

# 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  $\vartheta$  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

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

**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

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

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

Der Ablehnungsbereich der Statistik ist

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

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  $H_0$ .

**Lemma.** Seien  $(X_n), (Y_n)$  zwei 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

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

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

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

• 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

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

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

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

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

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

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

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

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

**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  $T_{n,s+1} > \chi_{s,1-\alpha}^2$ .

*Bemerkungen.* •  $T_{n,s+1}$  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	$h_{21}^{(n)}$	$h_{22}^{(n)}$	...	$h_{2l}^{(n)}$	$h_{2\bullet}^{(n)}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$
$k$	$h_{k1}^{(n)}$	$h_{k2}^{(n)}$	...	$h_{kl}^{(n)}$	$h_{k\bullet}^{(n)}$
	$h_{\bullet 1}^{(n)}$	$h_{\bullet 2}^{(n)}$	...	$h_{\bullet l}^{(n)}$	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  $Y$  durch Maximieren von

$$L(h_{1\bullet}^{(n)}, \dots, h_{k\bullet}^{(n)}, h_{\bullet 1}^{(n)}, \dots, h_{\bullet l}^{(n)}; p_{1\bullet}, \dots, p_{k-1,\bullet}, p_{\bullet 1}, \dots, p_{\bullet, l-1})$$

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

Wir haben dann

$$T_{k,l}^{(n)} := \sum_{i=1}^k \sum_{j=1}^l \frac{(h_{ij}^{(n)} - np_{i\bullet} p_{\bullet j})^2}{n \cdot p_{i\bullet} p_{\bullet j}}$$