Stichproben oder ungepaarter t-Test [mit Annahme gleicher Varianzen]). $\Theta = \mathbb{R}\times\mathbb{R}\times(0,\infty) \ni \vartheta = (\mu_1, \mu_2, \sigma^2)$, unter P_ϑ sind $X_1, \dots, X_m$ u.i.v. und davon unabhängig $Y_1, \dots, Y_n$ u.i.v. $(m, n \in \mathbb{N})$, $X_i \sim \mathcal{N}_{\mu_1,\sigma^2}$, $Y_j \sim \mathcal{N}_{\mu_2,\sigma^2}$. Seien

$$M_X := \frac{1}{m}\sum_{i=1}^m X_i, \ M_Y := \frac{1}{n}\sum_{j=1}^n Y_j$$

die jeweiligen Stichprobenmittelwerte,

$$S_X^2 = \frac{1}{m-1}\sum_{i=1}^m (X_i - M_X)^2, \ S_Y^2 = \frac{1}{n-1}\sum_{j=1}^n (Y_j - M_Y)^2,$$

die (korrigierten) Stichprobenvarianzen,

$$S^2 := \frac{(m-1)S_X^2 + (n-1)S_Y^2}{m+n-2} \quad \Big(= \frac{1}{m+n-2}\big(\sum_{i=1}^n (X_i - M_X)^2 + \sum_{j=1}^n (Y_j - M_Y)^2\big)\Big),$$

(die „gepoolte Stichprobenvarianz"),

$$T = \frac{M_X - M_Y}{S\sqrt{\frac{1}{m} + \frac{1}{n}}}.$$

(Beachte: Stets gilt $\mathbb{E}_{(\mu_1,\mu_2,\sigma^2)}[S^2] = \sigma^2$ [S^2 ist ein erwartungstreuer Schätzer für σ] und T ist unter $P_{(\mu_1,\mu_2,\sigma^2)}$ Student-$(n+m-2)$-verteilt, Argument analog zum Beweis von Satz 6.16).

1. Zweiseitiger ungepaarter t-Test : $\Theta_0 = \{(\mu_1, \mu_2, \sigma^2) \in \Theta : \mu_1 = \mu_2\}$ (oft knapp geschrieben als „$\Theta_0 : \mu_1 = \mu_2$"), $\Theta_1 = \Theta \setminus \Theta_0$.

Wähle $\alpha \in (0,1)$, mit $q := q_{m+n-2,1-\alpha/2} = (1-\alpha/2)$-Quantil der Student-$(m+n-2)$-Verteilung ist

$$\varphi(X_1, \dots, X_m, Y_1, \dots, Y_n) := \mathbf{1}_{\{|T|>q\}}$$

ein Test von Θ_0 gegen Θ_1 zum Niveau α.

1. Einseitiger Test : $\Theta_0 = \{(\mu_1, \mu_2, \sigma^2) \in \Theta : \mu_1 \le \mu_2\}$ (oft knapp geschrieben als „$\Theta_0 : \mu_1 \le \mu_2$“), $\Theta_1 = \Theta \smallsetminus \Theta_0$. Mit $q := q_{m+n-2,1-\alpha} = (1-\alpha)$-Quantil der Student-$(m+n-2)$-Verteilung ist

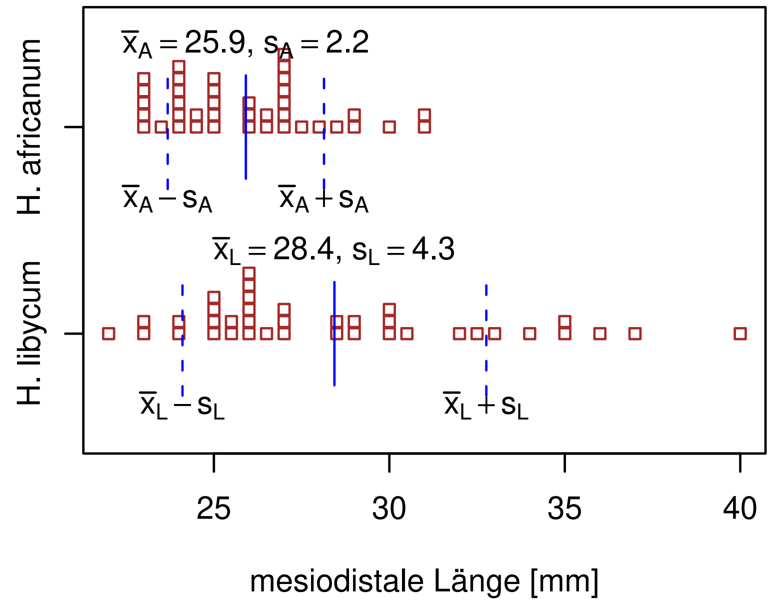
$$\varphi(X_1, \ldots, X_m, Y_1, \ldots, Y_n) := \mathbf{1}_{\{T>q\}}$$

ein Test von Θ_0 gegen Θ_1 zum Niveau α.
(Analog ist $\varphi(X_1, \ldots, X_m, Y_1, \ldots, Y_n) := \mathbf{1}_{\{T<-q\}}$ ein Test von $\Theta_0 = \{(\mu_1, \mu_2, \sigma^2) \in \Theta : \mu_1 \ge \mu_2\}$.)

(p-Werte werden analog zum ein-Stichproben-Fall (Bsp. 6.24) berechent, wobei $F_{T_{n-1}}$ durch $F_{T_{m+n-2}}$ ersetzt wird.)

Anwendungsbeispiel. Es wurden fossile Backenzähne gefunden, die zwei Arten von Urpferden zugeordnet wurden, und jeweils die („mesiodistale") Länge bestimmt. Wir möchten die (Null-)Hypothese prüfen, ob die mittlere Zahnlänge bei den beiden Arten gleich ist.

Die Daten:



Für *Hipparion africanum*:
$n_A = 39$, $\overline{x}_A = 25{,}9$, $s_A = 2{,}2$

für *Hipparion lybicum*:
$n_L = 38$, $\overline{x}_L = 28{,}4$, $s_L = 4{,}3$

Wir verwenden Signifikanzniveau $\alpha = 0{,}01$, das 99,5%-Quantil der Student-Vert. mit 75 Freiheitsgraden ist $\approxeq 2{,}64$. Es ist

$$s^2 = \frac{(n_A - 1)s_A^2 + (n_L - 1)s_L^2}{n_A + n_L - 1} \approxeq 11{,}94, \quad t = \frac{\overline{x}_A - \overline{x}_L}{s\sqrt{\frac{1}{n_A} + \frac{1}{n_L}}} \approxeq -3{,}229,$$

Wir können die Nullhypothese „die mittlere mesiodistale Länge bei *H. lybicum* und bei *H. africanum* sind gleich" zum Signifikanzniveau 1% ablehnen.
(Für ein Student-75-verteiltes T ist $P(|T| > 3{,}229) \approxeq 0{,}0018$, dies ist der p-Wert des Tests.)

Mögliche Formulierung dieses Befunds: „Die mittlere mesiodistale Länge war signifikant größer ($28{,}4$ mm) bei *H. libycum* als bei *H. africanum* ($25{,}9$ mm) (t-Test, $\alpha = 0{,}01$)."

Bericht 6.27. Die Tests aus Bsp. 6.23–6.26 sind (in ihrem jeweiligen statistischen Modell) optimal (sie sind „gleichmäßig beste Tests“) in dem Sinne, dass sie unter allen allen Tests von Θ_0 gegen Θ_1 mit Niveau α die größte Macht haben.

Dies ist das „Test-Analogon“ zu Bericht 6.8, vgl. z.B. die Diskussion in [G, Kap. 10.3 und 10.4].

Bericht 6.28 (zwei-Stichproben-t-Test ohne Annahme gleicher Varianz, Welchs t-Test[5]). Es gibt auch eine Version des zwei-Stichproben-t-Tests, der die Annahme gleicher Varianzen nicht trifft (wir werden ihn im Verlauf der Vorlesung allerdings nicht verwenden):

Man schätzt die Streuung von $M_X - M_Y$ durch

$$\sqrt{\frac{S_X^2}{n_X} + \frac{S_Y^2}{n_Y}} \quad \text{und bildet} \quad T = \frac{M_X - M_Y}{\sqrt{\frac{S_X^2}{n_X} + \frac{S_Y^2}{n_Y}}}.$$

$$\mathrm{Var}_{(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2)}[M_X - M_Y] = \frac{\sigma_1^2}{n_X} + \frac{\sigma_2^2}{n_Y}$$

Unter $P_{(\mu_0,\mu_0,\sigma_1^2,\sigma_2^2)}$ ist T „approximativ Student-verteilt mit g Freiheitsgraden", wobei

$$g = \frac{\left(\frac{s_X^2}{n_X} + \frac{s_Y^2}{n_Y}\right)^2}{\frac{s_X^4}{n_X^2(n_X-1)} + \frac{s_Y^4}{n_Y^2(n_Y-1)}}$$

aus den Daten geschätzt wird.

Seien die Werte $T = t$ und g beobachtet worden, man verwirft die Nullhypothese „$\mu_1 = \mu_2$" (zum Niveau α), wenn $1 - F_{T_g}(t) \leq \alpha/2$, wobei F_{T_g} die Verteilungsfunktion der Student-Verteilung mit g Freiheitsgraden, d.h. wenn die Wahrscheinlichkeit, dass eine Student-verteilte Zufallsgröße mit g Freiheitsgraden einen betragsmäßig mindestens so großen Wert wie den beobachteten t-Wert annimmt, $\leq \alpha$ ist.
(Wir hatten in Korollar 6.15 die Student-Verteilung nur für ganzzahlige Werte von n definiert, aber man kann dort allgemeine Werte $n > 0$ zulassen).

Dieser Test hat approximativ Niveau α und wird in der Praxis häufig verwendet. Beispielsweise führt der Befehl `t.test` in dem Statistikprogramm R automatisch diese Version des zwei-Stichproben-t-Tests durch, wenn man zwei Stichproben übergibt und keine weiteren Zusatzparameter setzt.

[5] B. L. Welch, The Significance of the Difference between Two Means When the Population Variances Are Unequal, Biometrika 29:350–362, (1938)

Bemerkung 6.29 (Zur „reinen Lehre“ des statistischen Testens). Nehmen wir an, wir möchten eine gewisse Aussage anhand experimenteller oder empirischer Daten statistisch prüfen. Das korrekte („lehrbuchmäßige“) Vorgehen sieht folgendermaßen aus:

1. Statistisches Modell formulieren, Nullhypothese und Alternative angeben (was die Nullhypothese ist, hängt von der konkreten Anwendungsfrage ab, oft ernennt man „das Gegenteil dessen, was man erhärten möchte“ zur Nullhypothese).

2. Dann einen Test (einschließlich gewünschtem Niveau) festlegen.

3. Dann erst: Daten erheben (bzw. Daten anschauen), Test-Entscheidung fällen.

Die Kontrolle der Fehlerwahrscheinlichkeiten, die die Theorie des statistischen Testens liefert, bezieht sich auf dieses Vorgehen. Wenn man die Reihenfolge herumdreht, also zuerst die Daten anschaut und dann einen Test wählt, verfälscht man strenggenommen zumindest das Signifikanzniveau, möglicherweise bis ins Unsinnige (Beispiel: zuerst den empirischen Mittelwert bestimmen, dann je nachdem, ob er links oder rechts von ϑ_0 liegt, entscheiden, ob man eine rechts- oder eine linksseitige Alternative wählt, ist offenbar „geschummelt“.)

Man sollte dieselben Daten nicht für explorative Statistik (d.h. Beobachtungen, die zu neuen Hypothesen führen [sollen]) und schließende Statistik (d.h. Beobachtungen, anhand denen eine Hypothese getestet werden soll) zugleich verwenden.

6.4.2 Alternativtests und das Lemma von Neyman-Pearson*

Wir betrachten ein Standardmodell (vgl. Def. 6.2) mit jeweils einpunktiger Nullhypothese und Alternative, d.h. $\Theta = \{0, 1\}$, $\mathcal{X} \subset \mathbb{R}^n$ oder $\mathcal{X}$ diskret und P_i besitzt Dichte bzw. $\rho(x, i)$ auf $\mathcal{X}$ für $i = 0, 1$.

Setze

$$R(x) := \begin{cases} \dfrac{\rho(x,1)}{\rho(x,0)} & \text{wenn } \rho(x,0) > 0, \\ \infty & \text{sonst.} \end{cases}$$

R heißt der *Likelihood-Quotient*.

Ein Test von P_0 gegen P_1 (formal hier wörtlich $\Theta_0 = \{0\}$ gegen $\Theta_1 = \{1\}$) der Form

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

für ein $c \geq 0$ heißt ein Neyman-Pearson-Test[6]. (Im Fall $R(x) = c$ kann Randomisierung notwendig sein.)

[6] nach Jerzy Neyman, 1894–1981 und Egon Pearson, 1895–1980

Satz 6.30 (Neyman-Pearson-Lemma). *Betrachte Standardmodell mit einpunktiger Nullhypothese und einpunktiger Alternative,* $\alpha \in (0,1)$.

1. Es gibt einen Neyman-Pearson-Test φ *mit* $\mathbb{E}_0[\varphi] = \alpha$.

2. Sei φ *ein Neyman-Pearson-Test* φ *mit* $\mathbb{E}_0[\varphi] = \alpha$, $\widetilde{\varphi}$ *irgendein Test von* P_0 *gegen* P_1 *zum Niveau* α. *Dann gilt* $\mathbb{E}_1[\varphi] \geq \mathbb{E}_1[\widetilde{\varphi}]$, *d.h. die Macht von* φ *ist mindestens so groß wie die von* $\widetilde{\varphi}$.

Man sagt: φ *ist (in dieser Situation) ein* gleichmäßig bester *Test.*

Beweis. 1. Wähle c mit $P_0(R \geq c) \geq \alpha$ und $P_0(R \leq c) \geq 1 - \alpha$.

Falls $P_0(R = c) = 0$, so ist $\varphi(x) := \mathbf{1}_{\{R(x) > c\}}$ ein Neyman-Pearson-Test mit $\mathbb{E}_0[\varphi] = P_0(R > c) = P_0(R \geq c) = \alpha$.

Falls $P_0(R = c) > 0$, so setze $\gamma := \dfrac{\alpha - P_0(R > c)}{P_0(R = c)}$ $(\in [0,1])$ und

$$\varphi(x) := \begin{cases} 1 & \text{wenn } R(x) > c, \\ \gamma & \text{wenn } R(x) = c, \\ 0 & \text{wenn } R(x) < c. \end{cases}$$

Dies ist ein Neyman-Pearson-Test mit

$$\mathbb{E}_0[\varphi] = 1 \cdot P_0(R > c) + \gamma P_0(R = c) + 0 \cdot P_0(R < c) = \alpha$$

2. Sei φ ein Neyman-Pearson-Test (mit Schwellenwert c), d.h.

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

mit $\mathbb{E}_0[\varphi] = \alpha$, $\widetilde{\varphi}$ irgendein Test mit $\mathbb{E}_0[\widetilde{\varphi}] \leq \alpha$. Es gilt

$$\text{für alle } x \in \mathcal{X} : \quad \big(\varphi(x) - \widetilde{\varphi}(x)\big)\big(\rho(x,1) - c\rho(x,0)\big) \geq 0$$

denn die beiden Faktoren haben dasselbe Vorzeichen (sofern der zweite $\neq 0$), da $\varphi(x) \geq \mathbf{1}(\rho(x,1) > c\rho(x,0))$. Somit

$$f_1(x) := \big(\varphi(x) - \widetilde{\varphi}(x)\big)\rho(x,1) \geq c\big(\varphi(x) - \widetilde{\varphi}(x)\big)\rho(x,0) =: cf_0(x) \tag{6.2}$$

und folglich

$$\mathbb{E}_1[\varphi] - \mathbb{E}_1[\widetilde{\varphi}] = \int\limits_{\mathcal{X}} f_1(x)\,dx \geq c\int\limits_{\mathcal{X}} f_0(x)\,dx = c(\alpha - \mathbb{E}_0[\widetilde{\varphi}]) \geq 0. \tag{6.3}$$

(Wenn $\mathcal{X}$ diskret ist, muss das Integral natürlich durch eine Summe ersetzt werden.) □

Bemerkung. Wir sehen aus dem Beweis auch: Wenn $\widetilde{\varphi}$ ebenfalls ein gleichmäßig bester Test von P_0 gegen P_1 mit $\mathbb{E}_0[\widetilde{\varphi}] = \alpha$ ist, so gilt Gleichheit in (6.3) und daher auch Gleichheit in (6.2) (möglicherweise mit Ausnahme einer Menge von [Lebesgue-]Maß 0). In diesem Sinne ist also hier ein gleichmäßig bester Test „identisch“ mit einem Neyman-Pearson-Test.

Beispiel. Beobachtungen $X_1, \ldots, X_n$ seien unter P_0 u.i.v. mit $X_i \sim \mathcal{N}_{\mu_0,\sigma^2}$, unter P_1 u.i.v. mit $X_i \sim \mathcal{N}_{\mu_1,\sigma^2}$, wobei $\sigma^2 > 0$ und $\mu_0 < \mu_1$ bekannt (und fest) sind.

Mit $x = (x_1, \ldots, x_n)$ und $\overline{x} := \frac{1}{n}\sum_{i=1}^n x_i$ ist

$$\begin{aligned} R(x) &= \exp\Big(-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i - \mu_1)^2 + \frac{1}{2\sigma^2}\sum_{i=1}^n (x_i - \mu_0)^2\Big) \\ &= \exp\Big(-\frac{n}{2\sigma^2}\big(2(\mu_1 - \mu_0)\overline{x} + \mu_1^2 - \mu_0^2\big)\Big). \end{aligned}$$

Offenbar ist $R(x)$ eine montone (fallende) Funktion von $\overline{x}$ und unter P_0 ist $\overline{X} := \frac{1}{n}\sum_{i=1}^n X_i \sim \mathcal{N}_{\mu_0,\sigma^2/n}$, also ist

$$\varphi(x) := \mathbf{1}_{\{\overline{x} > c\}}$$

mit der Wahl $c := \mu_0 + \sqrt{\sigma^2/n}\,\Phi^{-1}(1-\alpha)$ ein Neyman-Pearson-Test zum Niveau $\alpha \in (0,1)$ (denn dann ist $\mathbb{E}_0[\varphi] = P_0\big(\overline{X} > c\big) = \alpha$).

6.4.3 Tests für kategorielle Beobachtungen (zum χ^2-Test)*

Ein Experiment mit s möglichen Ausgängen werde n mal (unabhängig) wiederholt, Ausgang i habe die (unbekannte) Wahrscheinlichkeit ϑ_i, $i = 1, \dots, s$.
Angenommen, wir beobachten h_i-mal Ausgang i für $i = 1, \dots, s$. Passt dies zur (Null-)Hypothese, dass

$$\vartheta = (\vartheta_1, \dots, \vartheta_s) = (\rho_1, \dots, \rho_s) = \rho$$

gilt für einen vorgegebenen Vektor ρ von Wahrscheinlichkeitsgewichten (auf $\{1, \dots, s\}$)?

Satz 6.31. *Sei* $\rho \in \Delta_s := \{(\vartheta_1, \dots, \vartheta_s) \in [0,1]^s : \vartheta_1 + \dots + \vartheta_s = 1\}$,

$$(H_1^{(n)}, \dots, H_s^{(n)}) \sim \mathrm{Mult}_{n;\rho_1,\dots,\rho_s},$$

dann gilt

$$\sum_{i=1}^{s} \frac{(H_i^{(n)} - n\rho_i)^2}{n\rho_i} \xrightarrow[n\to\infty]{d} \chi^2_{s-1}$$

Beweisskizze. Seien $X_1^{(n)}, \dots, X_s^{(n)}$ u.a., $X_i^{(n)} \sim \mathrm{Poi}_{n\rho_i}$, dann ist

$$N_n := X_1^{(n)} + \dots + X_s^{(n)} \sim \mathrm{Poi}_n$$

(siehe Bsp. 2.26, 3.)
Beachte: Für $m \in \mathbb{N}$, $h_1, \dots, h_s \in \mathbb{N}_0$ mit $h_1 + \dots + h_s = m$ ist

$$\begin{aligned} &P\big(X_1^{(n)} = h_1, \dots, X_s^{(n)} = h_s \,\big|\, N_n = m\big) \\ &= \Big(e^{-n}\frac{n^m}{m!}\Big)^{-1} \prod_{i=1}^{s} e^{-n\rho_i}\frac{(n\rho_i)^{h_i}}{h_i!} = \binom{m}{h_1, h_2, \dots, h_s} \rho_1^{h_1} \cdots \rho_s^{h_s}, \end{aligned}$$

d.h. bedingt auf $\{N_n = m\}$ ist $(X_1^{(n)}, \dots, X_s^{(n)}) \sim \mathrm{Mult}_{m;\rho_1,\dots,\rho_s}$.

Sei

$$\widetilde{X}_i^{(n)} := \frac{X_i^{(n)} - n\rho_i}{\sqrt{n\rho_i}}, \qquad \widetilde{H}_i^{(n)} := \frac{H_i^{(n)} - n\rho_i}{\sqrt{n\rho_i}}$$

(beachte: $\mathbb{E}[\widetilde{X}_i^{(n)}] = 0$, $\mathrm{Var}[\widetilde{X}_i^{(n)}] = 1$),

$$\widetilde{N}_n := \frac{N_n - n}{\sqrt{n}} = \sum_{i=1}^{s} \sqrt{\rho_i}\widetilde{X}_i^{(n)}$$

(beachte: $N_n = n \iff \widetilde{N}_n = 0 \iff (\widetilde{X}_1^{(n)}, \dots, \widetilde{X}_s^{(n)})^T \in \mathbb{H}_\rho$)

Der zentrale Grenzwertsatz (Satz 5.5), angewendet auf jede der (unabhängigen) Koordinaten, liefert

$$\widetilde{X}^{(n)} := (\widetilde{X}_1^{(n)}, \dots, \widetilde{X}_s^{(n)})^T \xrightarrow[n\to\infty]{d} \mathcal{N}_{0,1}^{\otimes s}$$

und somit gilt auch

$$\sum_{i=1}^{s} \left(\widetilde{X}_1^{(n)}\right)^2 \xrightarrow[n\to\infty]{d} \chi_s^2.$$

Das macht zumindest plausibel, dass auch

$$\mathscr{L}\left(\sum_{i=1}^{s} \left(\widetilde{X}_1^{(n)}\right)^2 \,\middle|\, \widetilde{N}_n = 0 \right) \xrightarrow[n\to\infty]{d} \chi_{s-1}^2$$

gilt.

Hier sind etwas mehr Details: Sei O eine orthogonale $s \times s$-Matrix, deren letzte Spalte u_ρ ist (ergänze u_ρ zur ONB),

$$\Pi_\rho = O \cdot \begin{pmatrix} 1 & & & 0 \\ & \ddots & & \\ & & 1 & \\ 0 & & & 0 \end{pmatrix} O^T$$

ist die (orthogonale) Projektion(smatrix) auf $\mathbb{H}_\rho$.
Sei $\widetilde{N}_\rho := \mathcal{N}_{0,1}^{\otimes s} \circ (\Pi_\rho)^{-1}$,
$a_1, \dots, a_s \in \mathbb{R}$, $A := (-\infty, a_1] \times \times(-\infty, a_s]$,

$$q_{m,n} := P\big(\widetilde{X}^{(n)} \in A \,\big|\, N_n = m\big)$$

Es gilt

$$q_{m,n} \geq q_{m+1,n} \quad \text{für } \ m, n \in \mathbb{N}$$

(verwende die „natürliche" Kopplung der Multinomialverteilungen und die spezielle Form von A), somit für $\varepsilon > 0$

$$P\big(\widetilde{X}^{(n)} \in A \,\big|\, \widetilde{N}_n \in [0, \varepsilon]\big) \leq P\big(\widetilde{H}^{(n)} \in A\big) \leq P\big(\widetilde{X}^{(n)} \in A \,\big|\, \widetilde{N}_n \in [-\varepsilon, 0]\big)$$

(denn

$$q_{n,n} \geq \sum_{m=n}^{\lceil n+\varepsilon\sqrt{n}\rceil} q_{m,n} P\big(N_n = m \,\big|\, \widetilde{N}_n \in [0, \varepsilon]\big) = P\big(\widetilde{X}^{(n)} \in A \,\big|\, \widetilde{N}_n \in [0, \varepsilon]\big)$$

und analog für die andere Schranke).

Setze $\widetilde{Y}^{(n)} := O^T \widetilde{X}^{(n)}$, $U_\varepsilon := \{(x_1, \dots, x_s)^T \in \mathbb{R}^s : 0 \le x_s \le \varepsilon\} = \mathbb{R}^{s-1} \times [0, \varepsilon]$

$$\begin{aligned}
P\Big(\widetilde{X}^{(n)} \in A \,\Big|\, \widetilde{N}_n \in [0,\varepsilon]\Big) &= P\Big(\widetilde{Y}^{(n)} \in O^T A \,\Big|\, \widetilde{Y}_n \in U_\varepsilon\Big) \\
&= \frac{P\big(\widetilde{Y}_n \in O^T A \cap U_\varepsilon\big)}{P\big(\widetilde{Y}_n \in U_\varepsilon\big)} \\
&\xrightarrow[n\to\infty]{} \frac{\mathcal{N}_{0,1}^{\otimes s}(O^T A \cap U_\varepsilon)}{\mathcal{N}_{0,1}^{\otimes s}(U_\varepsilon)} \\
&= \frac{1}{\mathcal{N}_{0,1}([0,\varepsilon])} \int_0^\varepsilon \frac{1}{\sqrt{2\pi}} e^{-t^2/2} \mathcal{N}_{0,1}^{\otimes(s-1)}\Big(\{x \in \mathbb{R}^{s-1} : (x,t) \in O^T A\}\Big)\, dt
\end{aligned}$$

denn mit zentralem Grenzwertsatz (Satz 5.5) und Rotationssymmetrie der s-dim. Normalverteilung (Beispiel 2.20) folgt $\widetilde{Y}^{(n)} \longrightarrow^d_{n\to\infty} \mathcal{N}_{0,1}^{\otimes s}$. Mit $\varepsilon \downarrow 0$ konvergiert die rechte Seite gegen $\widetilde{\mathcal{N}}_\rho(A)$. $\square$

Korollar 6.32 (χ^2-Anpassungstest[7]). *Sei* $\vartheta \in \Theta = \Delta_s := \{(\vartheta_1, \dots, \vartheta_s) \in [0,1]^s : \vartheta_1 + \dots + \vartheta_s = 1\}$, *unter* P_ϑ *sei* $(H_1, \dots, H_s) \sim \mathrm{Mult}_{n;\vartheta_1,\dots,\vartheta_s}$.

Sei $\rho \in \Delta_s$,

$$D := \sum_{i=1}^{s} \frac{(H_i - n\rho_i)^2}{n\rho_i},$$

$\alpha \in (0,1)$, *q das* $(1-\alpha)$*-Quantil der* χ^2_{s-1}*-Verteilung.*

Der Test von $H_0 : \{\vartheta = \rho\}$ *gegen* $H_1 : \{\vartheta \neq \rho\}$ *mit Ablehnungsbereich* $\{D > q\}$ *hat (asymptotisches) Niveau* α.

Dies folgt aus Satz 6.31. Satz 6.31 macht allerdings keine Aussage darüber, wie groß n sein sollte, damit die Approximation plausibel ist. Eine oft zitierte Faustregel (für die Gültigkeit der χ^2-Approximation) ist $n\rho_i \geq 5$ für alle i.

[7] von Karl Pearson (1857–1936) im Jahr 1900 vorgeschlagen

Beispiel 6.33 (Mendels Erbsenexperimente[8]). Betrachte zwei Merkmale: Farbe: grün (rezessiv) vs. gelb (dominant), Form: rund (dominant) vs. runzlig (rezessiv)
Beim Kreuzen von Doppelhybriden erwarten wir folgende Phänotypwahrscheinlichkeiten unter Mendel'scher Segregation („rund" und „gelb" sind jeweils dominant, $n = 556$ Versuche):

Typ	rund/gelb	rund/grün	kantig/gelb	rund/gelb
Anteil	$9/16$	$3/16$	$3/16$	$1/16$
Erwartete Anzahl	315	104,25	104,25	34,75
beobachtet	315	108	101	32

Wir finden $D \approxeq 0{,}47$, $\chi^2_3([0, 0.47]) \approxeq 0{,}075$, ein χ^2-Test zum 1%-Niveau lehnt H_0 nicht ab (und auch zu „nahezu egal welchem Niveau" nicht). Insoweit passen die Daten sehr gut zu den theoretischen Häufigkeiten.

[8]Gregor Mendel, 1822–1884; G. Mendel, Versuche über Pflanzenhybriden, Verhandlungen des naturforschenden Vereines in Brünn, Bd. IV für das Jahr 1865, Abhandlungen: 3–47, (1866).

Beispiel. Wir vermuten, dass ein gegebener sechsseitiger Würfel unfair ist und möchten dies auf dem 5%-Niveau testen. Bei 120-maligem Würfeln finden wir folgende Häufigkeiten:

i	1	2	3	4	5	6
h_i	13	12	20	18	26	31

Es ist $D \approxeq 13{,}7$, das 95%-Quantil der χ^2_5-Verteilung ist $\approxeq 11{,}07$, wir können die Nullhypothese „$\vartheta = (1/6, 1/6, 1/6, 1/6, 1/6, 1/6)$“ also auf dem 5%-Niveau ablehnen.

```
> w <- c(13,12,20,18,26,31)
> chisq.test(w,p=c(1/6,1/6,1/6,1/6,1/6,1/6))

Chi-squared test for given probabilities

data:  w
X-squared = 13.7, df = 5, p-value = 0.01763
```

$$D := \frac{(13-20)^2}{20} + \frac{(12-20)^2}{20} + \dots + \frac{(31-20)^2}{20}$$

Exkurs: χ^2-Test auf Homogenität (auch: „auf Unabhängigkeit")

In einem Experiment werden zwei „Merkmale" beobachtet, wobei das erste Merkmal a und das zweite Merkmal b viele Ausprägungen besitzt (also insgesamt $s = a \cdot b$ mögliche Ausgänge).

Unter n u.a. Wiederholungen werde h_{ij} mal Ausgang (i, j) beobachtet ($i \in \{1, 2, \ldots, a\}$, $j \in \{1, 2, \ldots, b\}$), man fasst die Beobachtungen in einer $a \times b$-Kontingenztafel zusammen:

i \ j	1	2	3	
1	h_{11}	h_{12}	h_{13}	$h_{1\cdot}$
2	h_{21}	h_{22}	h_{23}	$h_{2\cdot}$
	$h_{\cdot 1}$	$h_{\cdot 2}$	$h_{\cdot 3}$	$h_{\cdot\cdot} = n$

mit Zeilensummen $h_{i\cdot} = \sum_{j=1}^{b} h_{ij}$, Spaltensummen $h_{\cdot j} = \sum_{i=1}^{a} h_{ij}$ und Gesamtsumme $h_{\cdot\cdot} = \sum_{i=1}^{a} \sum_{j=1}^{b} h_{ij} = n$.

Wir fassen die beobachteten Häufigkeiten als Realisierungen einer

$$\text{multinomial}(n, (\vartheta_{ij})_{i=1,\ldots,a;j=1,\ldots,b})\text{-verteilten ZV } (H_{ij})_{i=1,\ldots,a;j=1,\ldots,b}$$

auf, wobei

$(\vartheta_{ij})_{i=1,\ldots,a;j=1,\ldots,b}$ ein $a \cdot b$-dimensionaler Vektor von Wahrscheinlichkeitsgewichten ist.

Passen die Beobachtungen zur Nullhypothese, dass

$$\vartheta_{ij} = \eta_i \cdot \rho_j, \quad \text{für } i = 1, \ldots, a, \ j = 1, \ldots, b$$

mit $(\eta_i)_{i=1,\ldots,a}$, $(\rho_j)_{j=1,\ldots,b}$ gewissen a- bzw. b-dimensionalen Vektoren von Wahrscheinlichkeitsgewichten?

Beobachtungen:

$i \backslash j$	1	2	3	
1	h_{11}	h_{12}	h_{13}	$h_{1\cdot}$
2	h_{21}	h_{22}	h_{23}	$h_{2\cdot}$
	$h_{\cdot 1}$	$h_{\cdot 2}$	$h_{\cdot 3}$	$h_{\cdot\cdot} = n$

Wir bilden

$$\widehat{\vartheta}_{i\cdot} = \frac{H_{i\cdot}}{n}, \qquad \widehat{\vartheta}_{\cdot j} = \frac{H_{\cdot j}}{n}$$

und die Teststatistik

$$D = \sum_{i=1}^{a} \sum_{j=1}^{b} \frac{(H_{ij} - n\widehat{\vartheta}_{i\cdot}\widehat{\vartheta}_{\cdot j})^2}{n\widehat{\vartheta}_{i\cdot}\widehat{\vartheta}_{\cdot j}}$$

Bericht 6.34. Unter H_0 : „$(\vartheta_{ij})_{i=1,\ldots,a;j=1,\ldots,b}$ hat Produktform“ ist D (approximativ) $\chi^2_{(a-1)(b-1)}$-verteilt.

Wir würden also H_0 zum Niveau α ablehnen, falls der beobachtete Wert größer ist als das $(1-\alpha)$-Quantil der χ^2-Verteilung mit $(a-1)(b-1)$ Freiheitsgraden.

Das Simpson-Paradoxon

Durch Zusammenfassen von Gruppen können sich (scheinbare) statistische Trends in ihr Gegenteil verkehren. Dieses Phänomen heißt Simpson-Paradoxon oder Yule-Simpson-Effekt[9].

Beispiel (Zulassungsstatistik der UC Berkeley 1973)**.** Im Herbst 1973 haben sich an der Universität Berkeley 12763 Kandidaten für ein Studium beworben, davon 8442 Männer und 4321 Frauen. Es kam zu folgenden Zulassungszahlen:

	Aufgenommen	Abgelehnt
Männer	3738	4704
Frauen	1494	2827

Demnach betrug die Zulassungsquote
bei den Männern $\frac{3738}{8442} \approx 44\%$, bei den Frauen nur $\frac{1494}{4321} \approx 35\%$.
Ein χ^2-Test auf Homogenität (z.B. mit R) zeigt, dass eine solche Unverhältnismäßigkeit nur mit verschwindend kleiner Wahrscheinlichkeit durch „reinen Zufall" entsteht:

```
> berkeley <- matrix(c(3738,1494,4704,2827),nrow=2)
> berkeley
     [,1] [,2]
[1,] 3738 4704
[2,] 1494 2827
> chisq.test(berkeley,correct=FALSE)

Pearson's Chi-squared test

data:  berkeley
X-squared = 111.2497, df = 1, p-value < 2.2e-16
```

[9] nach Edward H. Simpson, *1922 und George Udny Yule, 1871–1951

Dieser Fall hat einiges Aufsehen erregt, s.a. P.J. Bickel, E.A. Hammel, J.W. O'Connell, Sex Bias in Graduate Admissions: Data from Berkeley, *Science* 187, no. 4175, 398–404, (1975).

Das Ungleichgewicht verschwindet, wenn man die Zulassungszahlen nach Departments aufspaltet:
Es stellt sich heraus, dass innerhalb der Departments die Aufnahmewahrscheinlichkeiten nicht signifikant vom Geschlecht abhängen, aber sich Frauen häufiger bei Departments mit (absolut) niedriger Aufnahmequote beworben haben als Männer – dies ist ein Beispiel für das *Simpson-Paradox*.

Die genauen nach Departments aufgeschlüsselten Bewerber- und Zulassungszahlen sind leider nicht öffentlich zugänglich (siehe aber Abb. 1 in Bickel et. al, loc. cit., für eine grafische Aufbereitung der Daten, die den Simpson-Effekt zeigt).
Bickel et. al demonstrieren das Phänomen mittels eines hypothetischen Beispiels:

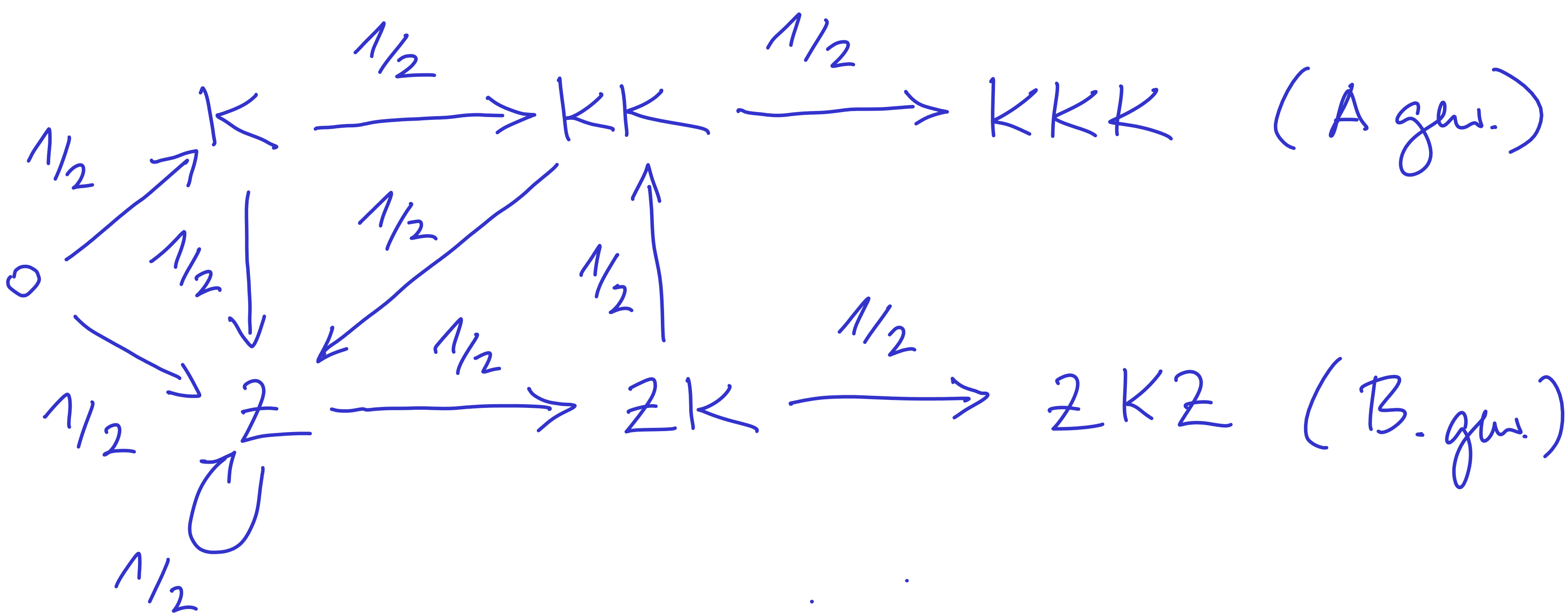
	Aufgenommen	Abgelehnt
	Department of machismathics	
Männer	200	200
Frauen	100	100
	Department of social warfare	
Männer	50	100
Frauen	150	300
	Gesamt	
Männer	250	300
Frauen	250	400

Kapitel 7

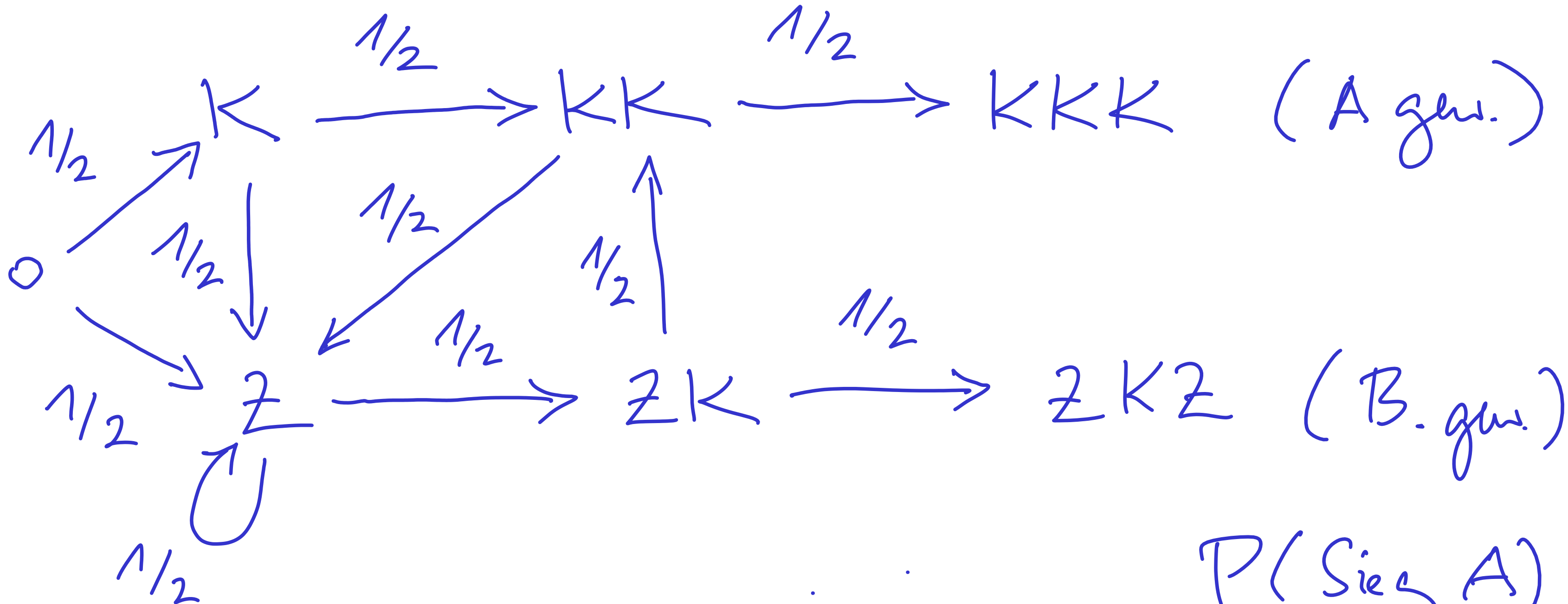
Markovketten

Beispiel 7.1. Eine faire Münze wird solange geworfen, bis entweder das Muster KKK ($\hat{=}$ Sieg von Spieler A) oder das Muster ZKZ ($\hat{=}$ Sieg von Spieler B) gefallen ist.

$$P(\text{Sieg A}) = ?$$



Sei $w(x) = P(\text{A gewinnt von Zustand } x \text{ aus})$, zerlege nach dem ersten Schritt:



$$P(\text{Sieg A}) = \tfrac{1}{2} w(K) + \tfrac{1}{2} w(Z)$$

$$w(KKK) = 1, \quad w(ZKZ) = 0$$

$$\boxed{\begin{aligned} w(KK) &= \tfrac{1}{2} w(Z) + \tfrac{1}{2} \cdot 1 \\ w(ZK) &= \tfrac{1}{2} w(KK) + \tfrac{1}{2} \cdot 0 \\ w(K) &= \tfrac{1}{2} w(KK) + \tfrac{1}{2} w(Z) \\ w(Z) &= \tfrac{1}{2} w(Z) + \tfrac{1}{2} w(ZK) \end{aligned}}$$

$w(KKK) = 1, \quad w(ZKZ) = 0$

$$\boxed{\begin{aligned}
&(1) \quad & w(KK) &= \tfrac{1}{2} w(Z) + \tfrac{1}{2} \cdot 1 \\
&(2) \quad & w(ZK) &= \tfrac{1}{2} w(KK) + \tfrac{1}{2} \cdot 0 \\
&(3) \quad & w(K) &= \tfrac{1}{2} w(KK) + \tfrac{1}{2} w(Z) \\
&(4) \quad & w(Z) &= \tfrac{1}{2} w(Z) + \tfrac{1}{2} w(ZK)
\end{aligned}}$$

$(4) \Leftrightarrow \quad \frac{1}{2} w(Z) = \frac{1}{2} w(ZK) \quad \Rightarrow \quad w(Z) = w(ZK)$

mit (2) $\Rightarrow \quad w(KK) = 2 w(ZK) = 2 w(Z)$

mit (1) $\Rightarrow \quad 2 w(Z) = \frac{1}{2} w(Z) + \frac{1}{2}$

$\Rightarrow \quad w(Z) = 1/3$

$\frac{5}{12}$

mit (3) $\Rightarrow \quad w(K) = \frac{1}{2} \cdot 2 \cdot \frac{1}{3} + \frac{1}{2} \cdot \frac{1}{3} = \frac{1}{2}$

$P(A \text{ gew.}) = \frac{1}{2} \cdot w(K) + \frac{1}{2} \cdot w(Z) = \frac{1}{2}\left(\frac{1}{3} + \frac{1}{2}\right)$

Definition 7.2. Sei S abzählbare Menge ($S \neq \emptyset$).

1. $A = (a_{x,y})_{x,y\in S}$ heißt eine *stochastische Matrix* (über S), wenn gilt

$$a_{x,y} \geq 0 \quad \text{für alle } x, y \in S \qquad \text{und} \qquad \sum_{y\in S} a_{x,y} = 1 \quad \text{für alle } x \in S.$$

2. Eine Folge $X = (X_n)_{n\in\mathbb{N}_0}$ von Zufallsvariablen mit Werten in S heißt eine *Markovkette*[1] (mit Zustandsraum S und Übergangsmatrix A), wenn

$$P(X_{n+1} = y \,|\, X_0 = x_0, \dots, X_{n-1} = x_{n-1}, X_n = x) = P(X_{n+1} = y \,|\, X_n = x) = a_{x,y}$$

   für alle $n \in \mathbb{N}, x_0, \dots, x_{n-1}, x, y \in S$ mit $P(X_0 = x_0, \dots, X_n = x) > 0$ gilt.

   Die Verteilung von X_0 heißt die *Startverteilung* (von X).

Die entscheidende Eigenschaft ist, dass der (bzw. die Verteilung des) „neue" Zustand X_{n+1} nur vom direkt vorhergehenden X_n abhängt, nicht von der „gesamten Vorgeschichte" $X_0, X_1, \dots, X_n$ – dies nennt man auch die „Gedächtnislosigkeit" einer Markovkette (s.a. Beob. 7.3, 3. unten).

[1] zu Ehren von Andrei Andreyevich Markov, 1852–1922 benannt