Zusammenfassung Stochastik 3

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

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

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

Def. Der **Ablehungsbereich** 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. 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 φ ist

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

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

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

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

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

Bem. Es gilt

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

Letzteres Intervall wird **Konfidenzintervall** für μ_0 zum Konfidenzniveau $1 - \alpha$ genannt.

Bsp. 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.
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.

Bsp. 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 $T\ \chi_{n-1}^2$. Falls μ bekannt ist, muss man

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

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

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

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

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

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

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

Bsp. 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 $H_0 : \rho = \rho_0 \in (0,1)$ 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 $T \sim \mathcal{N}(0,1)$ 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 von 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, 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 $X \sim \text{Exp}(\mu)$, $\hat{\mu}_n := \frac{1}{\overline{X}_n}$. Dann gilt

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

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

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

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

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

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

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

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

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

Bem. Bei $k = 2$ schreibt man oft

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

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

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

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

Die Klassenhäufigkeiten sind nun multinomialverteilt:

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

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

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

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

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

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

Bemn.

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

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

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

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

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

Ziel. Überprüfen, ob die Komponenten X und Y 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

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

Testregel: Die Nullhypothese 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

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

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

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

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

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

Satz (Gliwenko, Hauptsatz der math. Statistik).

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

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

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

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

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

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

unabhängig von F.

Satz. Falls F stetig ist, gilt

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

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 $H_1: F\ne F_1$. Dabei muss F_0 eine stetige VF sein. Wir verwenden dazu

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

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 $\hat F_n$.

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

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

Es 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.

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

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

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 $\omega_n^2(g) > \omega^2_{1-\alpha}(g)$.

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 $H_1: F\ne F^*$ gilt, indem wir die empirischen VFen $\hat F_n$ und $\hat F_m^*$ vergleichen. Dazu verwenden wir

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

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

$$T_{m,n}\overset{d}{=}\sqrt{\tfrac{m\cdot n}{m+n}}\sup_{0\le u\le 1}\left|\tfrac1n\sum_{i=1}^{n}\mathbb 1_{[0,u]}(U_i)-\tfrac1m\sum_{j=1}^{m}\mathbb 1_{[0,u]}(U_j^*)\right|,$$

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

$$F^-(t)\coloneqq\begin{cases}\min\{x\in\mathbb R^1\mid F(x)\ge t\} & 0<t\le 1,\\ \lim_{t\downarrow 0}F^-(t) & t=0.\end{cases}\qquad\text{(Quantilfunktion)}$$

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 $T_{m,n}>K_{1-\alpha}$.

Situation. 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. $H_1: F\ne F^*$. Konstruktion einer Rangstatistik für konkrete SPn $x_1,\dots,x_n$ 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 Gesamt-SP:

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

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. Testgröße:

$$W_{m,n}\coloneqq\sum_{i=1}^{n}\sum_{j=1}^{m}\mathbb 1_{\{X_i<X_j^*\}}=|\{(i,j)\mid X_i<X_j^*\}|=\sum_{j=1}^{m}|\{i\mid X_i<X_{j:m}^*\}|$$
$$=\sum_{j=1}^{m}(\nu_j-j)=\nu_1+\dots+\nu_m-\tfrac m2(m+1)$$

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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 $h(x_1, x_2) \coloneqq \frac{1}{2}(x_1 - x_2)^2$. Damit haben wir

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

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 (**Hoeffdings Projektionsmethode**)**.** Sei

$$\begin{aligned} g(x) &= \mathbb{E}[h(X_1, \ldots, X_n)|X_1 = x] = \mathbb{E}h(x, X_2, \ldots, X_m) \\ &= \int \cdots \int h(x, x_2, \ldots, x_n)\, \mathrm{d}F(x_2) \cdots \mathrm{d}F(x_n) \end{aligned}$$

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

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

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

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

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

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

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