# Statistik

**0.1 Definition** (statistisches Modell)**.** Ein Tripel $\mathcal{X} = (\Omega, \mathcal{F}, P_\vartheta)_{\vartheta \in \Theta}$ heißt *statistisches Modell*, falls gilt: $\Theta$ ist eine beliebige Indexmenge und für jedes $\vartheta \in \Theta$ ist $(\Omega, \mathcal{F}, P_\vartheta)$ ein W-Raum. Das Modell heißt

(i) *parametrisch*, falls $\Theta \subset \mathbb{R}^m$ für irgendein m $\in \mathbb{N}$.

(ii) *diskret*, falls $(\Omega, \mathcal{F})$ ein diskreter messbarer Raum ist.

(iii) *stetig*, falls $\Omega \subset \mathbb{R}^m$, $\mathcal{F} = \mathcal{B}^m_\Omega$ und jedes $P_\vartheta$ eine Lebesgue-Dichte $f_\vartheta$ besitzt.

(iv) *Standardmodell*, falls $\mathcal{X}$ stetig oder diskret ist.

**0.2 Definition** (Schätzer)**.** Sei $\mathcal{X} = (\Omega, \mathcal{F}, P_\vartheta)_{\vartheta \in \Theta}$ ein statistisches Modell, $(\Sigma, \mathcal{S})$ ein beliebiger messbarer Raum und $g : \Theta \to \Sigma$ eine beliebige Abbildung, genannt *Kenngröße*. Eine beliebige Zufallsvariable $T : (\Omega, \mathcal{F}) \to (\Sigma, \mathcal{S})$ heißt *Statistik* oder auch *Schätzer für* $g$.

**0.3 Remark.** Der Schätzer $T$ muss i.A. also nichts mit der Kenngröße $g$ zu tun haben. So ein Schätzer heißt dann *schlecht*.

**0.4 Definition** (Bias)**.** Seien $\mathcal{X} = (\Omega, \mathcal{F}, P_\vartheta)_{\vartheta \in \Theta}$ ein statistisches Modell, $g : \Theta \to \mathbb{R}$ eine Kenngröße und $T :: (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}_\mathbb{R})$ ein Schätzer für $g$. Dann heißt

$$\begin{aligned} b : \Theta &\to \mathbb{R} \\ \vartheta &\mapsto E_\vartheta(T - g(\vartheta)) \end{aligned}$$

*bias von* $T$. Ein Schätzer $T$ heißt *erwartungstreu*, falls $b \equiv 0$.

**0.5 Definition** (Likelihood-Funktion)**.** Sei $\mathcal{X} = (\Omega, \mathcal{F}, P_\vartheta)_{\vartheta \in \Theta}$ ein parametrisches Standardmodell. Die Abbildung

$$\begin{aligned} L : \Omega \times \Theta &\to \mathbb{R} \\ (\omega, \vartheta) &\mapsto \rho_\vartheta(\omega), \end{aligned}$$

wobei

$$\rho_\vartheta(\omega) := \begin{cases} P_\vartheta(\{\omega\}), & \text{falls } \mathcal{X} \text{ diskret,} \\ f_\vartheta(\omega), & \text{falls } \mathcal{X} \text{ stetig mit Dichten } f_\vartheta, \end{cases}$$

heißt *Likelihood-Funktion*. Ein Schätzer $T : (\Omega, \mathcal{F}) \to (\Theta, \mathcal{B}^m_\Omega)$ heißt *Maximum-Likelihood-Schätzer* (für id : $\Theta \to \Theta$), falls

$$\forall \omega \in \Omega : L(\omega, T(\omega)) = \max_{\vartheta \in \Theta} L(\omega, \vartheta).$$

Ist $g : \Theta \to \Sigma$ eine beliebige Kenngröße so heißt dann $g \circ T$ ein *Maximum-Likelihood-Schätzer* für $g$.

**0.6 Definition.** Sei $\mathcal{X} = (\Omega, \mathcal{F}, P_\vartheta)_{\vartheta \in \Theta}$ ein statistisches Modell, $\Sigma$ eine Menge, $g : \Theta \to \Sigma$ eine Kenngröße und $0 < \alpha < 1$. Eine Abbildung $C : \Omega \to \mathcal{P}(\Omega)$ heißt *Konfidenzbereich zum Irrtumsniveau* $\alpha$, falls gilt:

(i) Für alle $s \in \Sigma$ gilt $\{s \in C(\_)\} := \{\omega \in \Omega \mid s \in C(\omega)\} \in \mathcal{F}$.

(ii) Es gilt $\inf_{\vartheta\in\Theta} P_{\vartheta(g(\vartheta)\in C(\_))} \geq 1-\alpha$

Man nennt dann auch $1-\alpha$ das *Sicherheitsniveau.* Ist $\Sigma = \mathbb{R}$ und sind alle $C(\omega)$ Intervalle, dann nennt man $C$ auch ein *Konfidenzintervall.*

**0.7 Definition.** Sei $P$ ein W-Maß auf $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ mit Verteilungsfunktion $F$ und $0 < \alpha < 1$. Dann heißt

$$\inf_{x\in\mathbb{R}} F(x) \geq \alpha, \qquad\qquad \{\sup x \in \mathbb{R} \mid F(x) \leq \alpha\}$$

*unteres* bzw. *oberes $\alpha$-Quantil von $P$.*

**0.8 Definition** (Hypothesentest)**.** Sei $\mathcal{X} = (\Omega, \mathcal{F}, P_\vartheta)_{\vartheta\in\Theta}$ ein statistisches Modell. Es sei ferner $\Theta = \Theta_0 \dot\cup \Theta_1$ eine disjunkte Zerlegung mit $\Theta_0 \neq \emptyset$.

(i) Es heißt dann $\Theta_0$ *Nullhypothese* und $\Theta_1$ *Alternative.*

(ii) Eine Statistik $T : (\Omega, \mathcal{F}) \to ([0,1], \mathcal{B}_{[0,1]})$ heißt *Hypothesentest.* Falls $T(\Omega) \in \{0,1\}$, so heißt $T$ *nichtrandomisiert* und *randomisiert* andernfalls. Ist $T$ nichtrandomisiert, so heißt $\{\omega \in \Omega \mid T(\omega) = 1\}$ der *Verwerfungsbereich.*

(iii) Die Funktion

$$\begin{aligned} G_T : \Theta &\to [0,1] \\ \vartheta &\mapsto E_\vartheta(T) \end{aligned}$$

heißt *Gütefunktion* von $T$.

(iv) Die Größe $\sup_{\vartheta\in\Theta} G_T(\vartheta)$ heißt *(effektives) Niveau* von $T$. Wir sagen $T$ ist ein *Test zum Irrtumsniveau* $\alpha$, falls $\sup_{\vartheta\in\Theta} E_\vartheta(T) \leq \alpha$.

**0.9 Remark.** Fehler erster Art: Entscheidung für die Alternative, obwohl die Nullhypothese vorliegt. Fehler zweiter Art: Entscheidung für die Nullhypothese, obwohl die Alternative vorliegt.

**0.10 Definition.** Ein Test $T$ von $\Theta_0$ gegen $\Theta_1$ heißt *(gleichmäßig) bester Test zum Niveau* $\alpha$, falls er vom Niveau $\alpha$ ist und für jeden anderen solchen Test $S$ gilt:

$$\forall \vartheta \in \Theta_1 : G_T(\vartheta) \geq G_S(\vartheta).$$

**0.11 Definition.** Ein Test $T$ heißt *unverfälscht zum Niveau $\alpha$*, falls

$$\forall (\vartheta_0, \vartheta_1) \in \Theta_0 \times \Theta_1 : G_T(\vartheta_0) \leq \alpha \leq G_T(\vartheta_1)$$