

Zusammenfassung Stochastik 3

© Tim Baumann, <http://timbaumann.info/uni-spicker>

Hypothesentests mittels Stichprobenfktn

Modell. Gegeben sei ein parametrisches Modell, d.h. eine Zufallsgröße X , deren Verteilungsfunktion $P_X \in \{P_\vartheta \mid \vartheta \in \Theta \subset \mathbb{R}^n\}$ von einem Parameter ϑ abhängt.

Problem. Anhand einer **Stichprobe** $x_1, \dots, x_n \in \mathbb{R}^1$ von X (d.h. x_1, \dots, x_n sind Realisierung von iid ZGen $X_1, \dots, X_n \sim P_X$) ist zu entscheiden, ob die sogenannte **Nullhypothese** $H_0 : \vartheta \in \Theta_0 \subset \Theta$ oder eine **Gegenhypothese** $H_1 : \vartheta \in \Theta_1 = \Theta \setminus \Theta_0$ angenommen oder abgelehnt werden soll.

Def. Der **Stichprobenraum** ist $(\mathbb{R}^n, \mathfrak{B}(\mathbb{R}^n), P_\vartheta \times \dots \times P_\vartheta)$.

Terminologie. Die Hypothese H_i heißt **einfach**, falls $|\Theta_i| = 1$, andernfalls **zusammengesetzt**.

Def. Ein (nichtrandomisierter) **Test** für H_0 gegen H_1 ist eine Entscheidungsregel über die Annahme von H_0 basierend auf einer Stichprobe, die durch eine messbare Abbildung $\varphi : \mathbb{R}^n \rightarrow \{0, 1\}$ ausgedrückt wird und zwar durch

$$\varphi(x_1, \dots, x_n) = \begin{cases} 0 & \text{bei Annahme von } H_0, \\ 1 & \text{bei Ablehnung von } H_0. \end{cases}$$

Def. Der **Ablehnungsbereich** oder **kritische Bereich** von φ ist

$$K_n := \{(x_1, \dots, x_n) \in \mathbb{R}^n \mid \varphi(x_1, \dots, x_n) = 1\}.$$

Bem. Es gilt $\varphi = \mathbb{1}_{K_n}$.

Def. **Fehler 1. Art:** Ablehnung von H_0 , obwohl H_0 richtig ist
Fehler 2. Art: Annahme von H_0 , obwohl H_0 falsch ist

Def. Die **Güte- oder Machtfunktion** des Tests φ ist

$$\begin{aligned} m_\varphi : \Theta \rightarrow [0, 1], \quad m_\varphi(\vartheta) &:= \mathbb{E}_\vartheta \varphi(X_1, \dots, X_n) \\ &= \mathbb{P}_\vartheta((X_1, \dots, X_n) \in K_n) \\ &= (P_\vartheta \times \dots \times P_\vartheta)(K_n) \end{aligned}$$

Die Gegenwsk. $(1 - m_\varphi(\vartheta))$ heißt **Operationscharakteristik** von φ .

Bem. Es gilt $\mathbb{P}_\vartheta(\text{Fehler 1. Art}) = m_\varphi(\vartheta)$ für $\vartheta \in \Theta_0$,
 $\mathbb{P}_\vartheta(\text{Fehler 2. Art}) = 1 - m_\varphi(\vartheta)$ für $\vartheta \in \Theta_1$.

Def. Ein Test $\varphi : \mathbb{R}^n \rightarrow \{0, 1\}$ mit

$$\sup_{\vartheta \in \Theta_0} m_\varphi(\vartheta) \leq \alpha$$

heißt **α -Test** o. **Signifikanztest** zum **Signifikanzniveau** $\alpha \in (0, 1)$.

Ein α -Test φ heißt **unverfälscht** (erwartungstreu, unbiased), falls

$$\inf_{\vartheta \in \Theta_1} m_\varphi(\vartheta) \geq \alpha.$$

Situation. Sei nun eine Stichprobenfunktion oder **Teststatistik** $T : \mathbb{R}^n \rightarrow \mathbb{R}^1$ gegeben. Wir wollen einen Test der einfachen Nullhypothese $H_0 : \vartheta \in \Theta_0 = \{\vartheta_0\}$ entwickeln.

Def. $K_n^T \subset \mathbb{R}^1$ heißt **kritischer Bereich der Teststatistik**, falls

$$K_n = T^{-1}(K_n^T).$$

Bem. Es gilt

$$\begin{aligned} m_\varphi(\vartheta_0) &= \mathbb{P}_{\vartheta_0}((X_1, \dots, X_n) \in K_n) = \\ &= \mathbb{P}_{\vartheta_0}(T(X_1, \dots, X_n) \in K_n^T) = \int_{K_n^T} f_T(x) dx, \end{aligned}$$

wobei f_T die Dichte von $T(X_1, \dots, X_n)$ unter H_0 ist.

Test. Sei $X \sim \mathcal{N}(\mu, \sigma^2)$, σ bekannt und $\alpha \in (0, 1)$ vorgegeben. Zum Test von $H_0 : \mu = \mu_0$ **vs.** $H_1 : \mu \neq \mu_0$ wählen wir als Statistik

$$T(X_1, \dots, X_n) := \frac{\sqrt{n}}{\sigma} (\bar{X}_n - \mu_0) \quad \text{mit} \quad \bar{X}_n := \frac{1}{n} (X_1 + \dots + X_n).$$

Unter Annahme von H_0 gilt $T(X_1, \dots, X_n) \sim \mathcal{N}(0, 1)$.

Der Ablehnungsbereich der Statistik ist

$$K_n^T = \{t \in \mathbb{R}^1 \mid |t| > z_{1-\alpha/2}\} \quad \text{mit} \quad z_{1-\alpha/2} := \Phi^{-1}(1 - \alpha/2).$$

Für $\alpha = 0,05$ gilt beispielsweise $z_{1-\alpha/2} \approx 1,96$.

Bem. Es gilt

$$\begin{aligned} t \in (K_n^T)^c &\iff |t| \leq z_{1-\alpha/2} \iff |\bar{X}_n - \mu_0| \leq \frac{\sigma}{\sqrt{n}} z_{1-\alpha/2} \\ &\iff \mu_0 \in [\bar{X}_n - \frac{\sigma}{\sqrt{n}} z_{1-\alpha/2}, \bar{X}_n + \frac{\sigma}{\sqrt{n}} z_{1-\alpha/2}]. \end{aligned}$$

Def. Dieses Intervall heißt **Konfidenzintervall** für μ_0 zum Konfidenzniveau $1 - \alpha$.

Test. Sei wieder $X \sim \mathcal{N}(\mu, \sigma^2)$, σ^2 aber diesmal unbekannt. Zum Testen von $H_0 : \mu = \mu_0$ **vs.** $H_1 : \mu \neq \mu_0$ verwenden wir

$$\hat{T}(X_1, \dots, X_n) = \frac{\sqrt{n}}{S_n} (\bar{X}_n - \mu_0), \quad S_n^2 := \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2.$$

Dabei ist S_n die (**korrigierte**) **Stichprobenvarianz**. Man kann zeigen, dass $\hat{T}(X_1, \dots, X_n) \sim t_{n-1}$ unter H_0 . Dabei ist t_m die **Student'sche t -Verteilung** mit m Freiheitsgraden (siehe unten). Der Ablehnungsbereich ist

$$K_n^T = \{t \in \mathbb{R}^1 \mid |t| > t_{n-1, 1-\alpha/2}\}.$$

Bem. S_n^2 und \bar{X}_n sind unabhängig für $n \geq 2$.

Diskussion. • Je kleiner α ist, desto „nullhypotesenfreundlicher“ ist der Test. Häufig verwendet wird $\alpha \in \{10\%, 5\%, 1\%, 0,5\%\}$.

• Einseitige Tests: Die Gegenhypothese zu $H_0 : \mu = \mu_0$ ist $H_1 : \mu > \mu_0$. Die Nullhypothese wird nur abgelehnt, falls zu große Stichprobenmittelwerte \bar{x}_n vorliegen. Es ist dann $K_n^T = (z_{1-\alpha}, \infty)$.

Def. Es seien $X_1, \dots, X_n \sim \mathcal{N}(0, 1)$. Dann heißt die Summe $X_1^2 + \dots + X_n^2 \sim \chi_n^2$ **Chi-Quadrat-verteilt** mit n Freiheitsgraden.

Def. Falls $X \sim \mathcal{N}(0, 1)$ und $Y_n \sim \chi_n^2$ unabhängig sind, so heißt

$$\frac{X}{\sqrt{\frac{Y_n}{n}}} \sim t_n$$

t -verteilt mit n -Freiheitsgraden.

Lem. $\frac{n-1}{\sigma^2} S_n^2 \sim \chi_{n-1}^2$

Kor. \hat{T} aus dem zweiten obigen Bsp ist tatsächlich t -verteilt.

Def. Seien $Y_{n_i} \sim \chi_{n_i}^2$, $i = 1, 2$ zwei unabhängige ZGen. Dann heißt

$$\frac{Y_{n_1}/n_1}{Y_{n_2}/n_2} \sim F_{n_1, n_2}$$

F-verteilt (wie Fisher) mit (n_1, n_2) Freiheitsgraden.

Test. Sei $X \sim \mathcal{N}(\mu, \sigma^2)$ mit μ unbekannt. Wir testen $H_0 : \sigma = \sigma_0$ **vs.** $H_1 : \sigma \neq \sigma_0$ mit

$$T := \frac{n-1}{\sigma_0^2} S_n^2$$

Unter Annahme von H_0 gilt $T \sim \chi_{n-1}^2$. Falls μ bekannt ist, muss

$$\tilde{T} := \frac{n}{\sigma_0^2} \tilde{S}_n^2, \quad \tilde{S}_n^2 := \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2.$$

als Statistik gewählt werden. Unter Annahme von H_0 ist $\tilde{T} \sim \chi_n^2$.

Test. Seien Stichproben $X_1^{(i)}, \dots, X_{n_i}^{(i)} \sim \mathcal{N}(\mu_i, \sigma_i^2)$, $i = 1, 2$ gegeben. Wir testen $H_0 : \sigma_1 = \sigma_2$ **vs.** $H_1 : \sigma_1 \neq \sigma_2$. Dazu verwenden wir

$$T = \frac{S_{X^{(1)}}^2}{S_{X^{(2)}}^2}, \quad S_{X^{(j)}}^2 := \frac{1}{n-1} \sum_{i=1}^{n_j} (X_i^{(j)} - \bar{X}_n^{(j)})^2.$$

Falls H_0 gilt, so ist $T \sim F_{n_1-1, n_2-1}$.

Test. Situation wie im letzten Test mit $\sigma_1 = \sigma_2$.

Wir testen $H_0 : \mu_1 = \mu_2$ **vs.** $H_1 : \mu_1 \neq \mu_2$ mit

$$T = \sqrt{\frac{n_1 \cdot n_2}{n_1 + n_2}} \cdot \frac{\bar{X}_{n_1}^{(1)} - \bar{X}_{n_2}^{(2)}}{S_{n_1, n_2}}, \quad S_{n_1, n_2}^2 = \frac{(n_1-1)S_{X^{(1)}}^2 + (n_2-1)S_{X^{(2)}}^2}{n_1 + n_2 - 2}$$

Unter H_0 gilt $T \sim t_{n_1+n_2-2}$.

Test. Seien $(\begin{smallmatrix} X_1 \\ Y_1 \end{smallmatrix}), \dots, (\begin{smallmatrix} X_n \\ Y_n \end{smallmatrix}) \sim \mathcal{N}\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \sigma_1 \sigma_2 \rho \\ \sigma_1 \sigma_2 \rho & \sigma_2^2 \end{pmatrix}\right)$.

Wir testen $H_0 : \rho = 0$ **vs.** $H_1 : \rho \neq 0$ mit

$$T := \frac{\sqrt{n-2} \cdot \hat{\rho}_n}{\sqrt{1 - \hat{\rho}_n^2}}, \quad \hat{\rho}_n := \frac{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)(Y_i - \bar{Y}_n)}{S_{X,n} \cdot S_{Y,n}}.$$

Falls H_0 richtig ist, so gilt $T \sim t_{n-2}$.

Um $H_0 : \rho = \rho_0 \in (-1, 1)$ **vs.** $H_1 : \rho \neq \rho_0$ zu testen, kann man

$$T = \frac{\sqrt{n-3}}{2} \left(\log \frac{1 + \hat{\rho}_n}{1 - \hat{\rho}_n} - \log \frac{1 + \rho_0}{1 - \rho_0} \right)$$

verwenden. Für n groß gilt approx. $T \sim \mathcal{N}(0, 1)$ unter H_0 .

Lem (Slutzky). Seien $(X_n), (Y_n)$ Folgen von ZGn über $(\Omega, \mathfrak{A}, \mathbb{P})$ mit $X_n \xrightarrow[n \rightarrow \infty]{\mathbb{P}} c = \text{const}$ (d. h. $\forall \epsilon > 0 : \mathbb{P}(|X_n - c| > \epsilon) \rightarrow 0$) und $Y_n \xrightarrow[n \rightarrow \infty]{d} Y$ (d. h. $\mathbb{P}(Y_n \leq y) \rightarrow \mathbb{P}(Y \leq y)$ für alle Stetigkeitspunkte y der VF $y \mapsto \mathbb{P}(Y \leq y)$). Dann gilt:

$$X_n + Y_n \xrightarrow{d} c + Y, \quad X_n \cdot Y_n \xrightarrow{d} c \cdot Y, \quad Y_n / X_n \xrightarrow{d} Y / c \quad (\text{falls } c \neq 0)$$

und allgemeiner $f(X_n, Y_n) \xrightarrow[n \rightarrow \infty]{d} f(c, Y)$ für jede Fkt $f \in \mathcal{C}(\mathbb{R}^2, \mathbb{R})$.

Bem. Unabhängigkeit von (X_n) und (Y_n) wird nicht vorausgesetzt!

Situation. Sei $T_n = T(X_1, \dots, X_n)$ eine Statistik. Falls der ZGWS für T_n die Form

$$\sqrt{n}(T_n - \vartheta) \xrightarrow[n \rightarrow \infty]{d} \mathcal{N}(0, g(\vartheta))$$

besitzt, so benötigen wir für Hypothesentests eine Möglichkeit, die Abhängigkeit der Varianz vom Parameter ϑ zu beseitigen. Man sagt, man führt eine **varianzstabilisierende Transformation** durch. Wir suchen dazu eine stetig diff'bare Funktion $f : \Theta \rightarrow \mathbb{R}^1$, sodass

$$\sqrt{n}(f(T_n) - f(\vartheta)) \xrightarrow[n \rightarrow \infty]{d} \mathcal{N}(0, 1).$$

Man zeigt mit dem MWS und Slutzky, dass dafür gelten muss:

$$f'(\vartheta) = \frac{1}{\sqrt{g(\vartheta)}}, \quad \text{also} \quad f(\theta) = \int \frac{d\vartheta}{\sqrt{g(\vartheta)}}.$$

Bspe. • Sei $X \sim \text{Exp}(\mu)$ (also $\mathbb{E}X = \mu^{-1}$). Dann gilt

$$\sqrt{n}(\bar{X}_n - \frac{1}{\mu}) \xrightarrow[n \rightarrow \infty]{d} \mathcal{N}(0, g(\frac{1}{\mu})) \quad \text{mit} \quad g(\vartheta) := \vartheta^2.$$

$$\rightsquigarrow \text{Mit } f(\theta) := \int \frac{d\vartheta}{\sqrt{g(\vartheta)}} = \int \frac{d\vartheta}{\vartheta} = \log \theta$$

$$\text{gilt } \sqrt{n}(\log(\bar{X}_n) - \log(\frac{1}{\mu})) \xrightarrow[n \rightarrow \infty]{d} \mathcal{N}(0, 1).$$

- Wir wollen eine unbek. Wahrscheinlichkeit p schätzen, etwa durch Wurf einer Münze. Der ZGWS von de-Moivre-Laplace besagt

$$\sqrt{n}(\hat{p}_n - p) \xrightarrow[n \rightarrow \infty]{d} \mathcal{N}(0, p(1-p)),$$

wobei \hat{p}_n die relative Häufigkeit ist. Zur Stabilisierung der Varianz verwenden wir nun

$$f(\theta) := \int_0^\theta \frac{dp}{\sqrt{p(1-p)}} = 2 \arcsin(\sqrt{\theta}).$$

Def. Die k -dim **(Gaußsche) Normalverteilung** $\mathcal{N}_k(m, C)$ mit EW $m \in \mathbb{R}^k$ und einer nichtnegativ-definiten, symmetrischen Kovarianzmatrix $C \in \mathbb{R}^{k \times k}$ ist gegeben durch die Dichte

$$f_{\mathcal{N}_k(m, C)}(x) := \frac{1}{(2\pi)^{k/2} \sqrt{\det(C)}} \exp\left(-\frac{1}{2}(x - m)C^{-1}(x - m)^T\right).$$

Bem. Bei $k = 2$ schreibt man oft

$$C = \begin{pmatrix} \sigma_1^2 & \sigma_1 \sigma_2 \rho \\ \sigma_1 \sigma_2 \rho & \sigma_2^2 \end{pmatrix} \quad \text{mit} \quad \rho := \text{Cor}(X_1, X_2).$$

Def. Die **charakteristische Fkt** eines ZV $X = (X_1, \dots, X_k)^T$ ist

$$\varphi : \mathbb{R}^k \rightarrow \mathbb{R}, \quad t \mapsto \mathbb{E}e^{i\langle t, X \rangle} = \int_{\mathbb{R}^k} e^{i(t_1 x_1 + \dots + t_k x_k)} dF_X(x_1, \dots, x_k).$$

Bem. Die charakteristische Funktion von $\mathcal{N}_k(m, C)$ ist

$$\varphi_{\mathcal{N}_k(m, C)}(t) = \exp\left(i \sum_{i=1}^k t_i m_i - \frac{1}{2} \sum_{i,j=1}^k t_i c_{ij} t_j\right).$$

Satz. Für $A \in \mathbb{R}^{k \times l}$ gilt $\mathcal{N}_k(m, C) \cdot A = \mathcal{N}_l(m \cdot A, A^T C A)$.

Chi-Quadrat-Anpassungstest

Aufgabe. Prüfe, ob eine vorliegende Stichprobe x_1, \dots, x_n aus einer bestimmten (stetig oder diskret verteilten) Grundgesamtheit gezogen wurde. Wir testen also **$H_0 : F = F_0$ vs. $H_1 : F \neq F_0$** .

Verfahren. Wir teilen zunächst \mathbb{R} in Klassen ein,

$$\mathbb{R} = \bigcup_{i=1}^{s+1} I_j \quad \text{mit} \quad I_j := (y_{j-1}, y_j], \quad \text{wobei} \\ -\infty = y_0 < y_1 < \dots < y_s < y_{s+1} = +\infty.$$

Wir setzen

$$h_{n_j} := |\{k \in \{1, \dots, n\} \mid X_k \in I_j\}| \quad (\text{absolute Klassenhäufigkeit}) \\ p_j^{(0)} := \mathbb{P}(X \in I_j) = F_0(y_j) - F_0(y_{j-1}) \quad (\text{Klassenwktn unter } H_0)$$

Die Klassenhäufigkeiten sind multinomialverteilt unter H_0 :

$$\mathbb{P}(h_{n_1} = n_1, \dots, h_{n_{s+1}} = n_{s+1}) = \binom{n}{n_1, \dots, n_{s+1}} (p_1^{(0)})^{n_1} \dots (p_{s+1}^{(0)})^{n_{s+1}}.$$

Als (näherungsweise) Maß für die Abweichung einer empirischen Verteilung von F_0 bei gegebener Klasseneinteilung dient

$$T_{n,s+1} := \sum_{j=1}^{s+1} \frac{(h_{n_j} - np_j^{(0)})^2}{np_j^{(0)}}.$$

Satz. $T_{n,s+1} \xrightarrow[n \rightarrow \infty]{d} \chi_s^2$

Faustregel. Für $np_j^{(0)} \geq 5$, $j = 1, \dots, s+1$ ist $T_{n,s+1}$ mit guter Näherung χ_s^2 -verteilt.

Entscheidungsregel (χ^2 -Anpassungstest). Die Nullhypothese $H_0 : F = F_0$ wird genau dann verworfen, wenn $T_{n,s+1} > \chi_{s,1-\alpha}^2$.

Bemn. • $T_{n,s+1}$ misst eigentlich nicht die Abweichung von der VF F_0 , sondern von der Multinomialverteilung $\mathcal{M}(n, p^{(0)})$.
• Der χ^2 -Anpassungstest gilt als hypothesenfreundlich.
• Es ist üblich, zunächst die Parameter $\vartheta = (\vartheta_1, \dots, \vartheta_r)$ der VF F_0 durch MLE zu schätzen, also durch

$$\hat{\vartheta}_n := \arg \max L(h_{n_1}, \dots, h_{n_{s+1}}; \vartheta), \quad \text{wobei} \\ L(h_{n_1}, \dots, h_{n_{s+1}}; \vartheta) := \prod_{j=1}^{s+1} (p_j^{(0)})^{h_{n_j}}.$$

Es kann (unter „natürlichen“ Bedingungen) gezeigt werden, dass

$$T_{n,s+1}(\hat{\vartheta}_n) = \sum_{j=1}^{s+1} \frac{(h_{n_j} - np_j^{(0)}(\hat{\vartheta}_n))^2}{np_j^{(0)}(\hat{\vartheta}_n)} \xrightarrow[n \rightarrow \infty]{d} \chi_{s-r}^2,$$

wobei r die Anzahl der geschätzten Parameter ist.

- Manchmal wird die Parameter-Schätzung auch direkt aus der SP x_1, \dots, x_n ermittelt (z. B. $\bar{\mu}_n := \frac{1}{n}(x_1 + \dots + x_n)$ für den MW einer Normalverteilung). In manchen Fällen kann dann auf die Reduktion der Freiheitsgrade von s auf $s - r$ verzichtet werden.

Chi-Quadrat-Unabhängigkeitstest

Ziel. Überprüfen, ob die Komponenten $X \in \mathbb{R}^{n_1}$ und $Y \in \mathbb{R}^{n_2}$ eines zweidim. Zufallsvektors $(X, Y)^T$ **unabhängig** sind.

Verfahren. Seien $I_1, \dots, I_k \subset \mathbb{R}^{n_1}$ und $J_1, \dots, J_l \subset \mathbb{R}^{n_2}$ jeweils Familien paarweise disjunkter Mengen mit $\mathbb{P}(X \in I_1 \cup \dots \cup I_k) = 1$ bzw. $\mathbb{P}(Y \in J_1 \cup \dots \cup J_l) = 1$. Wir setzen

$$p_{ij} := \mathbb{P}((X, Y) \in I_i \times J_j) = \mathbb{P}(\{X \in I_i\} \cap \{Y \in J_j\}), \\ p_{i\bullet} := \sum_{j=1}^l p_{ij} = \mathbb{P}(X \in I_i), \quad p_{\bullet j} := \sum_{i=1}^k p_{ij} = \mathbb{P}(Y \in J_j).$$

Wir wollen nun die Nullhypothese $H_0 : \forall (i, j) : p_{ij} = p_{i\bullet} \cdot p_{\bullet j}$ gegen $H_1 : \exists (i, j) : p_{ij} \neq p_{i\bullet} \cdot p_{\bullet j}$ testen. Wir zählen dazu die Häufigkeiten einer Stichprobe $(X_1, Y_1), \dots, (X_n, Y_n)$:

$$h_{ij}^{(n)} := |\{m \in \{1, \dots, n\} \mid (X_m, Y_m) \in I_i \times J_j\}|, \\ h_{i\bullet} := \sum_{j=1}^l h_{ij}, \quad h_{\bullet j} := \sum_{i=1}^k h_{ij}.$$

Diese Häufigkeiten werden in einer **Kontingenztafel** dargestellt:

	1	2	...	l	
1	$h_{11}^{(n)}$	$h_{12}^{(n)}$...	$h_{1l}^{(n)}$	$h_{1\bullet}^{(n)}$
2	$h_{21}^{(n)}$	$h_{22}^{(n)}$...	$h_{2l}^{(n)}$	$h_{2\bullet}^{(n)}$
...
k	$h_{k1}^{(n)}$	$h_{k2}^{(n)}$...	$h_{kl}^{(n)}$	$h_{k\bullet}^{(n)}$
	$h_{\bullet 1}^{(n)}$	$h_{\bullet 2}^{(n)}$...	$h_{\bullet l}^{(n)}$	n

Wir können den Test nun wie folgt als Spezialfall des χ^2 -Anpassungstests verstehen: Die Nullhypothese ist, dass die Verteilung von (X, Y) das Produkt der Verteilungen von X und Y ist. Dabei schätzen wir zunächst die Verteilungen von X und Y mit

$$L(h_{1\bullet}^{(n)}, \dots, h_{k\bullet}^{(n)}, h_{\bullet 1}^{(n)}, \dots, h_{\bullet l}^{(n)}; p_{1\bullet}, \dots, p_{k-1,\bullet}, p_{\bullet 1}, \dots, p_{\bullet, l-1}) \\ := \prod_{i=1}^k (p_{i\bullet})^{h_{i\bullet}^{(n)}} \cdot \prod_{j=1}^l (p_{\bullet j})^{h_{\bullet j}^{(n)}}.$$

Diese Funktion wird maximal bei $\hat{p}_{i\bullet} = h_{i\bullet}^{(n)}/n$ und $\hat{p}_{\bullet j} = h_{\bullet j}^{(n)}/n$. Als Test-Statistik verwenden wir

$$\hat{T}_{k,l}^{(n)} := \sum_{i=1}^k \sum_{j=1}^l \frac{(h_{ij}^{(n)} - n\hat{p}_{i\bullet}\hat{p}_{\bullet j})^2}{n\hat{p}_{i\bullet}\hat{p}_{\bullet j}} = n \sum_{i=1}^k \sum_{j=1}^l \frac{(h_{ij}^{(n)} - h_{i\bullet}^{(n)} \cdot h_{\bullet j}^{(n)}/n)^2}{h_{i\bullet}^{(n)} \cdot h_{\bullet j}^{(n)}} \\ \xrightarrow[n \rightarrow \infty]{d} \chi_{kl-1-(k-1)-(l-1)}^2 = \chi_{(k-1)(l-1)}^2$$

Entscheidungsregel. H_0 wird genau dann abgelehnt, falls

$$\hat{T}_{k,l}^{(n)} > \chi_{(k-1)(l-1), 1-\alpha}^2.$$

Bem. • Zum Testen eines höherdim. ZV (X_1, \dots, X_r) auf Unabhängigkeit aller Komponenten untersuchen wir die Ereignisse

$$(X_1, \dots, X_r) \in I_{i_1}^{(1)} \times \dots \times I_{i_r}^{(r)} \quad \text{für } (i_1, \dots, i_r) \in \bigtimes_{j=1}^r \{1, \dots, k_j\}$$

für eine passende Intervalleinteilung. Wir verwenden dann

$$\hat{T}_{k_1, \dots, k_r}^{(n)} := n^{r-1} \sum_{i_1=1}^{k_1} \dots \sum_{i_r=1}^{k_r} \frac{(h_{i_1 \dots i_r}^{(n)} - n^{-r+1} \prod_{j=1}^r h_{\bullet \dots \bullet i_j}^{(n)})^2}{\prod_{j=1}^r h_{\bullet \dots \bullet i_j}^{(n)}} \\ \xrightarrow[n \rightarrow \infty]{d} \chi_{k_1 \dots k_s - k_1 - \dots - k_r + r - 1}^2$$

• Im Spezialfall $k=l=2$ (Vierfeldertafel) hat die Statistik die Form

$$\hat{T}_{2,2}^{(n)} = n \cdot \frac{(h_{11}^{(n)} \cdot h_{22}^{(n)} - h_{12}^{(n)} \cdot h_{21}^{(n)})^2}{h_{\bullet 1}^{(n)} \cdot h_{\bullet 2}^{(n)} \cdot h_{1\bullet}^{(n)} \cdot h_{2\bullet}^{(n)}} \xrightarrow[n \rightarrow \infty]{d} \chi_1^2 = \mathcal{N}^2(0, 1)$$

und wir lehnen H_0 genau dann ab, wenn $\hat{T}_{2,2}^{(n)} > \chi_{1, 1-\alpha}^2 = z_{1-\alpha/2}^2$.

Kolmogorow-Smirnow-1SP-Test

Situation. Sei $X_1, \dots, X_n \sim F$ eine math. SP. Wir sortieren die dabei gezogenen Werte aufsteigend: $X_{1:n} \leq X_{2:n} \leq \dots \leq X_{n:n}$.

Dann heißt $\hat{F}_n(x) := \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{(-\infty, x]}(X_{i:n})$ **empirische VF**.

Satz (**Gliwkeno-Cantelli**, Hauptsatz der math. Statistik).

$$\sup_{x \in \mathbb{R}^1} |\hat{F}_n(x) - F(x)| \xrightarrow[n \rightarrow \infty]{\mathbb{P}\text{-f. s.}} 0$$

Lem. Sei F stetig. Dann ist die Verteilung von $\sup_x |\hat{F}_n(x) - F(x)|$ nicht von der Verteilungsfunktion F abhängig. Genauer:

$$\sup_x |\hat{F}_n(x) - F(x)| \stackrel{d}{=} \sup_{0 \leq y \leq 1} |\hat{G}_n(y) - G(y)|,$$

wobei G die Verteilungsfunktion von $\mathcal{R}[0, 1]$ ist (also $G(y) = y$) und $\hat{G}_n(y) := \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{[0, y]}(U_i)$ für $U_1, \dots, U_n \sim \mathcal{R}[0, 1]$ i. i. d.

Kor. Sei F stetig, $n \geq 1$. Dann ist die Verteilungsfunktion

$$K_n(z) := \mathbb{P}(\sqrt{n} \cdot \sup_{x \in \mathbb{R}} |\hat{F}_n(x) - F(x)| \leq z)$$

unabhängig von F .

Satz. Falls F stetig ist, so gilt für alle $z \in \mathbb{R}^1$:

$$K_n(z) \xrightarrow[n \rightarrow \infty]{} K(z) := \sum_{k=-\infty}^{\infty} (-1)^k \exp(-2k^2 z^2).$$

Def. Dabei ist K die VF der **Kolmogorow-Verteilung**.

Bem. Man zeigt dazu, dass die Folge $X_n : y \mapsto \sqrt{n} \cdot (\hat{G}_n(x) - x)$ gegen die **Brownsche Brücke** \dot{B} konvergiert. Für diese gilt

$$\sup_{0 \leq x \leq 1} |\dot{B}(x)| \sim K.$$

Entscheidungsregel (Kolmogorow-(Smirnow-)1SP-Test). Wir testen $H_0 : F = F_0$ **gegen** $H_1 : F \neq F_0$. Dabei muss F_0 eine stetige VF sein. Wir verwenden dazu

$$T_n := \sqrt{n} \cdot \sup_{x \in \mathbb{R}} |\hat{F}_n(x) - F_0(x)|.$$

Wir lehnen H_0 genau dann ab, wenn $T_n > K_{1-\alpha}$.

Bem. • Für kleine $n \in \mathbb{N}$ sollte man $K_{n, 1-\alpha}$ verwenden.

• Für große z ist $K(z) \approx 1 - 2 \exp(-2z^2)$, also

$$K_{1-\alpha} \approx \sqrt{-1/2 \cdot \log(\alpha/2)} \quad \text{für } \alpha \text{ klein.}$$

• Das Supremum in T_n liegt bei einer Sprungstelle von \hat{F}_n .

Test (einseitiger Kolmogorow-(Smirnow-)1SP-Test).

Wir testen $H_0 : F = F_0$ **gegen** $H_1 : F > F_0$ mit

$$T_n^+ := \sqrt{n} \cdot \sup_{x \in \mathbb{R}} (\hat{F}_n(x) - F(x)).$$

Für alle $z \in \mathbb{R}^1$ gilt

$$K_n^+(z) := \mathbb{P}(T_n^+ \leq z) \xrightarrow[n \rightarrow \infty]{} K^+(z) := 1 - \exp(-2 \max(0, z)^2).$$

Wir lehnen H_0 ab, falls $T_n^+ > K_{1-\alpha}^+$.

Achtung. Der Kolmogorow-Test kann nicht verwendet werden, wenn die Parameter von F_0 aus der Stichprobe geschätzt werden.

Bem. Es gibt keine Entsprechung für mehrdimensionale ZVen

Cramér-von-Mises-Test

$$\text{Def. } \omega_n^2(g) = n \int_{\mathbb{R}^1} g(F(x)) (\hat{F}_n(x) - F(x))^2 dF(x)$$

heißt gewichtete **Cramér-von-Mises-Statistik** oder ω^2 -Statistik. Dabei ist $g : [0, 1] \rightarrow [0, \infty]$ eine Gewichtsfnkt. Häufig verwendet wird $g(x) := 1$ und die **Anderson-Darling-Statistik** $g(x) := \frac{1}{x(1-x)}$.

Satz. Sei F stetig. Dann ist

$$\omega_n^2(g) \stackrel{d}{=} n \int_0^1 g(u) (\hat{G}_n(u) - u)^2 du \xrightarrow[n \rightarrow \infty]{d} \int_0^1 g(u) (\dot{B}(u))^2 du =: \omega^2(g).$$

Entscheidungsregel (CvM-Test). Wir testen $H_0 : F = F_0$ **vs.** $H_1 : F \neq F_0$ anhand der CvM-Statistik. Wir lehnen H_0 genau dann ab, wenn $\omega_n^2(g) > \omega_{1-\alpha}^2(g)$.

Bem. Der rechte Wert ist tabelliert für wichtige Funktionen g .

Kolmogorow-Smirnow-2SP-Test

Situation. Gegeben seien zwei unabhängige SPn $X_1, \dots, X_n \sim F$ i. i. d. und $X_1^*, \dots, X_m^* \sim F^*$ i. i. d., wobei F und F^* stetig sind. Wir wollen $H_0 : F = F^*$ **vs.** $H_1 : F \neq F^*$ testen, indem wir die empirischen VFen \hat{F}_n und \hat{F}_m^* vergleichen. Dazu verwenden wir

$$T_{m,n} := \sqrt{\frac{m \cdot n}{m+n}} \sup_{x \in \mathbb{R}^1} |\hat{F}_n(x) - \hat{F}_m^*(x)|$$

Satz. Falls $F = F^*$ stetig ist, so gilt

$$T_{m,n} \stackrel{d}{=} \sqrt{\frac{m \cdot n}{m+n}} \sup_{0 \leq u \leq 1} \left| \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{[0, u]}(U_i) - \frac{1}{m} \sum_{j=1}^m \mathbb{1}_{[0, u]}(U_j^*) \right|,$$

wobei $U_k := F(X_k)$, $k = 1, \dots, n$ und $U_l^* := F(X_l^*)$, $l = 1, \dots, m$ jeweils $\mathcal{R}[0, 1]$ -verteilt sind.

Lem. $T_{m,n} \xrightarrow[n \rightarrow \infty]{d} \sup_{0 \leq u \leq 1} |\dot{B}(u)| \sim K$

Entscheidungsregel (Kolmogorow-(Smirnow-)2SP-Test). $H_0 : F = F^*$ wird genau dann abgelehnt, falls $T_{m,n} > K_{1-\alpha}$.

2SP-Test von Wilcoxon-Mann-Whitney

Situation (2-SP-Test von Wilcoxon-Mann-Whitney, U-Test). Geg. seien zwei unabh. SPn $X_1, \dots, X_n \sim F$ und $X_1^*, \dots, X_m^* \sim F^*$, wobei F und F^* stetig sind. Ziel: Prüfen von $H_0 : F = F^*$ **vs.** $H_1 : F \neq F^*$. Dazu konstruieren wir eine Rangstatistik für konkrete Stichproben x_1, \dots, x_n und x_1^*, \dots, x_m^* :

1. Ordnen: $x_{1:n} < \dots < x_{n:n}$ und $x_{1:m}^* < \dots < x_{m:m}^*$
2. $\nu_1, \dots, \nu_m \in \{1, \dots, m+n\}$ seien die Ränge der Werte $x_{i:m}^*$ innerhalb der Gesamtstichprobe, d. h.

$$x_{1:n} < \dots < x_{\nu_1-1:n} < x_1^* < x_{\nu_1:n} < \dots < x_{\nu_2-2:n} < x_{2:m}^* < x_{\nu_2-1:n} < \dots < x_{\nu_m-m:n} < x_{m:m}^* < x_{\nu_m-m+1:n} < \dots < x_{n:n}.$$

Heuristik: H_0 wird angenommen, falls sich die x - und x^* -Werte „gut durchmischen“, d. h. die Anzahl der x -Werte, die vor bzw. nach den x^* -Werten liegen, darf nicht zu groß werden. Die Testgröße dafür ist

$$W_{m,n} := \sum_{i=1}^n \sum_{j=1}^m \mathbb{1}_{\{X_i < X_j^*\}} = |\{(i, j) \mid X_i < X_j^*\}| = \sum_{j=1}^m |\{i \mid X_i < X_{j:m}^*\}| \\ = \sum_{j=1}^m (\nu_j - j) = \nu_1 + \dots + \nu_m - \frac{m(m+1)}{2}$$

Lem. Unter $H_0 : F = F^*$ stetig gilt:

$$\begin{aligned} \text{a) } \mathbb{E}W_{m,n} &= \frac{m \cdot n}{2} & \text{b) } \text{Var} W_{m,n} &= \frac{m \cdot n}{12} (m+n+1) \\ \text{c) } g_{m,n}(z) &:= \sum_{k=0}^{n \cdot m} \mathbb{P}(W_{m,n} = k) \cdot z^k = \\ &= \frac{z^{-m(m+1)/2}}{\binom{m+n}{m}} \sum_{1 \leq \nu_1 < \dots < \nu_m \leq m+n} z^{\nu_1 + \dots + \nu_m} = \frac{1}{\binom{m+n}{m}} \prod_{k=1}^m \frac{1 - z^{n+k}}{1 - z^k} \end{aligned}$$

Entscheidungsregel. Ablehnung von H_0 , falls $w_{m,n} \leq c_{\alpha/2}$ oder $w_{m,n} \geq m \cdot n - c_{\alpha/2}$, wobei

$$c_{\alpha/2} = \min\{k \geq 0 \mid \mathbb{P}(W_{m,n} \leq k) = \mathbb{P}(W_{m,n} \geq m \cdot n - k) \geq \alpha/2\}.$$

Annahme von H_0 genau dann, wenn $|w_{m,n} - \frac{m \cdot n}{2}| < \frac{m \cdot n}{2} - c_{\alpha/2}$.

Satz. Unter $H_0 : F = F^*$ stetig gilt

$$T_{m,n} := \frac{W_{m,n} - \frac{m \cdot n}{2}}{\sqrt{\frac{m \cdot n}{2} (m+n+1)}} \xrightarrow{m,n \rightarrow \infty} \mathcal{N}(0, 1).$$

Entscheidungsregel. Man erhält aus dem letzten Satz einen asymptotischen Test, den man für große m, n verwenden kann: Wir lehnen genau dann $H_0 : F = F^*$ ab, falls $|T_{m,n}| \geq z_{1-\alpha/2}$.

Kruskal-Wallis-Test

Test (Kruskal-Wallis). Gegeben seien k Messreihen $X_{i,1}, \dots, X_{i,n_i} \sim F_i$, $i = 1, \dots, k$ unabhängige SPn, F_i stetig. Ziel: Testen von $H_0 : F_1 = \dots = F_k$. Vorgehen:

1. Ordnen der Beobachtungen der Größe nach
2. $\nu_{i,1} < \dots < \nu_{i,n_i}$ Platznummern der n_i Beobachtungen der i -ten Messreihe in der Gesamt-SP
3. $\bar{\nu}_i := \frac{1}{n_i} (\nu_{i,1} + \dots + \nu_{i,n_i})$, $\bar{\nu} := \frac{1}{n} \sum_{i=1}^k n_i \bar{\nu}_i$ mit $n := n_1 + \dots + n_k$.

Heuristik: H_0 ist richtig, falls $\bar{\nu}_i \approx \bar{\nu}$ für alle i . Testgröße:

$$H := \frac{12}{n(n+1)} \sum_{i=1}^k n_i (\bar{\nu}_i - \frac{n+1}{2})^2 \xrightarrow{n_i \rightarrow \infty} \chi_{k-1}^2$$

Wir lehnen H_0 genau dann ab, wenn $H > \chi_{k-1, 1-\alpha}^2$.

Faustregel. Die Approx. ist gut, wenn $\min_{1 \leq i \leq k} n_i \geq 5$ und $k \geq 4$.

Theorie der U-Statistiken

Situation. Sei $n \geq m$, $X_1, \dots, X_n \sim F$ i. i. d., $h : \mathbb{R}^m \rightarrow \mathbb{R}^1$ Borel-messbar und symmetrisch, d. h.

$$h(x_1, \dots, x_m) = h(x_{\sigma(1)}, \dots, x_{\sigma(m)}) \quad \forall \sigma \in S_m.$$

Gelte $\mathbb{E}|h(X_1, \dots, X_m)| < \infty$.

Def. Die **U-Statistik der Ordnung m** mit Kernfunktion h ist

$$U_n^{(m)} := \frac{1}{\binom{n}{m}} \sum_{1 \leq i_1 < \dots < i_m \leq n} h(X_{i_1}, \dots, X_{i_m}).$$

Bem. Offenbar: $\mathbb{E}U_n^{(m)} = \mathbb{E}h(X_1, \dots, X_m)$.

Bsp. Für $m = 2$ gilt $\sigma^2 = \text{Var}(X_1) = \frac{1}{2} \mathbb{E}(X_1 - X_2)^2$. Davon inspiriert setzen wir $h(x_1, x_2) := \frac{1}{2} (x_1 - x_2)^2$. Damit haben wir

$$U_n^{(2)} = \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} \frac{1}{2} (X_i - X_j)^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2 = S_n^2$$

Ziel. Wir würden gerne den ZGWS auf $U_n^{(m)}$ anwenden. Problem dabei: Die Summanden in der Def. von $U_n^{(m)}$ sind nicht unabhängig. Wir approximieren deshalb $U_n^{(m)}$ mit einer Summe von i. i. d. ZGn.

Lem. Sei $\tilde{U}_n^{(m)} = \theta + \sum_{i=1}^n (\underbrace{\mathbb{E}(U_n^{(m)} | X_i)}_{\text{i. i. d.}} - \theta)$ mit $\theta := \mathbb{E}U_n^{(m)}$ und

$$\begin{aligned} g(x) &= \mathbb{E}[h(X_1, \dots, X_n) | X_1 = x] = \mathbb{E}h(x, X_2, \dots, X_n) \\ &= \int \dots \int h(x, x_2, \dots, x_n) dF(x_2) \dots dF(x_n). \end{aligned}$$

Falls $\mathbb{E}h^2(X_1, \dots, X_m) < \infty$, so gilt

- (1) $\text{Var}(U_n^{(m)} - \tilde{U}_n^{(m)}) = \text{Var}(U_n^{(m)}) - \text{Var}(\tilde{U}_n^{(m)})$
- (2) $\mathbb{E}(U_n^{(m)} | X_i = x) = \theta + \frac{m}{n} (g(x) - \theta)$

Lem. (2) $\text{Var}(\tilde{U}_n^{(m)}) = \frac{m^2}{n} \cdot \text{Var}(g(X_1)) = \frac{m^2}{2} (\mathbb{E}g^2(X_1) - \theta^2)$

(3) Falls $\mathbb{E}|h(X_1, \dots, X_m)| < \infty$, so gilt

$$\begin{aligned} \text{Var}(U_n^{(m)}) &= \left(\frac{n}{m}\right) \sum_{k=1}^m \binom{m}{k} \binom{n-m}{m-k} \cdot \zeta_k \quad \text{mit} \\ h_k(x_1, \dots, x_k) &:= \mathbb{E}[h(x_1, \dots, x_k, X_{k+1}, \dots, X_m)] \\ \zeta_k &:= \text{Var}(h_k(X_1, \dots, X_k)) \\ &= \mathbb{E}[h(X_1, \dots, X_k, X_{k+1}, \dots, X_m) \cdot h(X_1, \dots, X_k, X_{m+1}, \dots, X_{2m-k})] - \theta^2 \end{aligned}$$

Kor. Aus (1), (3) und (4) folgt für $m = 2$:

$$\text{Var}(U_n - \tilde{U}_n) = \text{Var}(U_n) - \text{Var}(\tilde{U}_n) = \dots = -\frac{4}{n(n-1)} \text{Var}(g(X_1))$$

Für $m \geq 2$ gilt $\text{Var}(U_n^{(m)} - \tilde{U}_n^{(m)}) \leq \frac{c(m)}{n^2} \text{Var}(h(X_1, \dots, X_m))$.

Satz (Hoeffding). Sei $U_n^{(m)}$ eine U-Statistik mit Kern $h : \mathbb{R}^m \rightarrow \mathbb{R}$, sodass $\mathbb{E}h^2(X_1, \dots, X_m) < \infty$ und $\sigma_g^2 := \text{Var}(g(X_1)) > 0$. Dann gilt

$$\sqrt{n}(U_n^{(m)} - \theta) \xrightarrow{n \rightarrow \infty} \mathcal{N}(0, \sigma_g^2).$$

Bem. • Der Fall $\text{Var}(g(X_1)) = 0$ (entarteter Fall) zieht eine kompliziertere Asymptotik nach sich.

• $\mathbb{E}g^2(X_1) < \infty$ ist schwächer als $\mathbb{E}h^2(X_1, \dots, X_m) < \infty$.

• Aus $\mathbb{E}|h(X_1, \dots, X_m)|^{1+q} < \infty$ für $0 < q \leq 1$ folgt

$$\mathbb{E}|\sqrt{n}(U_n^{(m)} - \tilde{U}_n^{(m)})|^{1+q} \leq \frac{c(q, m)}{n^{2q}} \mathbb{E}|h(X_1, \dots, X_m)|^{1+q}.$$

Mit einer Abschneidetechnik zeigt man, dass $\mathbb{E}g^2(X^1) < \infty$ und $\mathbb{E}|h(X_1, \dots, X_m)|^{\frac{4}{3}} < \infty$ schon für $\mathbb{P}(\sqrt{n}|U_n^{(m)} - \tilde{U}_n^{(m)}| < \epsilon) \rightarrow 0$ für alle $\epsilon > 0$ ausreichen und damit für den Satz von Hoeffding.

• U-Statistiken erweisen sich (unter gewissen Bedingungen) als suffiziente Schätzer mit minimaler Varianz.

Bsp. Wir betrachten die U-Statistik $S_n^2 = \left(\frac{n}{2}\right)^{-1} \sum_{i < j} \frac{1}{2} (X_i - X_j)^2$.

Dann ist $g(x) = \frac{1}{2} (x - \mathbb{E}X_1)^2 + \frac{1}{2} \sigma^2$ mit $\sigma^2 := \text{Var}(X_1)$. Es gilt

$$\sqrt{n}(S_n^2 - \sigma^2) \xrightarrow{n \rightarrow \infty} \mathcal{N}(0, 4\sigma_g^2)$$

mit $\sigma_g^2 = \mathbb{E}g^2(X_2) - (\mathbb{E}g(X_2))^2 = \frac{1}{4} \mu_4 - \frac{1}{4} \sigma^4$, $\mu_4 := \mathbb{E}(X_1 - \mathbb{E}X_2)^4$. Spezialfall: Ist $X_i \sim \mathcal{N}(\mu, \sigma^2)$, so gilt $\mu_4 = 3\sigma^4$.

Dann gilt $\sqrt{n}(S_n^2 - \sigma^2) \xrightarrow{n \rightarrow \infty} \mathcal{N}(0, 2\sigma^4)$. Es folgt

$$\frac{\sqrt{n}(S_n^2 - \sigma^2)}{\sqrt{2(S_n^2)^2}} = \sqrt{n/2} \left(1 - \frac{\sigma^2}{S_n^2}\right) \xrightarrow{n \rightarrow \infty} \mathcal{N}(0, 1).$$

Alternativ erhält man durch Anwenden einer Varianzstab. Trafo:

$$\sqrt{n/2} (\log S_n^2 - \log \sigma^2) \xrightarrow{n \rightarrow \infty} \mathcal{N}(0, 1).$$

Def. Die **Kumulante** oder **Semi-Invariante** m -ter Ordnung ist

$$\text{Cum}_m(X) = \frac{1}{m! 2^m} \frac{\partial^m}{\partial t^m} \log \mathbb{E}e^{itX}.$$

Bem. Falls X_1, \dots, X_n unabhängig sind, so gilt

$$\text{Cum}_m(X_1 + \dots + X_n) = \text{Cum}_m(X_1) + \dots + \text{Cum}_m(X_n).$$

Für $m = 3$ gilt $\text{Cum}_3(X) = \mathbb{E}X^3 - 3\mathbb{E}X \cdot \mathbb{E}X^2 + 2(\mathbb{E}X)^3$.

Bsp. Schätzung der Kumulante m -ter Ord. mit der SP X_1, \dots, X_n :

$$\begin{aligned} (\widehat{\text{Cum}_3(X)})_n &:= \frac{1}{n(n-1)(n-2)} (n^2 \hat{M}_3^{(n)} - 3n \hat{M}_1^{(n)} \hat{M}_2^{(n)} - 2(\hat{M}_1^{(n)})^3) \\ &= \left(\frac{1}{n}\right) \sum_{1 \leq i < j < k \leq n} h(X_i, X_j, X_k) \end{aligned}$$

mit $h(x, y, z) := -\frac{1}{2}(xy^2 + x^2y + xz^2 + x^2z + yz^2 + x^2z + yz^2 + y^2z) + \frac{1}{3}(x^3 + y^3 + z^3) + 2xyz$

wobei $\hat{M}_j^{(n)} := \frac{1}{n} \sum_{i=1}^n X_i^j$

Def. Eine VF F heißt **symmetrisch** bzgl. $\vartheta_0 \in \mathbb{R}^1$, falls

$$F(\vartheta_0 - x) = 1 - F(\vartheta_0 + x) \quad \forall x \in \mathbb{R}^1.$$

Bsp (Wilcoxon-1-SP-Test auf Symmetrie). Sei $X_1, \dots, X_n \sim F$ eine math. SP mit F stetig. Angenommen, F ist symmetrisch bzgl. $\vartheta_0 \in \mathbb{R}^1$. Dann sind $Z_i = X_i - \vartheta_0$ symmetrisch bzgl. 0. Seien ν_1^+, \dots, ν_n^+ die Ränge der ZGn $|Z_1|, \dots, |Z_n|$. Setze

$$T_n^+ = \sum_{i=1}^n \mathbb{1}_{\{Z_i > 0\}} \nu_i^+.$$

Unter $H_0 : F$ ist symmetrisch bzgl. ϑ_0 gilt

$$\mathbb{E}T_n^+ = \frac{1}{2} \sum_{i=1}^n \mathbb{E}\nu_i^+ = \frac{n(n+1)}{4}, \quad \text{Var}(T_n^+) = \frac{n}{24}(n+1)(2n+1).$$

Bsp. Alternativ können wir die U-Statistik

$$U_n = \left(\frac{1}{n}\right) \sum_{1 \leq i < j \leq n} \mathbb{1}_{\{Z_i + Z_j > 0\}}.$$

zum Test auf Symmetrie betrachten. Unter H_0 gilt

$$\begin{aligned} \mathbb{E}\mathbb{1}_{\{Z_i + Z_j > 0\}} &= \mathbb{P}(Z_1 > -Z_2) = \int (1 - F(-z)) \, dF(z) \\ &= \int F(z) \, dF(z) = \frac{1}{2}. \end{aligned}$$

Aus dem ZGWS für U-Statistiken folgt

$$\sqrt{n}(U_n - \frac{1}{2}) \xrightarrow{n \rightarrow \infty} \mathcal{N}(0, \frac{1}{3}).$$

Testregel: Wir lehnen H_0 genau dann ab, wenn $|U_n - \frac{1}{2}| > \frac{z_{1-\alpha/2}}{\sqrt{3n}}$.

Def. Sei $h : \mathbb{R}^{m_1} \times \mathbb{R}^{m_2} \rightarrow \mathbb{R}^1$ Borel-messbar, symmetrisch in den ersten m_1 und den letzten m_2 Argumenten. Seien $X_1, \dots, X_{n_1} \sim F$ und $X_1^*, \dots, X_{n_2}^* \sim F^*$ zwei unabh. math. SPn. Dann heißt

$$U_{n_1, n_2}^{(m_1, m_2)} := \left(\binom{n_1}{m_1} \binom{n_2}{m_2} \right)^{-1} \sum_{\substack{1 \leq i_1 < \dots < i_{m_1} \leq n_1 \\ 1 \leq j_1 < \dots < j_{m_2} \leq n_2}} h(X_{i_1}, \dots, X_{i_{m_1}}, X_{j_1}^*, \dots, X_{j_{m_2}}^*)$$

(verallg.) **U-Statistik** der Ordnung (m_1, m_2) mit Kernfunktion h .

Notation. Sei $m_1 = m_2 = 1$. Wir setzen

$$\begin{aligned} \theta &:= \mathbb{E}h(X_1, X_1^*) = \mathbb{E}U_{n_1, n_2}^{(1,1)} \\ g_1(x) &:= \mathbb{E}(h(X_1, X_1^*) \mid X_1 = x), \quad \sigma_1^2 := \text{Var } g(X_1), \\ g_2(y) &:= \mathbb{E}(h(X_1, X_1^*) \mid X_1^* = y), \quad \sigma_2^2 := \text{Var } g(X_1^*), \\ \tilde{U}_{n_1, n_2}^{(1,1)} &:= \frac{1}{n_1} \sum_{i=1}^{n_1} g_1(X_i) + \frac{1}{n_2} \sum_{j=1}^{n_2} g_2(X_j^*) - \theta \end{aligned}$$

Lem. Es seien $\mathbb{E}h^2(X_1, X_1^*) < \infty$ und $\sigma_1^2, \sigma_2^2 \in (0, \infty)$. Dann gilt

$$\sqrt{\frac{n_1 n_2}{n_2 \sigma_1^2 + n_1 \sigma_2^2}} \cdot (U_{n_1, n_2} - \theta) \xrightarrow{n_1, n_2 \rightarrow \infty} \mathcal{N}(0, 1).$$

Bsp. Die Wilcoxon-2-SP-Statistik ist eine U-Statistik mit

$$h(x, y) := |\{\heartsuit \mid x < y\}|.$$

Das allgemeine lineare Modell

Modell (allgemein). Für Zufallsgrößen X und Y gilt $Y = g(X) + \epsilon$ mit einer Funktion g , wobei $\mathbb{E}\epsilon = 0$ und $\sigma^2 := \text{Var}(Y - g(X)) = \mathbb{E}\epsilon^2$.

Modell (Lineare Regression). $Y = X\beta + \epsilon$, wobei

$$\begin{aligned} Y &= (Y_1, \dots, Y_n)^T & \text{Beobachtungsvektor,} \\ X &= (x_{ij}) \in \mathbb{R}^{n \times p} & \text{Einstellgrößen-, Versuchsplanmatrix,} \\ \beta &= (\beta_1, \dots, \beta_p)^T & \text{(unbek.) Parametervektor, Regressionskoeff.,} \\ \epsilon &= (\epsilon_1, \dots, \epsilon_n)^T & \text{(nicht beobachtbarer) Fehlervektor heißt.} \end{aligned}$$

Wir nehmen an, dass $\text{Cov}(Y_i, Y_j) = \text{Cov}(\epsilon_i, \epsilon_j) = \sigma^2 \delta_{ij}$ und $n > p$. Dabei heißt σ **Modellstreuung**.

Bem. Falls Y eine bekannte Kovarianzmatrix $K \in \mathbb{R}^{n \times n}$ besitzt, so können wir $X^* := K^{-1/2}X$, $Y^* := K^{-1/2}Y$, $\epsilon^* := K^{-1/2}\epsilon$ setzen und erhalten $Y^* = X^*\beta + \epsilon^*$ und $\text{Cov}(Y^*) = I_n$.

Problem. Gegeben seien $[Y, X\beta, \sigma^2 I_n]$.

Gesucht sind Schätzungen $\hat{\beta}(y) = (\hat{\beta}_1(y), \dots, \hat{\beta}_p(y))^T$ für β .

Def. Eine Schätzfunktion $\hat{\beta}(y)$ heißt **MkQ-Schätzung** (Methode der kleinsten Quadrate) für β , falls

$$S(y, \hat{\beta}) = \min_{\beta \in \mathbb{R}^p} S(y, \beta)$$

wobei $S(y, \beta) := \|y - X\beta\|^2 = \sum_{i=1}^n (y_i - \sum_{j=1}^p x_{ij}\beta_j)^2$.

Bem. $S(y, \beta)$ besitzt lokale Minima, da

$$\frac{\partial}{\partial \beta} S(y, \beta) = -2X^T y + 2X^T X \beta, \quad \frac{\partial^2}{\partial \beta^2} S(y, \beta) = 2X^T X.$$

Für die Minima gelten die Normalgleichungen

$$X^T X \beta = X^T Y \iff \sum_{j=1}^p \xi_{ij} \beta_j = \sum_{j=1}^n x_{ji} y_j \quad \text{mit } (\xi_{ij}) = X^T X. \quad (\text{N})$$

Satz. (N) ist stets lösbar und jede Lsg ist eine MkQ-Schätzung. Falls $\text{rk } X = p$, so ist $\hat{\beta}$ eind. bestimmt durch $\hat{\beta} = (X^T X)^{-1} X^T y$.

Bsp (Einfache lineare Regression).

Annahme: $Y_i = \beta_1 + \beta_2 x_i + \epsilon_i$, $i = 1, \dots, n$. Dann ist

$$\begin{aligned} X &= \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} & X^T X &= n \sum x_i^2 - (\sum x_i)^2 = n \cdot \sum (x_i - \overline{x_n})^2 > 0 \\ & & \hat{\beta} &= \det(X X^T)^{-1} \begin{pmatrix} \sum x_i^2 & -\sum x_i \\ -\sum x_i & n \end{pmatrix} \begin{pmatrix} \sum Y_i \\ \sum x_i Y_i \end{pmatrix} \end{aligned}$$

Bsp (Multiple linear Regression).

$$Y_i = \beta_0 + \beta_1 X_1^{(i)} + \dots + \beta_m X_m^{(i)} + \epsilon_i$$

Bsp (Quasilineare (multiple) Regression).

$$Y_i = \beta_0^{(i)} + \beta_1 f_1(X_1^{(i)}) + \dots + \beta_m f_m(X_m^{(i)}) + \epsilon_i$$

mit (nichtlinearen) Funktionen f_1, \dots, f_m

Def. Eine Matrix $A^- \in \mathbb{R}^{n \times m}$ heißt **g-Inverse** (g = generalized) von $A \in \mathbb{R}^{m \times n}$, wenn für jedes $y \in \mathbb{R}^m$, für welches $Ax = y$ lösbar ist, auch $x = A^- y$ eine Lösung ist.

Satz. A^- ist eine g-Inverse von $A \iff AA^- A = A$

Bem. • Falls $n = m$ und A^{-1} existiert, so ist $A^- = A^{-1}$ eindeutig.

• A^- ist im Allgemeinen nicht eindeutig. Man erhält Eindeutigkeit durch Zusatzforderungen:

Def. Eine **Moore-Penrose-Inverse** A^+ ist eine g-Inverse, welche folgende Bedingungen erfüllt:

$$A^+ A A^+ = A^+, \quad (A A^+)^T = A A^+, \quad (A^+ A)^T = A^+ A.$$

Satz. Die allgemeine Lösung von (N) lautet mit $S := X^T X$:

$$\beta = S^- X^T y + (S^- S - I_p) z, \quad \text{wobei } z \in \mathbb{R}^p.$$

Für die spezielle Lsg $z = 0$ gilt $\hat{\beta} = S^- X^T Y$ für die MkQ-Schätzung für β . Es gilt

$$\mathbb{E}\hat{\beta} = S^- S \beta \quad \text{und} \quad \text{Cov}(\hat{\beta}) = \sigma^2 S^- S S^-.$$

Bem. Bei Nichteindeutigkeit der Lsg von (N) gilt i. A. $S^- S \neq I_p$. Falls $\text{rk } X = \text{rk } S = p$, so gilt $\mathbb{E}\hat{\beta} = \beta$ und $\text{Cov } \hat{\beta} = \sigma^2 S^{-1}$

Def. Eine Linearkombination $l(\beta) = c^T \beta$ mit $c \in \mathbb{R}^p$, $\beta \in \mathbb{R}^p$ heißt bzgl. des linearen Modells $[Y, X\beta, \sigma^2 I_n]$ **schätzbare Funktion**, falls ein $a \in \mathbb{R}^n$ mit $c = X^T a$ existiert.

Satz. Es sind äquivalent:

- $l(\beta) = c^T \beta$ ist eine schätzbare Funktion
- $\hat{l} := l(\hat{\beta}) := c^T \hat{\beta}$ (wobei $\hat{\beta}$ MkQ-Schätzung) ist eine lin. Funktion von Y und eine erwartungstreue Schätzung für $l(\beta)$
- $c \in \text{im}(X^T) = \text{im}(X^T X)$
- $l(\hat{\beta}) = c^T \hat{\beta}$ ist konstant für alle $\hat{\beta}$, die Lösung von (N) sind, d. h. $X^T X \hat{\beta} = X^T y$.
- Es existiert ein $a \in \mathbb{R}^n$ mit $\mathbb{E}(a^T Y) = c^T \beta$.

Satz (Gauß-Markov). In einem lin. Modell $[Y, X\beta, \sigma^2 I_n]$ ex. für jede schätzbare (lin.) Funktion $l(\beta) = c^T \beta$ eine eindeutig bestimmte, in Y lin. erwartungstreue Schätzung $\hat{l} = a_*^T Y$ (für genau ein $a_* \in \text{im}(X) \subseteq \mathbb{R}^n$) und diese hat die Gestalt $\hat{l} = l(\hat{\beta}) = c^T \hat{\beta}$, wobei $\hat{\beta}$ eine MkQ-Schätzung ist. Außerdem besitzt \hat{l} minimale Varianz in der Klasse aller linearen erwartungstreuen Schätzungen $\hat{l} = a^T Y$.

Konstr. $a_* = X(X^T X)^- c$

Def. Der Schätzer heißt Best Linear Unbiased Estimation (**BLUE**).

Bem. Es gilt

$$\begin{aligned} S(Y, \hat{\beta}) &= \min_{\beta \in \mathbb{R}^p} \|Y - X\beta\|^2 = \|Y - X\hat{\beta}\|^2 = (Y - X\hat{\beta})^T (Y - X\hat{\beta}) = \\ &= Y^T Y - \underbrace{Y^T X \hat{\beta}}_{=(X\hat{\beta})^T X \hat{\beta}} - \underbrace{(X\hat{\beta})^T Y}_{=\hat{\beta}^T X^T X \hat{\beta} = \hat{\beta}^T X^T Y} + \underbrace{(X\hat{\beta})^T X \hat{\beta}}_{=\hat{\beta}^T X^T X \hat{\beta} = \hat{\beta}^T X^T Y} = \|Y\|^2 - \|X\hat{\beta}\|^2. \end{aligned}$$

Def. $(Y - X\hat{\beta})$ heißt **Restvektor** oder **Residuum**.

Lem. Falls $\hat{\beta}$ MkQ-Schätzung, so gilt

- $\mathbb{E}(Y - X\hat{\beta}) = 0$
- $c^T \beta$ ist eine schätzbare Funktion und $\mathbb{E}(c^T \hat{\beta}(Y - X\hat{\beta})) = 0$
- $\text{Cov}(Y - X\hat{\beta}) = \text{Cov}(Y) - \text{Cov}(X\hat{\beta})$.

Orthogonale Transformation von $[Y, X\beta, \sigma^2 I_n]$

$X = (\tilde{x}_1 \cdots \tilde{x}_p)$ mit $\tilde{x}_i \in \mathbb{R}^p$, $\text{rk } X = r \leq p$.

Es existiert eine orthonormale Basis o_1, \dots, o_r von $\text{im}(X)$ und $\sigma_{r+1}, \dots, \sigma_n \in \text{im}(X)^\perp$.

$Y = \sum_{i=1}^n o_i Z_i$, $Z = (Z_1 \cdots Z_n)^T$, d. h. $Y = OZ$, wobei $O(o_1 \cdots o_n)$,

$O_1 = (a_1 \cdots a_r)^T$, $O_2 = (a_{r+1} \cdots a_n)^T$

$Z = O^{-1}Z = O^T Y$

$\mathbb{E}Z = O^T \mathbb{E}Y = O^T X\beta = \begin{pmatrix} O_1^T \\ O_2^T \end{pmatrix} X\beta$

$\mathbb{E}Z_i = o_i^T X\beta = \begin{cases} o_i^T X\beta & \text{für } i = 1, \dots, r, \\ 0 & \text{für } i = r+1, \dots, n. \end{cases}$

$\mathbb{E}Z = \begin{pmatrix} O_1^T X\beta \\ 0 \end{pmatrix}$, $\text{Cov}(Z) = \text{Cov}(O^T Y) = O^T \text{Cov}(Y) O = \sigma^2 I$

Transformation: $[Y, X\beta, \sigma^2 I_n] \rightsquigarrow [Z, O^T X\beta, \sigma^2 I_n]$ mit $Z = O^T Y$

Satz. In einem linearen Modell $[Y, X\beta, \sigma^2 I_n]$ mit $\text{rk } X = r \leq p$ und einer MkQ-Schätzung $\hat{\beta}$ ist

$$\sigma^2 = \frac{1}{n-r} S(Y, \hat{\beta}) = \frac{1}{n-r} \|Y - X\hat{\beta}\|^2 = \frac{1}{n-r} \sum_{i=1}^n (Y_i - \sum_{j=1}^p x_{ij} \hat{\beta}_j)^2$$

eine erwartungstreue Schätzung für σ^2 .

Satz. Für ein normalverteiltes lineares Modell $[Y, \mathcal{N}_n(X\beta, \sigma^2 I_n)]$ mit $\text{rk } X = r \leq p$ gilt:

- Die ML-Schätzung für $\beta \in \mathbb{R}^p$ stimmt mit der MkQ-Schätzung $\hat{\beta}$ überein und es gilt $\hat{\beta} \sim \mathcal{N}_p(\mathbb{E}\hat{\beta}, \text{Cov}(\hat{\beta}))$.
- Die ML-Schätzung für σ^2 lautet $\hat{\sigma}_n^2 = \frac{S(Y, \hat{\beta})}{n} = \frac{n-r}{n} \hat{\sigma}^2$ mit $\mathbb{E}\hat{\sigma}_n^2 = \frac{n-r}{n} \sigma^2 \xrightarrow{n \rightarrow \infty} \sigma^2$ (asympt. erw.-treu) und $\frac{S(Y, \hat{\beta})}{\sigma^2} \sim \chi_{n-r}^2$
- Für einen Vektor $l^T(\beta) = (l_1(\beta), \dots, l_q(\beta))$ von $q \leq r$ linear unabhängigen schätzbaren Funktionen $l_i(\beta) = c_i^T \beta$, $c_i \in \mathbb{R}^p$ gilt $\hat{l} := l(\hat{\beta}) \sim \mathcal{N}_q(l(\beta), \sigma^2 A_* A_*^T)$ mit $\text{rk } A_* = q$, wobei

$$A_* = (a_{*,1}, \dots, a_{*,q})^T \quad \text{mit } a_{*,i} \in L(X) \quad \text{optimal gemäß}$$

Gauß-Markov-Theorem

- Die Schätzungen $\hat{l} = l(\hat{\beta})$ und $\hat{\sigma}^2$ (bzw. $\hat{\sigma}_n^2$) sind unabhängig.

Kor. Für $\text{rk } X = p$ gilt $\hat{\beta} \sim \mathcal{N}_p(\beta, \sigma^2(X^T X)^{-1})$ und $\hat{\beta}$ und $\hat{\sigma}^2$ sind unabhängig. (Grund: $\beta_i = e_i^T \hat{\beta}$ sind schätzbare Funktionen)

Test (σ^2 -Streuungstest im Modell $(Y, \mathcal{N}_n(X\beta, \sigma^2 I_n))$).

$H_0 : \sigma^2 = \sigma_0^2$, $\alpha > 0$ vorgegeben, $\text{rk } X = r \leq p$

$T := \frac{\|Y - X\hat{\beta}\|^2}{\sigma_0^2} \sim \chi_{n-r}^2$ (unter H_0) Kritischer Bereich für T :

$$K^* = [0, \chi_{n-r, \alpha/2^2}, \cup] [\chi_{n-r, 1-\alpha/2}, \infty)$$

Konfidenzschätzung für den Vektor der schätzbaren Funktionen

$l(\beta) = (l_1(\beta), \dots, l_q(\beta))^T$, $1 \leq q \leq r \leq p < n$, $l_1(\beta), \dots, l_q(\beta)$ lin. unabh.

$$\mathbb{P}((l(\hat{\beta}) - l(\beta))^T (A_* A_*^T)^{-1} (l(\hat{\beta}) - l(\beta))) \leq \frac{q}{n-r} \|Y - X\hat{\beta}\|^2 \cdot F_{q, n-r, 1-\alpha} = 1 - \alpha$$

Vergleich von Erwartungswerten von p Stufen (Populationen), $\mathcal{N}(\mu, \sigma^2)$ -verteilter unabhängiger Beobachtungen ($i = 1, \dots, p$)
Versuchsplan:

	1	2	...	n_i
1	y_{11}	y_{12}	\cdots	y_{1, n_1}
2	y_{21}	y_{22}	\cdots	y_{2, n_2}
\vdots	\vdots	\vdots		\vdots
p	y_{p1}	y_{p2}	\cdots	y_{p, n_p}

Mathematisches Modell: $Y_{ik} = \mu_i + \epsilon_{ik}$, $n = 1, \dots, n_i$, $i = 1, \dots, p$, $\beta = (\mu_1, \dots, \mu_p)^T$, $\epsilon_{ik} \sim \mathcal{N}(0, \sigma^2)$ i. i. d.

Wichtig: Prüfen, ob tatsächlich die Varianz von ϵ_{ik} gleich sind (Bartlett-Test).

Normalengleichung: $X^T X \beta = X^T Y$

$$X^T X = \begin{pmatrix} n_1 & 0 & 0 & 0 \\ 0 & n_2 & 0 & 0 \\ & 0 & \ddots & 0 \\ 0 & 0 & 0 & n_p \end{pmatrix}$$

$$X^T Y = \begin{pmatrix} Y_{1\bullet} = \sum_{k=1}^{n_1} Y_{1k} \\ \vdots \\ Y_{p\bullet} = \sum_{k=1}^{n_p} Y_{pk} \end{pmatrix}$$

$$\hat{\mu}_i = \bar{Y}_{i\bullet} = \frac{1}{n_i} \sum_{k=1}^{n_i} Y_{ik} \quad \text{für } i = 1, \dots, p$$

Schätzung der Modellstreuung:

$$\hat{\sigma}^2 = \frac{1}{n-p} \|Y - X\hat{\beta}\|^2 = \frac{1}{n-p} \sum_{i=1}^p \sum_{k=1}^{n_i} (Y_{ik} - \bar{Y}_{i\bullet})^2$$

Dann ist $(n-p) \frac{\hat{\sigma}^2}{\sigma^2} \sim \chi_{n-p}^2$

Test auf Gleichheit der Erwartungswerte:

$H_0 : \mu_1 = \mu_2 = \dots = \mu_p$

Testgröße: $T = \frac{n-p}{p-1} \frac{S_1^2 - S_0^2}{S_0^2} \sim F_{p-1, n-p}$

Ablehnung von H_0 , falls $T \geq T_{p-1, n-p, 1-\alpha}$

$$C = \begin{pmatrix} 1 & -1 & 0 & & \\ 0 & 1 & -1 & 0 & \\ & & \ddots & \ddots & \\ & & & 1 & -1 \end{pmatrix}$$

$$S_0^2 := \|Y - X\hat{\beta}\|^2 = (n-p)\hat{\sigma}^2 = \sum_{i=1}^p \sum_{k=1}^{n_i} (Y_{ik} - \bar{Y}_{i\bullet})^2$$

$$S_1^2 = \min_{\beta: C\beta=0} \|Y - X\beta\|^2$$

Berechnung von S_1^2 mittels Lagrange-Multiplikatoren ergibt:

$$S_1^2 = \sum_{i=1}^p \sum_{k=1}^{n_i} (Y_{ik} - \bar{Y}_{i\bullet})^2$$

$$S_0^2 = \dots = \sum_{i=1}^p n_i (\bar{Y}_{i\bullet} - \bar{Y}_{\bullet\bullet})^2$$

Sprechweise. Übliche Bezeichnungen:

$S_1^2 - S_0^2 = \text{SQA} = \text{Summe der Quadrate der Abweichungen}$

$S_0^2 = \text{SQR} = \text{Summe der Quadrate der Abweichungen innerhalb der Stufen}$

Bartlett-Test: Prüfen der Hypothese, dass p unabhängige ZGn

$X_i \sim \mathcal{N}(\mu_i, \sigma^2)$, $i = 1, \dots, p$ dieselbe Streuung σ^2 besitzen.

$H_0 : \sigma_1^2 = \dots = \sigma_p^2 = \sigma^2$ Gegeben seien p unabhängige Stichproben

X_{i1}, \dots, X_{in_i} , $i = 1, \dots, p$, $n = n_1 + \dots + n_p$.

Testgröße: $T_{n_1, \dots, n_p} = \frac{1}{D} \left((n-p) \log S^2 - \sum_{i=1}^p (n_i - 1) \log S_i^2 \right)$ mit

$$D := 1 + \frac{1}{3(-1)} \left(\sum_{i=1}^p \frac{1}{n_i - 1} - \frac{1}{n-p} \right), \quad S_i^2 = \frac{1}{n_i - 1} \sum_{k=1}^{n_i} (X_{ik} - \bar{X}_{i\bullet})^2,$$

$$S^2 = \frac{1}{n-p} \sum_{i=1}^p S_i^2$$

Für $\min(n_1, \dots, n_p) \rightarrow \infty$ gilt $T_{n_1, \dots, n_p} \xrightarrow{d} \chi_{p-1}^2$.

Faustregel: $\min(n_1, \dots, n_p) \geq 5$

Ablehnung von H_0 , falls $T_{n_1, \dots, n_p} > \chi_{p-1, 1-\alpha}^2$.

Zweifache Varianzanalyse (Zweiwegklassifikation, Kreuzwegklassifikation)

Wirkung eines Faktors A in p Stufen und Wirkung eines Faktors B in q Stufen mit s Wiederholungen in jeder Stufe von Faktor A und B .

Bsp. Faktor A: Düngemittel Faktor B: Bodenart Y_{ijk} : Ernteertrag in Stufe i von Faktor A, in Stufe j von Faktor B in k -ter Wiederholung. Es geht um den Vergleich von Mittelwerten bei eventueller Wechselwirkung zwischen den Stufen der Faktoren.

Modell. $Y_{ijk} = \mu_0 + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijk}$ α_i : mittlerer Effekt in Stufe i von Faktor A β_j : mittlerer Effekt in Stufe j von Faktor B γ_{ij} : mittlerer Effekt aus Wechselwirkung von Stufe i und Stufe j Voraussetzung: Alle Beobachtungen sind normalverteilt und unabhängig mit $\mathbb{E}\epsilon_{ijk} = 0$, $\mathbb{E}\epsilon_{ijk}^2 = \sigma^2$ (eventuell Prüfung mit Bartlett-Test)

Prüfen der folgenden Hypothese: $H_A : \alpha_1 = \dots = \alpha_p = 0$,
 $H_B : \beta_1 = \dots = \beta_q = 0$, $H_{AB} : \gamma_{11} = \dots = \gamma_{pq} = 0$

Schreibweise als lineares Modell:

$$Y = X\beta \text{ mit } \dim Y = p \dots q \cdot s, X \in \mathbb{R}^{pq \cdot s \times (1+p+q+pq)}.$$

$(p+1) \cdot (q+1)$ Parameter,

$$\beta = (\mu_0, \alpha_1, \dots, \alpha_p, \beta_1, \dots, \beta_q, \gamma_{11}, \dots, \gamma_{pq})$$

$$\text{rk } X = pq - 1 < \min\{pq, (p+1) \cdot (q+1)\}$$

Reparametrisierung ist notwendig, d. h. es werden Gleichungen zwischen den Parametern hinzugefügt, die die eindeutige Lösbarkeit von (N) garantieren:

$$\alpha_{\bullet} = 0, \quad \beta_{\bullet} = 0, \quad \gamma_{1\bullet} = \dots = \gamma_{p\bullet} = 0, \quad \gamma_{\bullet 1} = \dots = \gamma_{\bullet q} = 0.$$

Wegen $pq - 1 + (2 + p + q) = (p+1) \cdot (q+1)$ kann die Eindeutigkeit der MkQ-Schätzung gesichert werden. Diese Bedingungen bedeuten keine Einschränkung der Allgemeinheit der Darstellung

$Y_{ijk} = \mu_0 + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijk}$ denn:

$$\begin{aligned} \mu_0^* &= \mu_0 + \bar{\alpha}_{\bullet} + \bar{\beta}_{\bullet} + \bar{\gamma}_{\bullet\bullet}, \alpha_i^* = \alpha_i - \bar{\alpha}_{\bullet} + \bar{\gamma}_{i\bullet} - \bar{\gamma}_{\bullet\bullet}, \\ \beta_j^* &= \beta_j - \bar{\beta}_{\bullet} + \bar{\gamma}_{\bullet j} - \bar{\gamma}_{\bullet\bullet} \text{ und } \gamma_{ij}^* = \gamma_{ij} - \bar{\gamma}_{i\bullet} - \bar{\gamma}_{\bullet j} + \bar{\gamma}_{\bullet\bullet} \text{ für } \\ i &= 1, \dots, p, j = 1, \dots, q. \end{aligned}$$

Es ergeben sich die Gleichungen

$$Y_{ijk} = \mu_0^* + \alpha_i^* + \beta_j^* + \gamma_{ij}^* + \epsilon_{ijk} \text{ für } i = 1, \dots, p, j = 1, \dots, q, k = 1, \dots, s$$

Bezeichnungen für Mittelwert:

$$\hat{\mu}_0 = \bar{Y}_{\bullet\bullet\bullet}, \hat{\alpha}_i = \bar{Y}_{i\bullet\bullet} - \bar{Y}_{\bullet\bullet\bullet}, \hat{\beta}_j = \bar{Y}_{\bullet j\bullet} - \bar{Y}_{\bullet\bullet\bullet}$$

$$\hat{\gamma}_{ij} = \bar{Y}_{ij\bullet} - \bar{Y}_{i\bullet\bullet} - \bar{Y}_{\bullet j\bullet} + \bar{Y}_{\bullet\bullet\bullet}$$

Zum Prüfen der Hypothese H_A, H_B, H_{AB} verwendet man folgende Testgrößen:

$$S_{pqs}^2 := \sum_{i=1}^p \sum_{j=1}^q \sum_{k=1}^s (Y_{ijk} - \bar{Y}_{ij\bullet})^2,$$

$$F_A := \frac{pq \cdot (s-1)}{p-1} \frac{qs \sum_{i=1}^p (\bar{Y}_{i\bullet\bullet} - \bar{Y}_{\bullet\bullet\bullet})^2}{S_{pqs}^2} \sim F_{p-1, pq(s-1)}$$

$$F_B := \frac{pq \cdot (s-1)}{q-1} \frac{ps \sum_{i=1}^q (\bar{Y}_{\bullet j\bullet} - \bar{Y}_{\bullet\bullet\bullet})^2}{S_{pqs}^2} \sim F_{q-1, (pq(s-1))}$$

$$F_{AB} := \frac{pq(s-1)}{(p-1)(q-1)} \frac{s \sum_{i=1}^p \sum_{j=1}^q (\bar{Y}_{ij\bullet} - \bar{Y}_{i\bullet\bullet} - \bar{Y}_{\bullet j\bullet} + \bar{Y}_{\bullet\bullet\bullet})^2}{S_{pqs}^2} \sim F_{(p-1)(q-1), pq(s-1)}$$

Entscheidungsregel: Die Hypothesen H_A, H_B bzw. H_{AB} werden abgelehnt, falls

$$F_A > F_{p-1, pq(s-1), 1-\alpha}, \quad F_B > F_{q-1, pq(s-1), 1-\alpha} \text{ bzw. } F_{AB} > F_{(p-1)(q-1), pq(s-1), 1-\alpha}.$$

Bem. Die Anzahl der Wiederholungen kann auch in den einzelnen Stufen variieren.

Problem der Ausgleichsrechnung:

Zu n Messwerten $(x_1, y_1), \dots, (x_n, y_n)$ soll eine Funktion $\mu(x)$ gefunden werden, deren Funktionswerte $\mu(x_i)$ die y_i möglichst gut approximieren. Wir betrachten ein nichtparametrisches Regressionsmodell (Modell I)

$$Y_i = \mu(x_i) + \epsilon_i, \quad i = 1, \dots, n(*)$$

mit unbekannter Regressionsfunktion $\mu : [a, b] \rightarrow \mathbb{R}^1$. (Im Modell II werden die x_i 's durch ZGn X_i ersetzt)

Annahme: $a \leq x_1 < x_2 < \dots < x_n \leq b$ (bewählte Messstellen) und $\mathbb{E}\epsilon_i = 0, i = 1, \dots, n, \mathbb{E}\epsilon_i \epsilon_j = \sigma^2 \delta_{ij}$ für $i, j = 1, \dots, n$.

Kernschätzer für $\mu(x), a \leq x \leq b$

Kernfunktion: $K \in L^1(\mathbb{R}), \int K(x) dx = 1, K(-x) = K(x), \text{supp } K$ beschränkt, $\sup|K(x)| < \infty$. Bandweite: $h_n \rightarrow 0, h \cdot n \rightarrow \infty$.

$$\hat{\mu}_1(x) = \frac{1}{nh_n} \sum_{i=1}^n Y_i \int_{s_{i-1}}^{s_i} K\left(\frac{x-s}{h_n}\right) ds$$

mit $s_0 = a, s_{i-1} < x_i \leq s_i, s_n = b$. heißt **Gasser-Müller-Schätzer**.

$$\text{Verbessert: } \hat{\mu}_n(x) = \frac{b-a}{nh_n} \sum_{i=1}^n Y_i K\left(\frac{x-x_i}{h_n}\right).$$

Motivation:

$$\mathbb{E}\hat{\mu}_n(x) = \frac{1}{nh_n} \sum_{i=1}^n \mathbb{E}Y_i \int_{s_{i-1}}^{s_i} K\left(\frac{x-s}{h_n}\right) ds = \frac{1}{n} \sum_{i=1}^n \mu(x_i) \int_{s_{i-1}/h_n}^{s_i/h_n} K\left(\frac{x}{h_n} - \right)$$

$$t) dt = \frac{1}{h} \sum_{i=1}^n \mu(x_i) \int_{(x-s_{i-1})/h_n}^{(x-s_i)/h_n} K(y) dy \approx \mu(x)$$

Kernschätzer für Modell II: $(X_1, Y_1), \dots, (X_n, Y_n)$ i. i. d. mit Dichte $f_{(X,Y)}, K$ sei eine Kernfunktion wie oben und (h_n) eine Bandbreitenfolge.

$$\hat{\mu}_n(x) = \frac{\sum_{i=1}^n K\left(\frac{x-X_i}{h_n}\right)}{\sum_{i=1}^n K\left(\frac{x-X_i}{h_n}\right)}, \quad a \leq x \leq b$$

Nadaraya-Watson-Schätzer.

Bedingte Erwartung:

$$\mathbb{E}(Y - \mathbb{E}(Y|\tilde{\sigma}))^2 = \min_{Z \mathcal{F}\text{-messbar}, \mathbb{E}Z^2 < \infty} \mathbb{E}(Y - Z)^2, \mathcal{F} \subseteq \mathfrak{A}, (\Omega, \mathfrak{A}, \mathbb{P})$$

Falls $\mathcal{F} = \sigma(X) =$ von X erzeugte σ -Algebra, so schreiben wir $\mathbb{E}(Y|X)$ anstelle $\mathbb{E}(Y|\sigma(X))$ und es gilt: eine Borel-messbare Funktion $g: \mathbb{R} \rightarrow \mathbb{R}$ mit $\mathbb{E}(Y|X) = g(X)$ (Faktorisierungslemma). Wir können dann schreiben: $g(x) = \mathbb{E}(Y|X = x) = \int y f_{Y|X}(y|x) dy$, wobei $f_{Y|X}(y|x) =$ bedingte Dichte von Y bei gegebenen $X = x$

$$= \frac{f_{(X,Y)}(x, y)}{f_X(x)} \text{ mit } f_X(x) = \int f_{(X,Y)}(x, y) dy.$$

Motivation für (i_1) :

$$\mu(x) = \mathbb{E}(Y|X = x) = \int y \frac{f_{(X,Y)}(x, y)}{f_X(x)} dy \cdot f_X(x) = \int f_{(X,Y)}(x, y) dy$$

Schätzung von $f(x, y)$ mittels Produktkern: $K(x, y) = K_1(x) \cdot K_2(y)$ mit K_1, K_2 wie oben.

$$\hat{f}_n(x, y) := \frac{1}{nh_n^2}$$

$$\hat{f}_n(x) = \frac{1}{nh_n}$$

$$\hat{\mu}_n = \int y \frac{\hat{f}_n(x, y)}{\hat{f}_n(x)} dy = \frac{1}{\hat{f}_n(x)} \frac{1}{nh_n^2} \sum_{i=1}^n K_1\left(\frac{x-X_i}{h_n}\right) \int y K_2\left(\frac{y-Y_i}{h_n}\right) dy \xrightarrow{n \rightarrow \infty}$$

Eigenschaften des Gasser-Müller-Schätzers:

$$\text{MASE}(\hat{\mu}_n) = \frac{1}{n} \sum_{i=1}^n \text{MSE}(\hat{\mu}_n(x_i)),$$

$$\text{MSE}(\hat{\theta}_n) = \mathbb{E}(\hat{\theta}_n - \theta)^2 = \text{Var}(\hat{\theta}_n) + (\mathbb{E}(\hat{\theta}_n - \theta))^2$$

Satz. Unter den Bedingungen (*), (**), $K(x) = 0$ für $|x| > 1, a = 0, b = 1, |K(x) - K(x')| \leq L|x - x'|, h_n \downarrow 0, nh_n \rightarrow \infty, \mu \in C^2([0, 1]), x = (i - \frac{1}{2})/n, i = 1, \dots, n$. Dann gilt:

$$\mathbb{E}\hat{\mu}_n(x) = \mu(x) + \frac{h_n^2}{2} \int x^2 K(x) dx \mu''(x) + o(h_n^2) + O\left(\frac{1}{n}\right)$$

$$\text{MSE}(\hat{\mu}(x)) = \frac{\sigma^2}{nh_n} \int K^2(x) dx + \frac{h^4}{4} \int x^2 K(x) dx (\mu''(x))^2 + o\left(\frac{1}{nh_n}\right) + o(h_n^4)$$

Dichteschätzungen

Notation. Sei \mathcal{P} die Menge aller bezüglich des Lebesgue-Maß λ_1 absolut stetigen Wahrscheinlichkeitsmaße auf \mathbb{R}^1 und

$$\mathcal{F}_c := \{f \in \mathcal{C}(\mathbb{R}^1) \mid f = dP/d\lambda_1 \text{ für ein } P \in \mathcal{P}\}$$

die Menge der stetigen W-Dichtefunktionen.

Ziel. Finden einer „guten“ Dichteschätzung $\hat{f}_n(X, -) : \mathbb{R}^1 \rightarrow \mathbb{R}^1$, wobei $X = (X_1, \dots, X_n)$ eine math. Stichprobe ist, in Form einer Borel-messbaren Abbildung $\hat{f}_n(-, -) : \mathbb{R}^n \times \mathbb{R}^1 \rightarrow \mathbb{R}^1$.

Notation. $\hat{f}_n(t) := \hat{f}_n(X_1, \dots, X_n, t)$

Lem. Es gibt keinen Dichteschätzer $\hat{f}_n(-)$ mit

$$\mathbb{E}_f \hat{f}_n(t) = f(t) \quad \text{für } \lambda_1\text{-fast alle } t \text{ für alle } f \in \mathcal{F}_c.$$

Def. Sei $x_0 \in \mathbb{R}$ und $h > 0$ fest. Setze $I_j := [x_0 + jh, x_0 + (j+1)h)$ für $j \in \mathbb{Z}$. Das **Histogramm** ist der (naive) Dichte-Schätzer

$$\hat{f}_n(t) := \frac{1}{nh} \sum_{i=1}^n \mathbb{1}_{I_j}(X_i), \quad \text{wobei } j \in \mathbb{Z} \text{ so ist, dass } t \in I_j.$$

Bem. Der Graph der geschätzten Histogramm-Dichte ist ein Säulendiagramm. Verbindet man die Mitten der Säulen mit einer Linie, so bekommt man einen **Häufigkeitspolygonzug**.

Bem. Nach dem Gesetz der großen Zahlen gilt

$$\hat{f}_n(t) \xrightarrow[n \rightarrow \infty]{\mathbb{P}_f\text{-f.s.}} h^{-1} \int_{I_j} f(x) dx \quad \text{für } t \in I_j.$$

Def. Sei $\hat{f}_n(-)$ ein Dichteschätzer und $f \in \mathcal{F}_c$. Dann heißt

$$\begin{aligned} \text{MISE}(\hat{f}_n) &:= \Delta_n := \mathbb{E}_f \int_{\mathbb{R}} (\hat{f}_n(t) - f(t))^2 dt \\ &= \int_{\mathbb{R}} \underbrace{\mathbb{E}_f (\hat{f}_n(t) - f(t))^2}_{=: \text{MSE}(\hat{f}_n(t))} dt \end{aligned}$$

MISE (mean integrated squared error) von \hat{f}_n bzgl. f .

Satz (Freedman, Diaconis). Sei $f : \mathbb{R}^1 \rightarrow \mathbb{R}^1$ eine Dichtefkt mit

(i) $f \in L^2(\mathbb{R}^1)$ und f absolut stetig, d. h. f. ü. differenzierbar

(ii) $f' \in L^2(\mathbb{R}^1)$ und f' absolut stetig, d. h. f. ü. differenzierbar

(iii) $f'' \in L^p(\mathbb{R}^1)$ für ein $p \in [1, 2]$

Wir schreiben

$$\alpha := \sqrt[3]{6} \cdot \gamma^{-1/3}, \quad \beta := \frac{3}{2\sqrt[3]{6}} \gamma^{1/3}, \quad \gamma := \int_{\mathbb{R}^1} (f'(t))^2 dt.$$

Dann gilt $\min_{h=h_n > 0} \Delta_n^2 = \frac{\beta}{n^{2/3}} + O\left(\frac{1}{n^{2/3}}\right)$.

Das Minimum wird angenommen für $h_n = \frac{\alpha}{\sqrt[3]{n}} + O\left(\frac{1}{\sqrt[3]{n}}\right)$.

Falls nur (i) und $\gamma > 0$ erfüllt ist, so gilt

$\min_{h=h_n > 0} \Delta_n^2 = \frac{\beta}{n^{2/3}} + o\left(\frac{1}{n^{2/3}}\right)$ und das Minimum wird bei

$h_n = \frac{\alpha}{\sqrt[3]{n}} + o\left(\frac{1}{\sqrt[3]{n}}\right)$ angenommen.

Def. Sei $K \in L^1(\mathbb{R})$ eine Fktn mit $\int K(t) dt = 1$ und $(h_n)_{n \in \mathbb{N}}$ eine Folge in $(0, \infty)$ mit $h_n \downarrow 0$. Dann heißt

$$\hat{f}_n(t) = \hat{f}_n(X_1, \dots, X_n; t) := \frac{1}{n \cdot h_n} \sum_{i=1}^n K\left(\frac{X_i - t}{h_n}\right)$$

Kern dichteschätzer für f mit **Kernfunktion** K .

Bspe. Mit der **empirischen Dichte** $K(x) := \frac{1}{2} \mathbb{1}_{(-1,1]}(x)$ gilt

$$\hat{f}_n(t) = \frac{1}{2h_n} (\hat{F}_n(t + h_n) - \hat{F}_n(t - h_n)),$$

mit dem **Gauß-Kern** $K(x) := \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-x^2}{2}\right)$ gilt $\hat{f}_n(-) \in \mathcal{C}^\infty(\mathbb{R})$.

Lem. Sei zusätzlich $K \in L^2(\mathbb{R})$. Dann gilt:

- $\mathbb{E}_f \hat{f}_n(t) = \int K(x) f(t + h_n x) dx$
- $\text{Var}_f(\hat{f}_n(t)) = \frac{1}{n \cdot h_n} \cdot \int K^2(x) \cdot f(t + h_n x) dx - \frac{1}{n} \cdot \left(\int K(x) \cdot f(t + h_n x) dx\right)^2$

Satz. Sei f eine beschränkte W-Dichtefktn, $\forall x \in \mathbb{R} : f(x) \leq M$, mit Stetigkeitsstellen $C \subseteq \mathbb{R}$. Sei $K \in L^2(\mathbb{R})$ und $(h_n)_{n \in \mathbb{N}}$ eine Folge mit $h_n \downarrow 0$ für $n \rightarrow \infty$. Angenommen, $n \cdot h_n \xrightarrow[n \rightarrow \infty]{} \infty$. Dann gilt:

- $\mathbb{E}_f \hat{f}_n(t) \xrightarrow[n \rightarrow \infty]{} f(t) \quad \forall t \in C,$
- $n \cdot h_n \cdot \text{Var}_f(\hat{f}_n(t)) \xrightarrow[n \rightarrow \infty]{} f(t) \cdot \int K^2(x) dx \quad \forall t \in C,$

- $\sup_{t \in \mathbb{R}} \mathbb{E}_f (\hat{f}_n(t) - f(t))^2 \xrightarrow[n \rightarrow \infty]{} 0$, falls f glm. stetig auf \mathbb{R} .

Satz. Sei $f \in \mathcal{C}^2(\mathbb{R}^1)$, $f, |f'|, |f''| \leq M < \infty$ und $\int x^2 |K(x)| dx < \infty$. Dann gilt für alle $t \in \mathbb{R}^1$:

$$\mathbb{E}_f \hat{f}_n(t) = f(t) + h_n f'(t) \int x K(x) dx + \frac{1}{2} h_n^2 f''(t) \int x^2 K(x) dx + o(h_n^2).$$

wobei die Konvergenz von $o(-)$ sogar gleichmäßig in abgeschlossenen t -Intervallen ist.

Ziel. Bestimmung einer optimalen Bandbreite h_n

Satz. Sei $f \in \mathcal{C}^2(\mathbb{R}^1)$ mit $f'' \in L^2(\mathbb{R}^1)$ und $f, |f'| \leq A < \infty$. Für K gelte $0 \leq K(x) \leq B$, $K(-x) = K(x)$ und $\int x^2 K(x) dx < \infty$. Dann gilt für $(h_n)_{n \in \mathbb{N}}$ mit $h_n \downarrow 0$ und $nh_n \xrightarrow[n \rightarrow \infty]{} \infty$:

$$\begin{aligned} \text{MISE}(\hat{f}_n) &= \frac{1}{nh_n} \int K^2(x) dx + \frac{h_n^4}{4} \left(\int x^2 K(x) dx\right)^2 \int (f''(t))^2 dt \\ &\quad + o(h_n^4) + o\left(\frac{1}{nh_n}\right) \quad \text{für } n \rightarrow \infty. \end{aligned}$$

Optimale Wahl von h_n :

Gleichsetzen der beiden echten Fehlerterme:

$$\frac{1}{nh_n} \int K^2(x) dx = \frac{h_n^4}{4} \left(\int x^2 K(x) dx\right)^2 \int (f''(t))^2 dt$$

Es folgt $h_n^* = \frac{c^*}{n^{1/5}}$ mit $c^* := \left(\frac{4 \int K(x)^2 dx}{\left(\int x^2 K(x) dx\right)^2 \int (f''(t))^2 dt}\right)^{1/5}$

und $\text{MISE}(\hat{f}_n) =$

$$\frac{\sqrt[5]{8}}{n^{4/5}} \left(\int (f''(t))^2 dt\right)^{1/5} \left(\int x^2 K(x) dx\right)^{2/5} \left(\int K(x)^2 dx\right)^{4/5} + o(n^{-4/5})$$

Optimale Wahl der Kernfunktion K :

$$\int x^2 K(x) dx \left(\int K(x)^2 dx\right)^2 \rightarrow \text{Min}$$

unter der Nebenbedingung $\int K(x) dx = 1$ und $K(x) \geq 0$.

Satz. Sei K eine Kernfunktion mit $\int K(x) dx = 1$. Dann gilt für $\alpha > 0$:

$$\left(\int K(x)^2 dx\right)^\alpha \int |x|^\alpha |K(x)| dx \geq \frac{1}{2\alpha+1} \left(\frac{\alpha+1}{2\alpha+1}\right)^\alpha$$

Gleichheit gilt für $K_0(x) = \frac{\alpha+1}{2\alpha} (1 - |x|^\alpha) \mathbb{1}_{[-1,1]}(x)$

Def. Für $\alpha = 2$ heißt $K_0(x) := \frac{3}{4} (1 - x^2) \mathbb{1}_{[-1,1]}(x)$ **Epanetschnikow-Kern**.