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** $x_1, \ldots, x_n \in \mathbb{R}^1$ von X (d. h. $x_1, \ldots, x_n$ sind Realisierung von iid ZGen $X_1, \ldots, X_n \sim P_X$) 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 $(\mathbb{R}^n, \mathfrak{B}(\mathbb{R}^n), P_\vartheta \times \ldots \times P_\vartheta)$.

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 $\varphi : \mathbb{R}^n \to \{0, 1\}$ augedrückt wird und zwar durch

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

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

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

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

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

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

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

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

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

heißt **α-Test** o. **Signifikanztest** zum **Signifikanzniveau** $\alpha \in (0, 1)$.
Ein α-Test φ heißt **unverfälscht** (erwartungstreu, unbiased), falls

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

Situation. Sei nun eine Stichprobenfunktion oder **Teststatistik** $T : \mathbb{R}^n \to \mathbb{R}^1$ 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

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

wobei f_T die Dichte von $T(X_1, \ldots, X_n)$ 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

$$T(X_1, \ldots, X_n) := \tfrac{\sqrt{n}}{\sigma}\left(\overline{X}_n - \mu_0\right) \quad \text{mit } \overline{X}_n := \tfrac{1}{n}\left(X_1 + \ldots + X_n\right).$$

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

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

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

$$\hat{T}(X_1, \ldots, X_n) = \tfrac{\sqrt{n}}{S_n}\left(\overline{X}_n - \mu_0\right), \quad S_n^2 := \tfrac{1}{n-1} \sum_{i=1}^n \left(X_i - \overline{X}_n\right)^2.$$

Dabei ist S_n die **(korrigierte) Stichprobenvarianz**. Man kann zeigen, dass $\hat{T}(X_1, \ldots, X_n) \sim t_{n-1}$ 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 $\overline{X}_n$ sind unabhängig für $n \ge 2$.

Diskussion.
- Je kleiner α ist, desto „nullhypothesenfreundlicher" 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 $\overline{x}_n$ vorliegen. Es ist dann $K_n^T = (z_{1-\alpha}, \infty)$.

Def. Es seien $X_1, \ldots, X_n \sim \mathcal{N}(0, 1)$. Dann heißt die Summe $X_1^2 + \ldots + X_n^2 \sim \chi_n^2$ **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

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

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

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

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

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

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

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

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

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

Test. Seien $\binom{X_1}{Y_1}, \ldots, \binom{X_n}{Y_n} \sim \mathcal{N}\left(\binom{\mu_1}{\mu_2}, \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho \\ \sigma_1\sigma_2\rho & \sigma_2^2 \end{pmatrix}\right)$.
Wir testen $H_0 : \rho = 0$ **vs.** $H_1 : \rho \neq 0$ mit

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

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

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

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 $X_n \xrightarrow[n\to\infty]{\mathbb{P}} c = \text{const}$ (d. h. $\forall \epsilon > 0 : \mathbb{P}(|X_n - c| > \epsilon) \to 0$) und $Y_n \xrightarrow[n\to\infty]{d} Y$ (d. h. $\mathbb{P}(Y_n \leq y) \to \mathbb{P}(Y \leq y)$ für alle Stetigkeitspunkte y der VF $y \mapsto \mathbb{P}(Y \leq y)$). Dann gilt:

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

und allgemeiner $f(X_n, Y_n) \xrightarrow[n\to\infty]{d} f(c, Y)$ 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 $T_n = T(X_1, \ldots, X_n)$ eine Statistik. Falls der ZGWS für T_n die Form

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

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 $f : \Theta \to \mathbb{R}^1$, sodass

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

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

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

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

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

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

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

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

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

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

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

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

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

Bem. Bei $k = 2$ schreibt man oft

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

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

$$\varphi : \mathbb{R}^k \to \mathbb{R}, \ \ t \mapsto \mathbb{E}e^{i\langle t, X\rangle} = \int_{\mathbb{R}^k} e^{i(t_1x_1 + \ldots + t_kx_k)} \,\mathrm{d}F_X(x_1, \ldots, x_k).$$

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

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

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 $x_1, \ldots, x_n$ 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,

$$\mathbb{R} = \bigcup_{i=1}^{s+1} I_j \quad \text{mit} \quad I_j := (y_{j-1}, y_j], \quad \text{wobei}$$

$$-\infty = y_0 < y_1 < \ldots < y_s < y_{s+1} = +\infty.$$

Wir setzen

$$h_{n_j} := |\{k \in \{1, \ldots, n\} \mid X_k \in I_j\}| \qquad (\text{absolute Klassenhäufigkeit})$$

$$p_j^{(0)} := \mathbb{P}(X \in I_j) = F_0(y_j) - F_0(y_{j-1}) \qquad (\text{Klassenwktn unter } H_0)$$

Die Klassenhäufigkeiten sind multinomialverteilt unter H_0:

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

Als (näherungsweises) 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. $T_{n,s+1} \xrightarrow[n\to\infty]{d} \chi_s^2$

Faustregel. Für $np_j^{(0)} \geq 5$, $j = 1, \ldots, s+1$ 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 $T_{n,s+1} > \chi^2_{s,1-\alpha}$.

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 $\vartheta = (\vartheta_1, \ldots, \vartheta_r)$ der VF F_0 durch MLE zu schätzen, also durch

$$\hat{\vartheta}_n := \arg\max L(h_{n_1}, \ldots, h_{n_{s+1}}; \vartheta), \quad \text{wobei}$$

$$L(h_{n_1}, \ldots, h_{n_{s+1}}; \vartheta) := \prod_{j=1}^{s+1} \left(p_j^{(0)}\right)^{h_{n_j}}.$$

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

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

wobei r die Anzahl der geschätzten Parameter ist.

- Manchmal wird die Parameter-Schätzung auch direkt aus der SP $x_1, \ldots, x_n$ ermittelt (z. B. $\tilde{\mu}_n := \frac{1}{n}(x_1 + \ldots + x_n)$ 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 $I_1, \ldots, I_k \subset \mathbb{R}^{n_1}$ und $J_1, \ldots, J_l \subset \mathbb{R}^{n_2}$ jeweils Familien paarweise disjunkter Mengen mit $\mathbb{P}(X \in I_1 \cup \ldots \cup I_k) = 1$ bzw. $\mathbb{P}(Y \in J_1 \cup \ldots \cup J_l) = 1$. Wir setzen

$$p_{ij} := \mathbb{P}((X, Y) \in I_i \times J_j) = \mathbb{P}(\{X \in I_i\} \cap \{X_j \in J_j\}),$$

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

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 $(X_1, Y_1), \ldots, (X_n, Y_n)$:

$$h_{ij}^{(n)} := |\{m \in \{1, \ldots, n\} \mid (X_m, Y_m) \in I_i \times J_j\}|,$$

$$h_{i\bullet} := \sum_{j=1}^l h_{ij}, \quad h_{\bullet j} := \sum_{i=1}^k h_{ij}.$$

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

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

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

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

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

Entscheidungsregel. H_0 wird genau dann abgelehnt, falls

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

Bemn.
- Zum Testen eines höherdim. ZV $(X_1, \ldots, X_r)$ auf Unabhängigkeit aller Komponenten untersuchen wir die Ereignisse

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

  für eine passende Intervalleinteilung. Wir verwenden dann

  $$\hat{T}_{k_1,\ldots,k_r}^{(n)} \coloneqq n^{r-1} \sum_{i_1=1}^{k_1} \cdots \sum_{i_r=1}^{k_r} \frac{\left(h_{i_1\cdots i_r}^{(n)} - n^{-r+1} \prod_{j=1}^{r} h_{\bullet\cdots i_j \cdots\bullet}^{(n)}\right)^2}{\prod_{j=1}^{r} h_{\bullet\cdots i_j\cdots\bullet}}$$
  $$\xrightarrow[n\to\infty]{d} \chi^2_{k_1\cdots k_s - k_1 - \ldots - k_r + r - 1}$$

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

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

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

Kolmogorow-Smirnow-1SP-Test

Situation. Sei $X_1, \ldots, X_n \sim F$ eine math. SP. Wir sortieren die dabei gezogenen Werte aufsteigend: $X_{1:n} \leq X_{2:n} \leq \ldots \leq X_{n:n}$. Dann heißt $\hat{F}_n(x) \coloneqq \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{(-\infty,x]}(X_{i:n})$ **empirische VF**.

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

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

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

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

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

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

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

unabhängig von F.

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

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

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

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

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

Bemn.
- 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 $K_{1-\alpha} \approx \sqrt{-1/2 \cdot \log(\alpha/2)}$ für α 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

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

Für alle $z \in \mathbb{R}^1$ gilt

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

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

Def. $\omega_n^2(g) = n \int_{\mathbb{R}^1} g(F(x)) \left(\hat{F}_n(x) - F(x)\right)^2 \,\mathrm{d}F(x)$

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

Satz. Sei F stetig. Dann ist

$$\omega_n^2(g) \overset{\mathrm{d}}{=} n \int_0^1 g(u) \left(\hat{G}_n(u) - u\right)^2 \,\mathrm{d}u \xrightarrow[n\to\infty]{d} \int_0^1 g(u) (\dot{B}(u))^2 \,\mathrm{d}u \eqqcolon \omega^2(g).$$

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 $X_1, \ldots, X_n \sim F$ i. i. d. und $X_1^*, \ldots, X_m^* \sim F^*$ 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

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

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

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

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

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

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

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

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

a) $\mathbb{E}W_{m,n} = \frac{m \cdot n}{2}$ b) $\operatorname{Var} W_{m,n} = \frac{m \cdot n}{12}(m+n+1)$

c) $g_{m,n}(z) := \sum_{k=0}^{n\cdot m} \mathbb{P}(W_{m,n} = k) \cdot z^k = \frac{z^{-m(m+1)/2}}{\binom{m+n}{m}} \sum_{1 \le \nu_1 < \dots < \nu_m \le m+n} z^{\nu_1 + \dots + \nu_m} = \frac{1}{\binom{m+n}{m}} \prod_{k=1}^m \frac{1 - z^{n+k}}{1 - z^k}$

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

$$c_{\alpha/2} = \min\{k \ge 0 \,|\, \mathbb{P}(W_{m,n} \le k) = \mathbb{P}(W_{m,n} \ge m \cdot n - k) \ge \alpha/2\}.$$

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

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

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 $|T_{m,n}| \ge z_{1-\alpha/2}$.

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. $\overline{\nu}_i := \frac{1}{n_i}(\nu_{i,1} + \dots + \nu_{i,n_i})$, $\overline{\nu} := \frac{1}{n}\sum_{i=1}^k n_i \overline{\nu}_i$ mit $n := n_1 + \dots + n_k$.

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

$$H := \frac{12}{n(n+1)} \sum_{i=1}^k n_i \left(\overline{\nu}_i - \frac{n+1}{2}\right)^2 \xrightarrow[n_i\to\infty]{d} \chi^2_{k-1}$$

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

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

Theorie der U-Statistiken

Situation. Sei $n \ge m$, $X_1,\dots,X_n \sim F$ i. i. d., $h : \mathbb{R}^m \to \mathbb{R}^1$ 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

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

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

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

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

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

$$g(x) = \mathbb{E}[h(X_1,\dots,X_n)|X_1 = x] = \mathbb{E}h(x, X_2,\dots,X_m) = \int \cdots \int h(x, x_2,\dots,x_n)\,\mathrm{d}F(x_2)\cdots\mathrm{d}F(x_n).$$

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

(1) $\operatorname{Var}(U_n^{(m)} - \tilde{U}_n^{(m)}) = \operatorname{Var}(U_n^{(m)}) - \operatorname{Var}(\tilde{U}_n^{(m)})$

(2) $\mathbb{E}(U_n^{(m)}|X_i = x) = \theta + \frac{m}{n}(g(x) - \theta)$

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

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

$$\operatorname{Var}(U_n^{(m)}) = \frac{1}{\binom{n}{m}} \sum_{k=1}^m \binom{m}{k}\binom{n-m}{m-k} \cdot \zeta_k \quad \text{mit}$$

$$\begin{aligned} h_k(x_1,\dots,x_k) &:= \mathbb{E}(h(x_1,\dots,x_k, X_{k+1},\dots,X_m) \\ \zeta_k &:= \operatorname{Var}(h_k(X_1,\dots,X_k)) \\ &= \mathbb{E}[h(X_1,\dots,X_k,X_{k+1},\dots,X_m)\cdot \\ &\qquad h(X_1,\dots,X_k,X_{m+1},\dots,X_{2m-k})] - \theta^2 \end{aligned}$$

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

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

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

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

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

Bemn.
- Der Fall $\operatorname{Var}(g(X_1)) = 0$ (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 $E|h(X_1,\dots,X_m)|^{1+q} < \infty$ für $0 < q \le 1$ folgt
$$\mathbb{E}|\sqrt{n}(U_n^{(m)} - \tilde{U}_n^{(m)})|^{1+q} \le \frac{c(q,m)}{n^{2q}}\mathbb{E}|h(X_1,\dots,X_m)|^{1+q}.$$
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 $\mathbb{P}(\sqrt{n}|U_n^{(m)} - \tilde{U}_n^{(m)}| < \epsilon) \to 0$ 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 $S_n^2 = \binom{n}{2}^{-1} \sum_{i<j} \frac{1}{2}(X_i - X_j)^2$. Dann ist $g(x) = \frac{1}{2}(x - \mathbb{E}X_1)^2 + \frac{1}{2}\sigma^2$ mit $\sigma^2 := \operatorname{Var}(X_1)$. Es gilt

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

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 $\sqrt{n}(S_n^2 - \sigma^2) \xrightarrow[n\to\infty]{d} \mathcal{N}(0, 2\sigma^4)$. Es folgt

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

Alternativ erhält man durch Anwenden einer varianzstab. Trafo:

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

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

$$\operatorname{Cum}_m(X) = \frac{1}{m! 2^m} \frac{\partial^m}{\partial t^m}|_{t=0} \log \mathbb{E}e^{itX}.$$

Bem. Falls X_1, …, X_n unabhängig sind, so gilt

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

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

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

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

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 $\nu_1^+, \dots, \nu_n^+$ 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

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

Bsp. Alternativ können wir die U-Statistik

$$U_n = \frac{1}{\binom{n}{2}} \sum_{1 \le i<j \le n} \mathbb{1}_{\{Z_i + Z_j > 0\}}.$$

zum Test auf Symmetrie betrachten. Unter H_0 gilt

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

Aus dem ZGWS für U-Statistiken folgt

$$\sqrt{n}(U_n - \tfrac{1}{2}) \xrightarrow[n\to\infty]{d} \mathcal{N}(0, \tfrac{1}{3}).$$

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

Def. Sei $h : \mathbb{R}^{m_1} \times \mathbb{R}^{m_2} \to \mathbb{R}^1$ 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

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

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

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

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

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

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

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 $\sigma^2 := \mathrm{Var}(Y - g(X)) = \mathbb{E}\epsilon^2$.

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

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

Wir nehmen an, dass $\mathrm{Cov}(Y_i, Y_j) = \mathrm{Cov}(\epsilon_i, \epsilon_j) = \sigma^2\delta_{ij}$ 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 $\mathrm{Cov}(Y*) = I_n$.

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

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

$$\tfrac{\partial}{\partial\beta} S(y,\beta) = -2X^Ty + 2X^TX\beta, \quad \tfrac{\partial^2}{\partial\beta^2} S(y,\beta) = 2X^TX.$$

Für die Minima gelten die Normalengleichungen

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

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

Bsp (**Einfache lineare Regression**)**.**
Annahme: $Y_i = \beta_1 + \beta_2 x_i + \epsilon_i$, $i = 1, \dots, n$. Dann ist

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

Bsp (**Multiple lineare 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:

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

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

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

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

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

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

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 $c = X^Ta$ 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 $l(\beta)$
- $c \in \mathrm{im}(X^T) = \mathrm{im}(X^TX)$
- $l(\hat{\beta}) = c^T\hat{\beta}$ ist konstant für alle $\hat{\beta}$, die Lösung von (N) sind, d. h. $X^TX\hat{\beta} = X^Ty$.
- Es existiert ein $a \in \mathbb{R}^n$ mit $\mathbb{E}(a^TY) = c^T\beta$.

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 $\hat{l} = a_*^TY$ (für genau ein $a_* \in \mathrm{im}(X) \subseteq \mathbb{R}^n$) 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 $\hat{l} = a^TY$.

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

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

Bem. Es gilt

$$S(Y,\hat\beta) = \min_{\beta\in\mathbb{R}^p} \|Y - X\beta\|^2 = \|Y - X\hat\beta\|^2 = (Y - X\hat\beta)^T(Y - X\hat\beta) =$$

$$= Y^TY - \underbrace{Y^TX\hat\beta}_{=(X\hat\beta)^TX\hat\beta} - (X\hat\beta)^TY + \underbrace{(X\hat\beta)^TX\hat\beta}_{=\hat\beta^TX^TX\hat\beta=\beta^TX^TY} = \|Y\|^2 - \|X\hat\beta\|^2.$$

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 $\mathbb{E}(c^T\hat\beta(Y - X\hat\beta)) = 0$
- $\operatorname{Cov}(Y - X\hat\beta) = \operatorname{Cov}(Y) - \operatorname{Cov}(X\hat\beta)$.

Orthogonale Transformation von $[Y, X\beta, \sigma^2 I_n]$
$X = (\tilde x_1 \cdots \tilde x_p)$ mit $\tilde x_i \in \mathbb{R}^p$, $\operatorname{rk} X = r \le p$.
Es existiert eine orthonormale Basis $o_1, \dots, o_r$ von $\operatorname{im}(X)$ und $\sigma_{r+1}, \dots, \sigma_n \in \operatorname{im}(X)^\perp$.

$Y = \sum_{i=1}^n o_i Z_i$, $Z = (Z_1 \cdots Z_n)^T$, d. h. $Y = OZ$, wobei $O(o_1 \cdots o_n)$,
$O_1 = (a_1 \cdots a_r)^T$, $O_2 = (a_{r+1} \cdots a_n)^T$
$Z = O^{-1}Z = O^TY$
$\mathbb{E}Z = O^T\mathbb{E}Y = O^TX\beta = \begin{pmatrix} O_1^T \\ O_2^T \end{pmatrix} X\beta$

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

$\mathbb{E}Z = \begin{pmatrix} O_1^TX\beta \\ 0 \end{pmatrix}$, $\operatorname{Cov}(Z) = \operatorname{Cov}(O^TY) = O^T\operatorname{Cov}(Y)O = \sigma^2 I$
Transformation: $[Y, X\beta, \sigma^2 I_n] \rightsquigarrow [Z, O^TX\beta, \sigma^2 I_n]$ mit $Z = O^TY$

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

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

eine erwartungstreue Schätzung für σ^2.

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

- Die ML-Schätzung für $\beta \in \mathbb{R}^p$ stimmt mit der MkQ-Schätzung $\hat\beta$ überein und es gilt $\hat\beta \sim \mathcal{N}_p(\mathbb{E}\hat\beta, \operatorname{Cov}(\hat\beta))$.
- 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 $\mathbb{E}\hat\sigma_n^2 = \frac{n-r}{n}\sigma^2 \xrightarrow[n\to\infty]{} \sigma^2$ (asympt. erw.-treu) und $\frac{S(Y,\hat\beta)}{\sigma^2} \sim \chi^2_{n-r}$
- Für einen Vektor $l^T(\beta) = (l_1(\beta), \dots, l_q(\beta))$ von $q \le r$ 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 $\operatorname{rk} A_* = q$, wobei

  $A_* = (a_{*,1}, \dots, a_{*,q})^T$ mit $a_{*,i} \in L(X)$ 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 $\operatorname{rk} X = p$ gilt $\hat\beta \sim \mathcal{N}_p(\beta, \sigma^2(X^TX)^{-1})$ und $\hat\beta$ und $\hat\sigma^2$ sind unabhängig. (Grund: $\beta_i = e_i^T\beta$ 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, $\operatorname{rk} X = r \le p$
$T := \frac{\|Y - X\hat\beta\|^2}{\sigma_0^2} \sim \chi^2_{n-r}$ (unter H_0) Kritischer Berech 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

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

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

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

	1	2	$\cdots$	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: $X^TX\beta = X^TY$

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

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

$\hat\mu_i = \overline{Y}_{i\bullet} = \frac{1}{n_i}\sum_{k=1}^{n_i} Y_{ik}$ für $i = 1, \dots, p$
Schätzung der Modellstreuung:

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

Dann ist $(n-p)\dfrac{\hat\sigma^2}{\sigma^2} \sim \chi^2_{n-p}$
Test auf Gleichheit der Erwartungswerte:
$H_0: \mu_1 = \mu_2 = \ldots = \mu_p$
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 $T \ge T_{p-1,n-p,1-\alpha}$

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

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

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

Berechnung von S_1^2 mittels Lagrange-Multiplikatoren ergibt:

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

$$S_1^2 - S_0^2 = \ldots = \sum_{i=1}^p n_i(\overline{Y}_{i\bullet} - \overline{Y}_{\bullet\bullet})^2$$

Sprechweise. Übliche Bezeichnungen:

$S_1^2 - S_0^2 = \mathrm{SQA} =$ Summe der Quadrate der Abweichunger

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

Bartlett-Test: Prüfen der Hypothese, dass p unabhänige ZGn $X_i \sim \mathcal{N}(\mu_i, \sigma^2)$, $i = 1, \dots, p$ dieselbe Streuung σ^2 besitzen.
$H_0: \sigma_1^2 = \ldots = \sigma_p^2 = \sigma^2$ Gegeben seien p unabhängige Stichproben $X_{i1}, \dots, X_{in_i}$, $i = 1, \dots, p$, $n = n_1 + \ldots + n_p$.

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

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

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

Für $\min(n_1, \dots, n_p) \to \infty$ gilt $T_{n_1,\dots,n_p} \xrightarrow{d} \chi^2_{p-1}$.
Faustregel: $\min(n_1, \dots, n_p) \ge 5$
Ablehnung von H_0, falls $T_{n_1,\dots,n_p} > \chi^2_{p-1,1-\alpha}$.
Zweifache Varianzanalyse (Zweiwegklassifikation, Kreuzwegkassifikation)
Wirkung eines Faktors A in p Stufen und Wirkung eines Faktors B in q Stufen mit s Wiederholungen in jdeder 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: $H_A : \alpha_1 = \ldots = \alpha_p = 0$, $H_B : \beta_1 = \ldots = \beta_q = 0$, $H_{AB} : \gamma_{11} = \ldots = \gamma_{pq} = 0$

Schreibweise als lineares Modell:

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

$(p+1)\cdot(q+1)$ Parameter, $\beta = (\mu_0, \alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q, \gamma_{11}, \ldots, \gamma_{pq})$

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

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

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

Wegen $pq - 1 + (2 + p + q) = (p+1)\cdot(q+1)$ kann die Eindeutigkeit der MkQ-Schätzung gesichtert werden. Diese Bedingungen bedeuten keine Einschränkung der Allgemeinheit der Darstellung $Y_{ijk} = \mu_0 + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijk}$ denn:

$\mu_0^* = \mu_0 + \overline{\alpha}_\bullet + \overline{\beta}_\bullet + \overline{\gamma}_\bullet$, $\alpha_i^* = \alpha_i - \overline{\alpha}_\bullet + \overline{\gamma}_{i\bullet} - \overline{\gamma}_{\bullet\bullet}$, $\beta_j^* = \beta_j - \overline{\beta}_\bullet + \overline{\gamma}_{\bullet j} - \overline{\gamma}_{\bullet\bullet}$ und $\gamma_{ij}^* = \gamma_{ij} - \overline{\gamma_{i\bullet}} - \overline{\gamma_{\bullet j}} + \overline{\gamma}_{\bullet\bullet}$ für $i = 1, \ldots, p$, $j = 1, \ldots, q$.

Es ergeben sich die Gleichungen

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

Bezeichnungen für Mittelwert:

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

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

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

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

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

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

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

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 $(x_1, y_1), \ldots, (x_n, y_n)$ 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)

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

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

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

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

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

$$\hat{\mu}_1(x) = \tfrac{1}{nh_n} \sum_{i=1}^n Y_i \int_{s_{i-1}}^{s_i} K(\tfrac{x-s}{h_n})\,\mathrm{d}s$$

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

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

Motivation:

$\mathbb{E}\hat{\mu}_n(x) = \frac{1}{nh_n} \sum_{i=1}^n \mathbb{E}Y_i \int_{s_{i-1}}^{s_i} K(\frac{x-s}{h_n})\,\mathrm{d}s = \frac{1}{n} \sum_{i=1}^n \mu(x_i) \int_{s_{i-1}/h_n}^{s_i/h_n} K(\frac{x}{h_n} - t)\,\mathrm{d}t = \frac{1}{h} \sum_{i=1}^n \mu(x_i) \int_{(x-s_{i-1})/h_n}^{(x-s_i)/h_n} K(y)\,\mathrm{d}y \approx \mu(x)$

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

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

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 $g : \mathbb{R} \to \mathbb{R} \cup \{\pm\infty\}$ mit $\mathbb{E}(Y|X) = g(X)$ (Faktorisierungslemma). Wir können dann schreiben: $g(x) = \mathbb{E}(Y|X = x) = \int y f_{Y|X}(y|x)\,\mathrm{d}y$, wobei $f_{Y|X}(y|x)$ = bedingte Dichte von Y bei gegebenen $X = x$ $= \dfrac{f_{(X,Y)}(x,y)}{f_X(x)}$ mit $f_X(x) = \int f_{(X,Y)}(x,y)\,\mathrm{d}y$.

Motivation für (i_1):

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

Schätzung von $f(x,y)$ mittels Produktkern: $K(x,y) = K1(x) \cdot K_2(y)$ mit K_1, K_2 wie oben.

$$\hat{f}_n(x,y) \coloneqq \tfrac{1}{nh_n^2}$$

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

$$\hat{\mu}_n = \int y \tfrac{\hat{f}_n(x,y)}{\hat{f}_n(x)}\,\mathrm{d}y = \tfrac{1}{\hat{f}_n(x)} \tfrac{1}{nh_n^2} \sum_{i=1}^n K_1(\tfrac{x-X_i}{h_n}) \int y K_2(\tfrac{y-Y_i}{h_n})\,\mathrm{d}y \xrightarrow[n\to\infty]{}$$

Eigenschaften des Gasser-Müller-Schätzers:

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

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

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

$$\mathbb{E}\hat{\mu}_n(x) = \mu(x) + \tfrac{h_n^2}{2} \int_0^1 x^2 K(x)\,\mathrm{d}x \mu''(x) + o(h_n^2) + O(\tfrac{1}{n})$$

$$\operatorname{MSE}(\hat{\mu}(\mathrm{x})) = \tfrac{\sigma^2}{nh_n} \int_0^1 K^2(x)\,\mathrm{d}x + \tfrac{h_n^4}{4} \int x^2 K(x)\,\mathrm{d}x (\mu''(x))^2 + o(\tfrac{1}{nh_n}) + o(h_n^4)$$

Dichteschätzungen

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

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

die Menge der stetigen W-Dichtefunktionen.

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

Notation. $\hat{f}_n(t) \coloneqq \hat{f}_n(X_1, \ldots, X_n, t)$

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 $I_j \coloneqq [x_0 + jh, x_0 + (j+1)h)$ für $j \in \mathbb{Z}$. Das **Histogramm** ist der (naive) Dichte-Schätzer

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

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

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

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

$$\mathrm{MISE}(\hat{f}_n) \coloneqq \Delta_n \coloneqq \mathbb{E}_f \int_{\mathbb{R}} (\hat{f}_n(t) - f(t))^2 \, \mathrm{d}t$$

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

Satz (Freedman, Diaconis)**.** Sei $f : \mathbb{R}^1 \to \mathbb{R}^1$ 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

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

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

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 $\min\limits_{h=h_n>0} \Delta_n^2 = \frac{\beta}{n^{2/3}} + o\left(\frac{1}{n^{2/3}}\right)$ 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 $\int K(t) \, \mathrm{d}t = 1$ und $(h_n)_{n \in \mathbb{N}}$ eine Folge in $(0, \infty)$ mit $h_n \downarrow 0$. Dann heißt

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

Kerndichteschätzer für f mit **Kernfunktion** K.