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

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

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** $T : \mathbb{R}^n \to \mathbb{R}^1$ 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

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

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

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

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

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

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 $n \geq 2$.

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 $X_1^{(i)}, ..., X_{n_i}^{(i)} \sim \mathcal{N}(\mu_i, \sigma_i^2)$, $i = 1, 2$ 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:

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

Bspe.

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

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

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

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

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

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

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

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:

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

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

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

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 $(X_1,\dots,X_r)$ 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

  $$\hat{T}_{k_1,\dots,k_r}^{(n)}\coloneqq n^{r-1}\sum_{i_1=1}^{k_1}\cdots\sum_{i_r=1}^{k_r}\frac{\left(h_{i_1\cdots i_r}^{(n)}-n^{-r+1}\prod_{j=1}^{r}h_{\bullet\cdots i_j\cdots\bullet}^{(n)}\right)^2}{\prod_{j=1}^{r}h_{\bullet\cdots i_j\cdots\bullet}}$$

  $$\xrightarrow[n\to\infty]{d}\chi^2_{k_1\cdots k_s-k_1-\dots-k_r+r-1}$$

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

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

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 $U_1,\dots,U_n\sim\mathcal{R}[0,1]$ i. i. d.

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

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

unabhängig von F.

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

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

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

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

Entscheidungsregel (Kolmogorow-(Smirnow-)1SP-Test). Wir testen $H_0:F=F_0$ **gegen** $H_1:F\ne F_0$. 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

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

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.** $H_1:F\ne F_0$ 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.** $H_1:F\ne F^*$ testen, indem wir die empirischen VFen $\hat{F}_n$ und $\hat{F}_m^*$ vergleichen. Dazu verwenden wir

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

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

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

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

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

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

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

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) $g_{m,n}(z) := \sum_{k=0}^{n \cdot m} \mathbb{P}(W_{m,n} = k) \cdot z^k =$

$$= \frac{z^{-m(m+1)/2}}{\binom{m+n}{m}} \sum_{1 \leq \nu_1 < \dots < \nu_m \leq m+n} z^{\nu_1 + \dots + \nu_m} = \frac{1}{\binom{m+n}{m}} \prod_{k=1}^{m} \frac{1 - z^{n+k}}{1 - z^k}$$

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

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

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

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

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

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

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., $h : \mathbb{R}^m \to \mathbb{R}^1$ 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

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

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

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

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

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

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

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

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

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

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 $0 < q \leq 1$ folgt
$$\mathbb{E}|\sqrt{n}(U_n^{(m)} - \tilde{U}_n^{(m)})|^{1+q} \leq \tfrac{c(q,m)}{n^{2q}} \mathbb{E}|h(X_1, \dots, X_m)|^{1+q}.$$
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

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

Bem. Falls $X_1, \dots, X_n$ 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 $X_1,\dots,X_n$:

$$\begin{aligned}(\widehat{\mathrm{Cum}_3(X)})_n &:= \tfrac{1}{n(n-1)(n-2)}(n^2\hat M_3^{(n)} - 3n\hat M_1^{(n)}\hat M_2^{(n)} - 2(\hat M_1^{(n)})^3)\\ &= \tfrac{1}{\binom{n}{3}}\sum_{1\le i<j<j\le n} h(X_i,X_j,X_k)\end{aligned}$$

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 mathematische Stichprobe mit stetiger VF F. Wir wollen **$H_0 : F$ ist symmetrisch bzgl. ϑ_0** testen. Es reicht dazu, die VF der $Z_i = X_i - \vartheta_0$ auf Symmetrie bzgl. 0 zu prüfen.
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

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

Bsp. Alternativ können wir zum Test auf Symmetrie die U-Statistik

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

Unter H_0 gilt für $h(x_1,x_2) := \mathbb{1}_{\{x_1+x_2>0\}}$:

$$\mathbb{E}h(Z_i,Z_j) = \mathbb{P}(Z_1>-Z_2) = \textstyle\int(1-F(-z))\,\mathrm{d}F(z) = \int F(z)\,\mathrm{d}F(z) = \frac{1}{2}.$$

Aus dem ZGWS für U-Statistiken folgt

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

Entscheidungsregel. Ablehnung von $H_0 \iff |U_n - \frac{1}{2}| > \frac{z_{1-\alpha/2}}{\sqrt{3n}}$.

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

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

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

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

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

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

$$\sqrt{\frac{n_1n_2}{n_2\sigma_1^2+n_1\sigma_2^2}}\cdot(U_{n_1,n_2}-\theta) \xrightarrow[n_1,n_2\to\infty]{d} \mathcal{N}(0,1).$$

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.

Bem. Falls Y eine bek. Kovarianzmatrix $K\in\mathbb{R}^{n\times n}$ hat, 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 $\mathrm{Cov}(Y^*) = I_n$. Wir dürfen daher annehmen:

Voraussetzung. $\mathrm{Cov}(Y_i,Y_j) = \mathrm{Cov}(\epsilon_i,\epsilon_j) = \sigma^2\delta_{ij}$. Dabei heißt σ **Modellstreuung**. Üblicherweise gilt $n>p$.

Problem. Gegeben seien $[Y,X\beta,\sigma^2I_n]$.
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: $Y_i = \beta_1+\beta_2x_i+\epsilon_i$, $i = 1,\dots,n$. 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).

$$Y_i = \beta_0+\beta_1X_1^{(i)}+\ldots+\beta_mX_m^{(i)}+\epsilon_i$$

Bsp (Quasilineare (multiple) Regression).

$$Y_i = \beta_0^{(i)}+\beta_1f_1(X_1^{(i)})+\ldots+\beta_mf_m(X_m^{(i)})+\epsilon_i$$

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 spez. Lsg $\hat\beta = S^-X^TY$ (mit $z = 0$) der MkQ-Schätzung gilt

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

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 $\mathrm{Cov}\,\hat\beta = \sigma^2S^{-1}$

Schätzbare Funktionen

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

Satz. Es sind äquivalent:
- $\ell(\beta) = c^T\beta$ ist eine schätzbare Funktion.
- $\hat\ell := \ell(\hat\beta) := c^T\hat\beta$ (wobei $\hat\beta$ MkQ-Schätzung) ist eine lineare Funktion von Y und eine erwartungstreue Schätzung für $\ell(\beta)$
- $c\in\mathrm{im}(X^T) = \mathrm{im}(X^TX)$
- $\ell(\hat\beta) = c^T\hat\beta$ ist konstant für alle $\hat\beta$, die Lösung von (N) sind.
- Es existiert ein $a\in\mathbb{R}^n$ mit $\mathbb{E}(a^TY) = c^T\beta$.

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

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

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

Schätzung der Modellstreuung σ^2

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. Für die MkQ-Schätzung $\hat\beta$ gilt

- $\mathbb{E}(Y - X\hat\beta) = 0$,
- $c^T\beta$ ist eine schätzbare Funktion und $\mathbb{E}(c^T\hat\beta(Y - X\hat\beta)) = 0$
  TODO: Was ist c???,
- $\mathrm{Cov}(Y - X\hat\beta) = \mathbb{E}S(Y,\hat\beta) = \mathbb{E}[\|Y\|^2 - \|X\beta\|^2] = \mathrm{Cov}(Y) - \mathrm{Cov}(X\hat\beta)$.

Verfahren (Orthogonale Transformation eines linearen Modells). Sei $[Y, X\beta, \sigma^2 I_n]$ geg. und $r = \mathrm{rk}\, X \leq p$. Wähle eine orthonormale Basis $o_1, \ldots, o_r$ von $\mathrm{im}(X) \subseteq \mathbb{R}^n$. Ergänze diese zu einer ONB $o_1, \ldots, o_n$ von $\mathbb{R}^n$. Wir setzen

$$O_1 = (o_1 \cdots o_r), \quad O_2 = (o_{r+1} \cdots o_n), \quad O = (O_1\, O_2) = (o_1, \ldots, o_n).$$

Wir betrachten nun das lineare Modell $[Z, O^TX\beta, \sigma I_n]$, wobei

$$Z := O^{-1}Y = O^TY = \begin{pmatrix} O_1^T \\ O_2^T \end{pmatrix}.$$

Es gilt $\quad \mathrm{Cov}(Z) = \mathrm{Cov}(O^TY) = O^T\,\mathrm{Cov}(Y)O = \sigma^2 I$

$$\mathbb{E}Z = O^T\mathbb{E}Y = O^TX\beta = \begin{pmatrix} O_1^TX\beta \\ O_2^TX\beta \end{pmatrix} = \begin{pmatrix} O_1^TX\beta \\ 0 \end{pmatrix}$$

Satz. Sei $[Y, X\beta, \sigma^2 I_n]$ geg., $r := \mathrm{rk}\, X$ und $\hat\beta$ eine MkQ-Schätzung. Dann ist eine erwartungstreue Schätzung für σ^2 gegeben durch

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

Normalverteilte lineare Modelle

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

- Die ML-Schätzung für $\beta \in \mathbb{R}^p$ stimmt mit der MkQ-Schätzung $\hat\beta$ überein und es gilt $\hat\beta \sim \mathcal{N}_p(\mathbb{E}\hat\beta, \mathrm{Cov}(\hat\beta))$.
- Die ML-Schätzung für σ^2 lautet $\hat\sigma_n^2 = \frac{S(Y,\hat\beta)}{n} = \frac{n-r}{n}\hat\sigma^2$. Es gilt $\mathbb{E}\hat\sigma_n^2 = \frac{n-r}{n}\sigma^2 \xrightarrow[n\to\infty]{} \sigma^2$ (asympt. erw.-treu) und $\frac{S(Y,\hat\beta)}{\sigma^2} \sim \chi^2_{n-r}$.
- Für einen Vektor $\ell^T(\beta) = (\ell_1(\beta), \ldots, \ell_q(\beta))$ von $q \leq r$ linear unabhängigen schätzbaren Funktionen $\ell_i(\beta) = c_i^T\beta$, $c_i \in \mathbb{R}^p$ gilt
  $$\hat\ell := \ell(\hat\beta) \sim \mathcal{N}_q(\ell(\beta), \sigma^2 A_* A_*^T) \quad \text{mit } q = \mathrm{rk}\, A_*.$$
  Dabei ist $A_* = (a_{*,1}, \ldots, a_{*,q})^T$ mit $a_{*,i} \in L(X)$ optimal gemäß dem Gauß-Markov-Theorem.
- Die Schätzungen $\hat\ell = \ell(\hat\beta)$ und $\hat\sigma^2$ (bzw. $\hat\sigma_n^2$) sind unabhängig.

Kor. Für $\mathrm{rk}\, X = p$ 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 $[Y, \mathcal{N}_n(X\beta, \sigma^2 I_n)]$). Wir testen $H_0 : \sigma^2 = \sigma_0^2$ **vs.** $H_1 : \sigma^2 \neq \sigma_0^2$. Wir verwenden dazu

$$T := \tfrac{\|Y - X\hat\beta\|^2}{\sigma_0^2}$$

Unter H_0 gilt $T \sim \chi^2_{n-r}$, wobei $r := \mathrm{rk}\, X \leq p$.

Entscheidungsregel. Wir lehnen H_0 genau dann ab, falls

$$T \in K^* = \left[0, \chi^2_{n-r,\alpha/2}\right] \cup \left[\chi^2_{n-r,1-\alpha/2}, \infty\right).$$

Bem. Sei $\ell(\beta) = (\ell_1(\beta), \ldots, \ell_q(\beta))^T$ ein Vektor von linear unabh. schätzbaren Fktn, wobei $1 \leq q \leq r \leq p < n$. Setze $w := \ell(\hat\beta) - \ell(\beta)$. Die Konfidenzschätzung für $\ell(\beta)$ ist dann

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

Anwendung auf das Modell I der Varianzanalyse

Bsp. Ziel ist der Vergleich von Erwartungswerten von p Stufen (*Populationen*), je $\mathcal{N}(\mu_i, \sigma^2)$-verteilter unabhängiger Beobachtungen ($i = 1, \ldots, p$). Für alle Populationen i tragen wir alle n_i Messergebnisse in den *Versuchsplan* (rechts) ein.

	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}

In unserem math. Modell gibt es einen (unbek.) Vektor $\mu \in \mathbb{R}^p$ mit

$$Y_{ik} = \mu_i + \epsilon_{ik}, \quad \epsilon_{ik} \sim \mathcal{N}(0, \sigma^2) \text{ (i. i. d.)} \quad \text{für } n = 1, \ldots, n_i,\ i = 1, \ldots, p.$$

(Wichtig ist, zu prüfen, ob tatsächlich die Varianz der ϵ_{ik} gleich ist, etwa mit dem *Bartlett-Test*.) Die Transponierte X^T der Versuchsplanmatrix $X \in \mathbb{R}^{n\times p}$, $n := n_1 + \ldots + n_p$, ist in Zeilenstufenform, wobei die i-te Zeile aus genau n_i Einsen besteht. Es gilt

$$X^TX = \begin{pmatrix} n_1 & & & 0 \\ & n_2 & & \\ & & \ddots & \\ 0 & & & n_p \end{pmatrix}, \quad X^TY = \begin{pmatrix} Y_{1\bullet} = \sum_{k=1}^{n_1} Y_{1k} \\ \vdots \\ Y_{p\bullet} = \sum_{k=1}^{n_p} Y_{pk} \end{pmatrix}$$

Aus der Normalengleichung $X^TX\beta = X^TY$ folgt

$$\hat\mu_i = \overline{Y}_{i\bullet} = \tfrac{1}{n_i}\textstyle\sum_{k=1}^{n_i} Y_{ik} \quad \text{für } i = 1, \ldots, p.$$

Die Schätzung der Modellstreuung ist

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

Es gilt $(n-p)\frac{\hat\sigma^2}{\sigma^2} \sim \chi^2_{n-p}$.

Test. Wir testen $H_0 : \mu_1 = \mu_2 = \ldots = \mu_p$ **vs.** $H_1 : \exists\, i, j \,:\, \mu_i \neq \mu_j$. Als Testgröße verwenden wir

$$T := \tfrac{n-p}{p-1} \cdot \tfrac{S_1^2 - S_0^2}{S_0^2} \quad \text{mit } S_0^2 := \|Y - X\hat\beta\|^2 = \sum_{i=1}^p\sum_{k=1}^{n_i}(Y_{ik} - \overline{Y}_{i\bullet})^2,$$

$$S_1^2 := \sum_{i=1}^p\sum_{k=1}^{n_i}(Y_{ik} - \overline{Y}_{\bullet\bullet})^2 \quad \rightsquigarrow S_1^2 - S_0^2 = \ldots = \sum_{i=1}^p n_i(\overline{Y}_{i\bullet} - \overline{Y}_{\bullet\bullet})^2$$

Unter H_0 gilt $T \sim F_{p-1,n-p}$.

Entscheidungsregel. Ablehnung von $H_0 \iff T \geq F_{p-1,n-p,1-\alpha}$

Sprechweise. Die üblichen Bezeichnungen sind

$$\begin{aligned} S_1^2 &= \text{SQG} = \text{S. d. Q. d. A. in der Gesamtheit} \\ S_1^2 - S_0^2 &= \text{SQA} = \text{S. d. Q. d. A. zwischen den Stufen des Faktors } A \\ S_0^2 &= \text{SQR} = \text{S. d. Q. d. A. innerhalb der Stufen des Faktors } A \\ &= \text{Restquadratesumme,} \end{aligned}$$

wobei „S. d. Q. d. A.“ = „Summe der Quadrate der Abweichungen“.

Test (**Bartlett**). Seien $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$, $i = 1, \ldots, p$, unabh. ZGn. Wir prüfen $H_0 : \sigma_1^2 = \ldots = \sigma_p^2 = \sigma^2$ **vs.** $H_1 : \exists\, i, j \,:\, \sigma_i^2 \neq \sigma_j^2$. Dazu verwenden wir die Testgröße

$$\begin{aligned} T_{n_1,\ldots,n_p} &:= \tfrac{1}{D}\left((n-p)\log S^2 - \textstyle\sum_{i=1}^p (n_i - 1)\log S_i^2\right), \\ D &:= 1 + \tfrac{1}{3(-1)}\left(\textstyle\sum_{i=1}^p \tfrac{1}{n_i-1} - \tfrac{1}{n-p}\right) \\ S_i^2 &:= \tfrac{1}{n_i-1}\textstyle\sum_{k=1}^{n_i}(X_{ik} - \overline{X}_{i\bullet})^2 \\ S^2 &:= \tfrac{1}{n-p}\textstyle\sum_{i=1}^p (n_i - 1)S_i^2 \end{aligned}$$

Unter H_0 gilt $T_{n_1,\ldots,n_p} \xrightarrow[\min(n_1,\ldots,n_p)\to\infty]{d} \chi^2_{p-1}$.

Faustregel. Die Näherung ist gut, falls $\min(n_1, \ldots, n_p) \geq 5$.

Entscheidungsregel. Ablehnung v. $H_0 \Leftrightarrow T_{n_1,\ldots,n_p} > \chi^2_{p-1,1-\alpha}$.

Zweifache Varianzanalyse (*Zweiwegklassifikation*)

Situation. Wir wollen die Wirkung eines Faktors A in p Stufen und die Wirkung eines Faktors B in q Stufen mit s Wiederholungen in jeder Stufe von Faktor A und B untersuchen.

Bsp. Wir untersuchen den Ernteertrag abhängig vom Düngemittel (Faktor A) und der Bodenart (Faktor B). Insbesondere sind wir an den Wechselwirkungen der Faktoren interessiert.

Modell. $Y_{ijk} = \mu_0 + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijk}$, wobei

μ_0	Grundniveau
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 der Stufen i und j
ϵ_{ijk}	ZG mit $\mathbb{E}\epsilon_{ijk} = 0$ und $\mathbb{E}\epsilon_{ijk}^2 = \sigma^2$ (i. i. d.)

Dies lässt sich als lineares Modell $Y = X\beta + \epsilon$ mit

$$\beta = (\mu_0, \alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q, \gamma_{11}, \ldots, \gamma_{pq}) \in \mathbb{R}^{1+p+q+pq},$$
$$X \in \mathbb{R}^{pqs\times(1+p+q+pq)}, \quad \mathrm{rk}\, X = pq - 1 < \min\{pqs, 1 + p + q + pq\}$$

schreiben. Wegen des zu kleinen Ranges ist eine Reparametrisierung notwendig, d. h. es werden Gleichungen zwischen den Parametern hinzugefügt, die die eindeutige Lösbarkeit von (N) garantieren:

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

Notation. $\hat\mu_0 = \overline{Y}_{\bullet\bullet\bullet}$, $\hat\alpha_i = \overline{Y}_{i\bullet\bullet} - \overline{Y}_{\bullet\bullet\bullet}$, $\hat\beta_j = \overline{Y}_{\bullet j\bullet} - \overline{Y}_{\bullet\bullet\bullet}$, $\hat\gamma_{ij} = \overline{Y}_{ij\bullet} - \overline{Y}_{i\bullet\bullet} - \overline{Y}_{\bullet j\bullet} + \overline{Y}_{\bullet\bullet\bullet}$

Test. Wir testen die Hypothesen

$$H_{AB}: \gamma_{11} = \ldots = \gamma_{pq}, \quad H_A: \alpha_1 = \ldots = \alpha_p, \quad H_B: \beta_1 = \ldots = \beta_q$$

mit den Testgrößen

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

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

$$F_B := \frac{pq\cdot(s-1)}{q-1} \cdot \frac{ps}{S^2_{pqs}} \cdot \sum_{j=1}^{q} (\overline{Y}_{\bullet j\bullet} - \overline{Y}_{\bullet\bullet\bullet})^2 \quad \text{wobei}$$

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

Unter H_A gilt $F_A \sim F_{p-1,pq(s-1)}$, unter H_B gilt $F_B \sim F_{q-1,pq(s-1)}$ und unter H_{AB} gilt $F_{AB} \sim F_{(p+1)(q+1),pq(s-1)}$.

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

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

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

Regressionskurvenschätzer

Problem. Zu n Messwerten $(x_1, y_1), \ldots, (x_n, y_n)$ soll eine Fktn $\mu(x)$ gefunden werden, deren Funktionswerte $\mu(x_i)$ die y_i möglichst gut approximieren.

Modell (nichtparametrisches Regressionsmodell, **Modell I**).

$$Y_i = \mu(x_i) + \epsilon_i \quad \text{für } i = 1, \ldots, n$$

mit unbekannter Regressionsfunktion $\mu : [a, b] \to \mathbb{R}^1$.
Dabei gilt $\mathbb{E}\epsilon_i = 0$ und $\mathbb{E}\epsilon_i\epsilon_j = \sigma^2\delta_{ij}$.

Bem. Im **Modell II** werden die x_i's durch ZGn X_i ersetzt.

Voraussetzung. $K \in L^1(\mathbb{R})$ ist eine **Kernfktn**, d. h. $\int K(x)\,\mathrm{d}x = 1$, $\forall x : K(-x) = K(x)$, $\operatorname{supp} K$ ist beschränkt und $\sup|K(x)| < \infty$.

Notation. $K_h(x) := \frac{1}{h} K(\frac{x}{h})$

Def. Seien *Messstellen* $a \leq x_1 < x_2 < \ldots < x_n \leq b$ gewählt. Der **Gasser-Müller-Schätzer** mit *Bandweite* $h_n > 0$ ist

$$\hat{\mu}_n^{(\mathrm{GM})}(x) := \frac{1}{nh_n} \sum_{i=1}^{n} Y_i \int_{s_{i-1}}^{s_i} K(\tfrac{x-s}{h_n})\,\mathrm{d}s, \quad a \leq x \leq b$$

wobei $s_0 := a$, $s_{n+1} := b$ und $s_i := {}^{(x_{i-1}+x_i)}/_2$.

Bem. Es gilt $\hat{\mu}_n(x) = (K_{h_n} * h)(x)$ mit $h(x) = Y_i$ für $x \in [s_i, s_{i+1})$.

Für Modell II gibt es folgenden Kernschätzer:

Def. Seien $(X_1, Y_1), \ldots, (X_n, Y_n)$ i. i. d. mit Dichte $f_{(X,Y)}$, K eine Kernfunktion und $hn > 0$. Der **Nadaraya-Watson-Schätzer** ist

$$\hat{\mu}_n^{(\mathrm{NW})}(x) = \frac{\sum_{i=1}^n Y_i K(\frac{x-X_i}{h_n})}{\sum_{i=1}^n K(\frac{x-X_i}{h_n})}, \qquad a \leq x \leq b.$$

Motivation. Unter geeigneten Voraussetzungen gilt

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

Wir schätzen $f_{(X,Y)}(x, y)$ mit einem *Produktkern*, also $K(x, y) = K_1(x) \cdot K_2(y)$ mit Kernfunktionen K_1, K_2:

$$\hat{f}_n(x, y) := \frac{1}{nh_n^2} \sum_{i=1}^{n} K_1(\tfrac{x-X_i}{h_n}) K_2(\tfrac{y-Y_i}{h_n}),$$

$$\hat{f}_n(x) = \frac{1}{nh_n} \sum_{i=1}^{n} K_1(\tfrac{x-X_i}{h_n}) = \int \hat{f}_n(x, y)\,\mathrm{d}y$$

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

$$\xrightarrow[n\to\infty]{} \mu(x) \qquad (\text{bei } h_n \to 0,\ h_n \cdot n \to \infty \text{ für } n \to \infty)$$

Def. $\mathrm{MSE}(\hat{\theta}_n) = \mathbb{E}(\hat{\theta}_n - \theta)^2$ heißt *mean squared error*,
$\mathrm{MASE}(\hat{\mu}_n) = \frac{1}{n} \sum_{i=1}^{n} \mathrm{MSE}(\hat{\mu}_n(x_i))$ heißt *mean averaged squared error*

Bem. Es gilt $\mathrm{MSE}(\hat{\theta}_n) = \mathrm{Var}(\hat{\theta}_n) + (\underbrace{\mathbb{E}(\hat{\theta} - \theta)}_{\textbf{Bias}})^2$

Satz. Seien die Annahmen des Modell I erfüllt und $\mu \in \mathcal{C}^2([0, 1])$. Die Messstellen seien $x_i = {}^{(i-1/2)}/_n$ für $i = 1, \ldots, n$. Der Kern K sei Lipschitz-stetig und es gelte $\operatorname{supp} K \subseteq [-1, 1]$. Angenommen, $h_n \downarrow 0$, $nh_n \to \infty$ für $n \to \infty$. Dann gilt für den Gasser-Müller-Schätzer:

$$\mathbb{E}\hat{\mu}_n(x) - \mu(x) = \tfrac{h_n^2}{2} \int x^2 K(x)\,\mathrm{d}x \mu''(x) + o(h_n^2) + O(\tfrac{1}{n})$$
$$\operatorname{Var} \hat{\mu}_n(x) = \tfrac{1}{nh_n} \sigma^2 \int K^2(x)\,\mathrm{d}x + O(\tfrac{1}{n^2h_n^2})$$

Kor. $\mathrm{MSE}(\hat{\mu}(x)) = \frac{\sigma^2}{nh_n} \int K^2(x)\,\mathrm{d}x + \frac{h_n^4}{4} \left(\int x^2 K(x)\,\mathrm{d}x \cdot \mu''(x)\right)^2 + o(\frac{1}{nh_n} + h_n^4)$

Dichteschätzungen

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

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

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

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

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

Histogramm-Schätzer

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

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

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

$$\underbrace{\mathrm{MISE}(\hat{f}_n)}_{\Delta_n :=} := \mathbb{E}_f \int_{\mathbb{R}} (\hat{f}_n(t) - f(t))^2\,\mathrm{d}t = \int_{\mathbb{R}} \underbrace{\mathbb{E}_f (\hat{f}_n(t) - f(t))^2}_{=:\mathrm{MSE}(\hat{f}_n(t))}\,\mathrm{d}t$$

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

Satz (Freedman, Diaconis)**.** Sei $f : \mathbb{R} \to \mathbb{R}$ eine Dichtefkt mit

(i) $f \in L^2(\mathbb{R})$ und f absolut stetig, d. h. f. ü. diff'bar,
(ii) $f' \in L^2(\mathbb{R})$ und f' absolut stetig, d. h. f. ü. diff'bar und
(iii) $f'' \in L^p(\mathbb{R})$ für ein $p \in [1, 2]$.

Wir schreiben

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

Gelte $\gamma > 0$. Dann ist $\min_{h=h_n>0} \Delta_n^2 = \frac{\beta}{n^{2/3}} + O(\frac{1}{n})$.
Das Minimum wird angenommen für $h_n = \frac{\alpha}{\sqrt[3]{n}} + O(\frac{1}{\sqrt{n}})$.

Angenommen, es gilt nur (i), $f' \in L^2(\mathbb{R})$ und $\gamma > 0$. Dann gilt

$$\min_{h=h_n>0} \Delta_n^2 = \frac{\beta}{n^{2/3}} + o\left(\frac{1}{n^{2/3}}\right) \quad \text{mit Minimum bei} \quad h_n = \frac{\alpha}{\sqrt[3]{n}} + o\left(\frac{1}{\sqrt[3]{n}}\right).$$

Kerndichteschätzer

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

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

Kerndichteschätzer für f mit *Kernfunktion* K.

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

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

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

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

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

Satz. Sei f eine beschränkte W-Dichtefktn, $\forall x\in\mathbb{R} : f(x) \le M$, 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 $n\to\infty$. Angenommen, $n\cdot h_n \xrightarrow[n\to\infty]{} \infty$. Dann gilt:

- $\mathbb{E}_f \hat{f}_n(t) \xrightarrow[n\to\infty]{} f(t) \quad \forall t \in C$,
- $n\cdot h_n \cdot \operatorname{Var}_f(\hat{f}_n(t)) \xrightarrow[n\to\infty]{} f(t)\cdot\int K^2(x)\,dx \quad \forall t\in C$,
- $\sup_{t\in\mathbb{R}} \mathbb{E}_f\left(\hat{f}_n(t) - f(t)\right)^2 \xrightarrow[n\to\infty]{} 0$, falls f glm. stetig auf $\mathbb{R}$.

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

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

wobei die Konvergenz von $o(-)$ sogar glm. in abg. 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 $f, |f'| \le A < \infty$. Für K gelte $0 \le K(x) \le B$, $K(-x) = K(x)$ und $\int x^2 K(x)\,dx < \infty$. Dann gilt für $(h_n)_{n\in\mathbb{N}}$ mit $h_n\downarrow 0$ und $nh_n \xrightarrow[n\to\infty]{} \infty$:

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

Bem. Wir wollen zunächst h_n irgendwie optimal wählen. Setzen wir die beiden echten Fehlerterme gleich, also

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

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

$$\begin{aligned}\operatorname{MISE}(\hat{f}_n) =& \frac{\sqrt[5]{8}}{n^{4/5}}\left(\int (f''(t))^2\,dt\right)^{1/5}\left(\int x^2 K(x)\,dx\right)^{2/5}\left(\int K(x)^2\,dx\right)^{4/5} \\ &+ o(n^{-4/5}).\end{aligned}$$

Bem. Nun wollen wir die Kernfunktion K optimal wählen, also so, dass der IMSE in der letzten Gleichung möglichst klein wird. Dazu suchen wir eine Funktion K, sodass

$$\int x^2 K(x)\,dx \left(\int K(x)^2\,dx\right)^2$$

minimal wird unter den Nebenbed. $\int K(x)\,dx = 1$ und $K(x)\ge 0$.

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

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

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

Def. Im Spezialfall $\alpha = 2$ heißt

$$K_0(x) := \tfrac{3}{4}(1-x^2)\mathbb{1}_{[-1,1]}(x) \quad \textbf{Epanetschnikow-Kern}.$$

Bem. Wichtige Kerne neben dem Epanetschnikow-Kern:

- **Dirichlet-Kern**: $K_N(x) = \frac{1}{2}\sum_{k=-N}^N e^{i2\pi kx} = \frac{\sin((2N+1)\pi x)}{2\sin(\pi x)}$
- **Fejér-Kern**: $K_N(x) = \frac{1}{2N}|\sum_{k=0}^N e^{ik\pi x}|^2 = \frac{\sin^2((2N+1)\pi x)}{2N\sin^2(\pi x)}$
- **Dreieckskern**: $K(x) = (1-|x|)\mathbb{1}_{[-1,1]}(x)$
- **Kosinuskern**: $K(x) = \begin{cases}\frac{\pi}{4}\cos(\frac{\pi}{2}x) & |x|\le 1\\ 0 & |x|>1\end{cases}$
- **Sobolev-Kern**: $K_\epsilon(x) = c_d\epsilon^{-d}\exp\left(-\frac{\epsilon^2}{\epsilon^2 - \|x\|^2}\right)\mathbb{1}_{B_\epsilon(0)}(x)$
  Das Besondere an ihm ist $K_\epsilon \in \mathcal{C}^\infty(\mathbb{R}^d)$.

Bemn.
- Es gibt Kerne mit der Eigenschaft $\int x^k K(x)\,dx = 0$ für $k = 1,\dots,m-1$ und $\int |x|^m K(x)\,dx < \infty$ für $m\ge 2$.
- Es gibt Kerne mit $\int x^n K(x)\,dx = 0$ für alle $n\in\mathbb{N}$, z. B.

$$K(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\cos(tx)g(t)\,dt \quad \text{mit} \quad g(t) := \begin{cases}1 - e^{-1/t^2} & \text{falls } t\ne 0,\\ 1 & \text{falls } t = 0\end{cases}$$

Satz (H. Müntz)**.** Sei $K:\mathbb{R}\to\mathbb{R}$ eine Fkt mit $\int x^{n_i}K(x)\,dx = 0$ für $0 < n_1 < n_2 < \dots$ mit $\sum_{i=1}^\infty \frac{1}{n_i} = \infty$, $\operatorname{supp}(K) = [a,b]$, $K\in\mathcal{C}[a,b]$. Dann gilt $K\equiv 0$.

Orthogonale Schätzer für W-Dichten

Voraussetzung. Sei $M = [a,b]$ mit $-\infty < a < b < \infty$.

Bem. Der Hilbertraum $L^2(M)$ besitzt eine abzählbare ONB $e_1, e_2, \dots$, sodass für alle $f\in L^2(M)$ gilt:

$$f(x) = \sum_{k=1}^\infty \alpha_k e_k(x) \quad \text{mit } \alpha_k = \langle f, e_k\rangle \quad \text{für } x\in[a,b]$$

Problem. Gegeben sei eine math. SP $X_1,\dots,X_n$ aus einer Grundgesamtheit für ein zuf. Merkmal mit Dichte $f\in L^2(M)$. Gesucht ist eine Schätzung der Dichte f.

Verfahren. Schätzung von f erfolgt in zwei Schritten:

1. Wähle einen Parameter $N = N_n$. Sei f_N die Projektion von f auf $L(e_1,\dots,e_N) \subseteq L^2(M)$,
$$f_N(x) = \sum_{k=1}^N \alpha_k e_k(x),\ \ x\in M \quad \text{mit } \alpha_k = \langle f, e_k\rangle$$
2. Schätzung der Koeffizienten:
$$\hat{\alpha}_k = \hat{\alpha}_{k,n} = \frac{1}{n}\sum_{k=1}^n e_k(X_i), \ \ k = 1,\dots,N_n$$

Def. Der **ON-Schätzer** (oder *Projektionsschätzer*) ist

$$\hat{f}_n(x) = \hat{f}_{n,N}(x) := \sum_{k=1}^N \hat{\alpha}_{k,n}e_k(x) = \frac{1}{n}\sum_{i=1}^n\sum_{k=1}^N e_k(X_i)e_k(x).$$

Eigenschaften des Schätzers $\hat{f}_n$:

Lem. Für $f\in L^2(M)$ und $e_1(x) \equiv \text{const} = \frac{1}{\sqrt{|M|}}$, $|M| = b - a$ gilt

$$\int_M \hat{f}_n(x)\,dx = 1 \quad \text{und} \quad \mathbb{E}\hat{f}_n(x) = f_N(x)\,\forall x\in M.$$

Kor (Pktweise asympt. Erwartungstreue)**.** Für $N_n \xrightarrow[n\to\infty]{} \infty$ gilt $\mathbb{E}\hat{f}_n(x) \xrightarrow[n\to\infty]{\text{f. ü.}} f(x)$, falls $f_N(x) \xrightarrow[N\to\infty]{\text{f. ü.}} f(x)$.

Lem. Sei $f\in L^2(M)$. Dann gilt für ON-Schätzer

$$\|\hat{f}_n - f\|_2 = \sum_{k=1}^N(\hat{\alpha}_{k,n} - \alpha_k)^2 + \sum_{k=N+1}^\infty \alpha_k^2$$

Satz. Es sind äquivalent:

- $J_n = \mathbb{E}\|\hat{f}_{n,N_n} - f\|_2^2 = \mathbb{E}\int_M(\hat{f}_{n,N_n}(x) - f(x))^2\,dx \xrightarrow[n\to\infty]{} 0$
- $\frac{1}{n}\sum_{i=1}^{N_n}\int_M e_k^2(x)f(x)\,dx \xrightarrow[n\to\infty]{} 0$

Kor. Aus $J_n \xrightarrow[n\to\infty]{} 0$ folgt $\|\hat{f}_{n,N_n} - f\|_2 \xrightarrow[n\to\infty]{\mathbb{P}} 0$.

Simulation von Zufallszahlen und Simulationstests

1. Möglichkeit: Quantilfunktionen, Inversionsmethode

Sei F eine VF auf $\mathbb{R}^1$. Dann heißt die Funktion $F^- : [0,1] \to [-\infty, \infty)$ mit $F^-(y) := \min\{x \in \mathbb{R}^1 \mid F(x) \geq y\}$ für $y \in (0,1]$, $F^-(0) = \lim_{y \downarrow 0} F^-(y)$ **Quantilfunktion** oder (verallgemeinerte) **Pseudo-Inverse** zu F.

Eigenschaften:

- F^- ist monoton und linksseitig stetig
- $F(F^-(y)) \geq y$ für alle $y \in [0,1]$
- $U \sim \mathcal{R}\,[0,1] \implies F^-(U) \sim F$. Falls $(U_i)_{i\in\mathbb{N}}$ eine Folge unabhängiger gleichverteilter Zufallszahlen auf $[0,1]$ sind, dann ist $(F^-(U_i))_{i\in\mathbb{N}}$ eine Folge F-verteilter Zufallszahlen.

1. Verwerfungsmethode (rejection method)
Voraussetzung: X besitze eine Dichte f mit einem beschränkten Träger $\subseteq [a,b]$ und $f(x) \leq M$ für alle $x \in [a,b]$. Wir erzeugen Zufallszahlen $V_1 = a + (b-a)U_1 \sim \mathcal{R}\,[a,b]$ und $V_2 = MU_2$. Wir nehmen $X := V_1$ falls $(V_1, V_2) \in \text{Graph}(f) \iff V_2 \leq f(V_1)$, andernfalls wiederholen wir das Verfahren mit neuen Werten V_1 und V_2.
2. Box-Muller-Verfahren
Seien $U_1, U_2 \sim \mathcal{R}\,[0,1]$ unabhängig. Dann sind
$X_1 := g_1(U_1, U_2) := \sqrt{-2\log U_1}\sin(2\pi u_2) \sim \mathcal{N}(0,1)$,
$X_2 := g_2(U_1, U_2) := \sqrt{-2\log U_1}\cos(2\pi U_2) \sim \mathcal{N}(0,1)$ unabhängig.

Allgemein sei $h = \binom{h_1}{h_2}$ die Umkehrfunktion von $g = \binom{g_1}{g_2}$, also $\binom{u_1}{u_2} = \binom{h_1(X_1,X_2)}{h_2(X_1,X_2)}$. Dann gilt
$f_{(X_1,X_2)}(x_1,x_2) = \dfrac{f_{U_1,U_2}(h_1(x_1,x_2), h_2(x_1,x_2))}{\det(J_g(h_1(x_1,x_2), h_2(x_1,x_2)))}$, wobei $J_g(x,y)$ die Jacobi-Matrix von g im Punkt (x,y) ist.

Erzeugung eines n-dim. ZV $\mathcal{N}(n)(m,S)$, $m \in \mathbb{R}^n$, $S \in \mathbb{R}^{n\times n}$ Kovarianzmatrix, positiv semidefinit und symmetrisch. Mit Cholesky-Zerlegung bekommt man eine untere Dreiecksmatrix $L \in \mathbb{R}^{n\times n}$ mit $S = L \cdot L^T$.

$Y = (Y_1, \ldots, Y_n)^T := m + X \cdot XL^T$, $X = (X_1, \ldots, X_n)$,
$X_1, \ldots, X_n \sim \mathcal{N}(0,1)$ i. i. d.. Dann ist
$\text{Cov}(Y) = \text{Cov}(XL^T) = L\,\text{Cov}(X)L^T = LL^T = S$.