

# 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  $\vartheta$  abhängt.

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

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

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

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

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

**Def.** Der **Ablehnungsbereich** oder **kritische Bereich** von  $\varphi$  ist

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

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

**Def.** Ein **Fehler 1. Art** ist eine Ablehnung der Nullhypothese  $H_0$ , obwohl  $H_0$  richtig ist; ein **Fehler 2. Art** ist eine Annahme von  $H_0$ , obwohl  $H_0$  falsch ist.

**Def.** Die **Güte- oder Machtfunktion** des Tests  $\varphi$  ist

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

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

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

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

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

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

Ein  $\alpha$ -Test  $\varphi$  heißt **unverfälscht** (erwartungstreu, unbiased), falls

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

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

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

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

*Bem.* Es gilt

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

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

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

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

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

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

Für  $\alpha = 0,5$  gilt beispielsweise  $z_{1-\alpha/2} \approx 1,96$ .

*Bem.* Es gilt

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

Letzteres Intervall wird **Konfidenzintervall** für  $\mu_0$  zum Konfidenzniveau  $1 - \alpha$  genannt.

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

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

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

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

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

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

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

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

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

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

**$t$ -verteilt** mit  $n$ -Freiheitsgraden.

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

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

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

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

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

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

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

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

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

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

**Bsp.** Seien Stichproben  $X_1^{(1)}, \dots, X_{n_1}^{(1)} \sim \mathcal{N}(\mu_1, \sigma_1^2)$  und  $X_1^{(2)}, \dots, X_{n_2}^{(2)} \sim \mathcal{N}(\mu_2, \sigma_2^2)$  gegeben. Wir wollen  $H_0 : \sigma_1 = \sigma_2$  gegen  $H_1 : \sigma_1 \neq \sigma_2$  testen. Dazu verwenden wir

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

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

**Bsp.** Situation wie im letzten Beispiel mit  $\sigma_1 = \sigma_2$ . Wir testen  $H_0 : \mu_1 = \mu_2$  vs.  $H_1 : \mu_1 \neq \mu_2$  mit

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

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

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

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

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

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

Um  $H_0 : \rho = \rho_0 \in (0, 1)$  vs.  $H_1 : \rho \neq \rho_0$  zu testen, kann man

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

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

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

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

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

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

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

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

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

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

Man zeigt, dass dafür gelten muss:

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

**Bspe.** • Sei  $X \sim \text{Exp}(\mu)$ ,  $\hat{\mu}_n := \frac{1}{\bar{X}_n}$ . Dann gilt

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

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

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

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

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

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

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

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

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

*Bem.* Bei  $k = 2$  schreibt man oft

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

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

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

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

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

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

## Chi-Quadrat-Anpassungstest

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

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

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

Wir setzen

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

Die Klassenhäufigkeiten sind nun multinomialverteilt:

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

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

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

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

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

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

*Bem.* •  $T_{n,s+1}$  misst eigentlich nicht die Abweichung von der VF  $F_0$ , sondern von der Multinomialverteilung  $\mathcal{M}(n, p^{(0)})$ .

- Der  $\chi^2$ -Anpassungstest gilt als hypothesenfreundlich.

- Es ist üblich, zunächst die Parameter  $\vartheta = (\vartheta_1, \dots, \vartheta_r)$  der VF  $F_0$  durch MLE zu schätzen, also durch

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

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

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

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

wobei  $r$  die Anzahl der geschätzten Parameter ist.

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

**Ziel.** Überprüfen, ob die Komponenten  $X$  und  $Y$  eines zweidim. Zufallsvektors  $(X, Y)^T$  unabhängig sind.

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

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

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

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

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

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

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

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

Wir können den Test nun wie folgt als Spezialfall des  $\chi^2$ -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

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

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

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

Testregel: Die Nullhypothese wird genau dann abgelehnt, falls

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

*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

$$\begin{aligned} \hat{T}_{k_1, \dots, k_r}^{(n)} &:= n^{r-1} \sum_{i_1=1}^{k_1} \dots \sum_{i_r=1}^{k_r} \frac{\left(h_{i_1 \dots i_r}^{(n)} - n^{-r+1} \prod_{s=1}^r h_{\bullet \dots \bullet i_s}^{(n)}\right)^2}{\prod_{s=1}^r h_{\bullet \dots \bullet i_s}^{(n)}} \\ &\xrightarrow{n \rightarrow \infty} \chi_{k_1 \dots k_s - k_1 - \dots - k_r + r - 1}^2 \end{aligned}$$

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

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

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

## Ein- und Zwei-SP-Tests für VFen

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

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

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

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

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

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

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

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

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

unabhängig von  $F$ .

**Satz.** Falls  $F$  stetig ist, gilt

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

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

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

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

**Entscheidungsregel (Kolmogorow(-Smirnow)-1SP-Test).** Wir testen  $H_0 : F = F_0$  gegen  $H_1 : F \neq F_1$ . Dabei muss  $F_0$  eine stetige VF sein. Wir verwenden dazu

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

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

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

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

**Entscheidungsregel.** Um  $H_0 : F = F_0$  gegen  $H_1 : F > F_0$  mit

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

Es gilt

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

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

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

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

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

**Satz.** Sei  $F$  stetig. Dann ist

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

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

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

**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 testen, ob  $H_0 : F = F^*$  oder  $H_1 : F \neq F^*$  gilt, indem wir die empirischen VFen  $\hat{F}_n$  und  $\hat{F}_m^*$  vergleichen. Dazu verwenden wir

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

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

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

wobei  $X_i \stackrel{d}{=} F^-(U_i)$ ,  $i = 1, \dots, n$ ,  $X_j^* \stackrel{d}{=} F^{*, -}(U_j^*)$ ,  $j = 1, \dots, m$  und

$$F^-(t) := \begin{cases} \min\{x \in \mathbb{R}^1 \mid F(x) \geq t\} & 0 < t \leq 1, \\ \lim_{t \downarrow 0} F^-(t) & t = 0. \end{cases} \quad (\text{Quantilfunktion})$$

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

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

**Situation (2-SP-Test von Wilcoxon-Mann-Whitney, U-Test).** Gegeben seien zwei unabhängige 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.  $H_1 : F \neq F^*$ . Konstruktion einer Rangstatistik für konkrete SPn  $x_1, \dots, x_n$  und  $x_1^*, \dots, x_m^*$ .

- Ordnen:  $x_{1:n} < \dots < x_{n:n}$  und  $x_{1:m}^* < \dots < x_{m:m}^*$
- $\nu_1, \dots, \nu_m \in \{1, \dots, m+n\}$  seien die Ränge der Werte  $x_{i:m}^*$  innerhalb der Gesamt-SP:

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

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. Testgröße:

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

Unter  $H_0$  gilt  $\mathbb{E}W_{m,n} = \frac{m \cdot n}{2}$ ,  $\text{Var } W_{m,n} = \frac{m \cdot n}{12}(m+n+1)$ .

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

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

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

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

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

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

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

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

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

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

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

## Theorie der U-Statistiken

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

$$h(x_1, \dots, x_n) = h(x_{\sigma(1)}, \dots, x_{\sigma(n)}) \quad \forall \sigma \in S_n.$$

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

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

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

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

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

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

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

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

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

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

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

**Lem.** (2)  $\text{Var}(\tilde{U}_n^{(m)}) = \frac{m^2}{n} \cdot \text{Var}(g(X_1)) = \frac{m^2}{n} (\mathbb{E}g^2(X_1) - \theta^2)$   
(3) Falls  $\mathbb{E}|h(X_1, \dots, X_m)| < \infty$ , so gilt

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

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

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

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

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

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

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

- $\mathbb{E}g^2(X_1) < \infty$  ist schwächer als  $\mathbb{E}h^2(X_1, \dots, X_m) < \infty$ .
- Aus  $\mathbb{E}|h(X_1, \dots, X_m)|^{1+q} < \infty$  für  $0 < q \leq 1$  folgt

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

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

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

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

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

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

mit  $\sigma_g^2 = \mathbb{E}g^2(X_2) - (\mathbb{E}g(X_2))^2 = \frac{1}{4}\mu_4 - \frac{1}{4}\sigma^4$ ,  $\mu_4 := \mathbb{E}(X_1 - \mathbb{E}X_2)^4$ .

Spezialfall: Ist  $X_i \sim \mathcal{N}(\mu, \sigma^2)$ , so gilt  $\mu_4 = 3\sigma^4$ .

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

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

Alternativ erhält man durch Anwenden einer varianzstab. Trafo:

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

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

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

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

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

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

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

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

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

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

**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.  $\vartheta_0$  gilt

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



**Bsp.** Alternativ können wir die U-Statistik

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

zum Test auf Symmetrie betrachten. Unter  $H_0$  gilt

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

Aus dem ZGWS für U-Statistiken folgt

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

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

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

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

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

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

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

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

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

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

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

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

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

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

**Problem.** Gegeben seien  $[Y, X\beta, \sigma^2 I_n]$ . Gesucht sind Schätzungen  $\hat{\beta}(y) = (\hat{\beta}_1(y), \dots, \hat{\beta}_p(y))^T$  für  $\beta$ .

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

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

Für die Minima gelten die Normalengleichungen

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

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

**Bsp (Einfache lineare Regression).**

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

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

**Bsp (Multiple lineare Regression).**

$$Y_i = \beta_0 + \beta_1 X_{i1}^{(i)} + \dots + \beta_m X_{im}^{(i)} + \epsilon_i$$

**Bsp (Quasilineare (multiple) Regression).**

$$Y_i = \beta_0^{(i)} + \beta_1 f_1(X_{i1}^{(i)}) + \dots + \beta_m f_m(X_{im}^{(i)}) + \epsilon_i$$

mit (nichtlinearen) Funktionen  $f_1, \dots, f_m$

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

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

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

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

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

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

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

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

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

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

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

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

**Satz.** Es sind äquivalent:

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

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

*Konstr.*  $a_* = X(X^T X)^{-1} c$

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

*Bem.* Es gilt

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

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

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

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

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

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

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

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

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

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

$\mathbb{E}Z = O^T \mathbb{E}Y = O^T X\beta = \begin{pmatrix} o_1^T \\ o_2^T \end{pmatrix} X\beta$

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

$\mathbb{E}Z = \begin{pmatrix} o_1^T X\beta \\ o_2^T X\beta \end{pmatrix}$ ,  $\text{Cov}(Z) = \text{Cov}(O^T Y) = O^T \text{Cov}(Y) O = \sigma^2 I$

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

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

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

eine erwartungstreue Schätzung für  $\sigma^2$ .

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

- Die ML-Schätzung für  $\beta \in \mathbb{R}^p$  stimmt mit der MkQ-Schätzung  $\hat{\beta}$  überein und es gilt  $\hat{\beta} \sim \mathcal{N}_p(\mathbb{E}\hat{\beta}, \text{Cov}(\hat{\beta}))$ .

- Die ML-Schätzung für  $\sigma^2$  lautet  $\hat{\sigma}_n^2 = \frac{S(Y, \hat{\beta})}{n} = \frac{n-r}{n} \hat{\sigma}^2$  mit  $\mathbb{E}\hat{\sigma}_n^2 = \frac{n-r}{n} \sigma^2 \xrightarrow{n \rightarrow \infty} \sigma^2$  (asympt. erw.-treu) und  $\frac{S(Y, \hat{\beta})}{\sigma^2} \sim \chi_{n-r}^2$

- Für einen Vektor  $l^T(\beta) = (l_1(\beta), \dots, l_q(\beta))$  von  $q \leq r$  linear unabhängigen schätzbaren Funktionen  $l_i(\beta) = c_i^T \beta$ ,  $c_i \in \mathbb{R}^p$  gilt  $\hat{l} := l(\hat{\beta}) \sim \mathcal{N}_q(l(\beta), \sigma^2 A_* A_*^T)$  mit  $\text{rk } A_* = q$ , wobei

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

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

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

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

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

Konfidenzschätzung für den Vektor der schätzbaren Funktionen

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

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