1 Grundbegriffe, Motivation

Sei $(\Omega, \mathcal{A}, \mathbb{P})$ Wahrscheinlichkeitsraum, $(\mathfrak{X}, \mathcal{B})$ messbarer Raum[1] (sogenannter Stichprobenraum).
$X: \ \Omega \to \mathfrak{X}$ Zufallsvariable

$$P(B) := P^X(B) := \mathbb{P}(X^{-1}(B)), \ B \in \mathcal{B}$$

Verteilung von X ($\hookrightarrow$ Wahrscheinlichkeitsraum $(\mathfrak{X}, \mathcal{B}, P)$)

Statistischer Entscheidung liegt Datenmaterial (Beobachtung) $x \in \mathfrak{X}$ zugrunde.
Grundannahme:

1) $x = X(\omega)$ für ein $\omega \in \Omega$, d.h. x ist Realisierung von X
2) P ist (teilweise) unbekannt

Ziel: Aufgrund von x Aussagen über P machen!

Sei $\mathcal{M}^1(\mathfrak{X}, \mathcal{B}) := \{P: \ P \text{ ist Wahrscheinlichkeitsmaß auf } \mathcal{B}\}$.

1.1 Definition

Eine Verteilungsannahme ist eine Teilmenge $\wp \subset \mathcal{M}^1(\mathfrak{X}, \mathcal{B})$.
Das Tripel $(\mathfrak{X}, \mathcal{B}, \wp)$ heißt statistischer Raum (statistisches Modell).

1.2 Beispiel

$(\mathfrak{X}, \mathcal{B}) = (\mathbb{R}^n, \mathcal{B}^n)$

$$\wp := \{P: \ \exists \text{ Wahrscheinlichkeitsmaß } Q \text{ auf } \mathcal{B}^1 \text{ mit } P = \underbrace{Q \otimes \ldots \otimes Q}_{n \text{ Faktoren}}\}$$

Mit anderen Worten $X = (X_1, \ldots, X_n)$, $X_1, \ldots, X_n$ stochastisch unabhängig mit gleicher Verteilung Q, $X_1, \ldots, X_n \overset{\text{uiv}}{\sim} Q$.

[1] $\mathcal{B}$ steht hier für eine beliebige σ-Algebra, die Borelsche σ-Algebra wird mit $\mathcal{B}^d$ bezeichnet, wobei d die Dimension angibt

1.3 Beispiel

$(\mathfrak{X}, \mathcal{B}) = (\mathbb{R}^n, \mathcal{B}^n)$

$$\wp := \{P : \exists(\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}_{>0} \text{ mit } P = \mathcal{N}(\mu, \sigma^2) \otimes \ldots \otimes \mathcal{N}(\mu, \sigma^2)\}$$

Also $X_1, \ldots, X_n \overset{\text{uiv}}{\sim} \mathcal{N}(\mu, \sigma^2)$.
Ein-Stichproben-Normalverteilungs-Annahme

1.4 Beispiel

$(\mathfrak{X}, \mathcal{B}) = (\mathbb{R}^{m+n}, \mathcal{B}^{m+n})$

$$\wp := \{P : \exists \text{ W'maße } Q_1, Q_2 : \ P = \underbrace{Q_1 \otimes \ldots \otimes Q_1}_{m \text{ Faktoren}} \otimes \underbrace{Q_2 \otimes \ldots \otimes Q_2}_{n \text{ Faktoren}}\}$$

Also $X = (X_1, \ldots, X_m, Y_1, \ldots, Y_n)$, $X_1, \ldots, X_m, Y_1, \ldots, Y_n$ unabhängig,
$X_1, \ldots, X_m \overset{\text{uiv}}{\sim} Q_1, Y_1, \ldots, Y_n \overset{\text{uiv}}{\sim} Q_2$.

1.5 Beispiel

$(\mathfrak{X}, \mathcal{B})$ wie in 1.4

$$\wp := \{P : \exists(\mu, \nu, \sigma^2) \in \mathbb{R}\times\mathbb{R}\times\mathbb{R}_{>0} : \ P = \mathcal{N}(\mu, \sigma^2)\otimes\ldots\otimes\mathcal{N}(\mu, \sigma^2)\otimes\mathcal{N}(\nu, \sigma^2)\otimes\ldots\otimes\mathcal{N}(\nu, \sigma^2)\}$$

$X_1, \ldots, X_m, Y_1, \ldots, Y_n$ unabhängig
$X_i \overset{\text{uiv}}{\sim} \mathcal{N}(\mu, \sigma^2), Y_j \overset{\text{uiv}}{\sim} \mathcal{N}(\nu, \sigma^2)$
2 unabhängige normalverteilte Stichproben mit gleicher Varianz

1.6 Definition

Eine **Parametrisierung** von $\wp \subset \mathcal{M}^1(\mathfrak{X}, \mathcal{B})$ ist eine bijektive Abbildung $\Theta \ni \vartheta \to P_\vartheta \in \wp$.

Ist X eine Zufallsvariable mit Verteilung P_ϑ, so schreibt man auch

$$\left.\begin{array}{l} E_\vartheta(X) \\ \mathrm{Var}_\vartheta(X) \\ (*) \ F_\vartheta(t) := P_\vartheta(X \le t) = P_\vartheta((-\infty, t]) \end{array}\right\} \text{falls X reellwertig}$$

$$(**)\ P_\vartheta(B) = P_\vartheta(X \in B),\ B \in \mathcal{B}$$

Schreibweisen $(*), (**)$ unterstellen

$$(\Omega, \mathcal{A}, \mathbb{P}) = (\mathfrak{X}, \mathcal{B}, P),\ X = \mathbf{id}_\Omega$$

[eigentlich: $P_\vartheta(B) := \mathbb{P}(X^{-1}(B)) = \mathbb{P}(X \in B)$]

1.7 Definition

Eine Verteilungsklasse $\wp = \{P_\vartheta :\ \vartheta \in \Theta\}$ heißt ϑ-parametrisch, wenn sie sich „zwanglos“ durch einen Parameterraum $\Theta \subset \mathbb{R}^k$ parametrisieren lässt. Ist $\vartheta = (\vartheta_1, \vartheta_2)$ und interessiert nur ϑ_1, so heißt ϑ_1 Hauptparameter und ϑ_2 Nebenparameter oder Störparameter.

1.8 Beispiele

a) In Beispiel 1.3:
   2-parametrige Verteilungsannahme, wobei $\vartheta = (\mu, \sigma^2)$, $\Theta = \mathbb{R} \times \mathbb{R}_{>0}$.

b) In Beispiel 1.5:
   3-parametrig, $\vartheta = (\mu, \nu, \sigma^2)$, $\Theta = \mathbb{R} \times \mathbb{R} \times \mathbb{R}_{>0}$
   Hier meistens: (μ, ν) Hauptparameter

Häufig interssiert von $\wp$ der Wert eines reellwertigen Funktionals $\gamma :\ \wp \to \mathbb{R}$ anstelle von P, z.B. (falls P Wahrscheinlichkeitsmaß auf $\mathcal{B}^1$)

$$\gamma(P) := \int_{\mathbb{R}} x dP(x)$$

(Erwartungswert von X)

Falls $\wp = \{P_\vartheta :\ \vartheta \in \Theta\}$, so schreibt man auch $\gamma(\vartheta) := \gamma(P_\vartheta)$, fasst also γ als Abbildung $\gamma :\ \Theta \to \mathbb{R}$ auf.

Problem:
„Enge“ Verteilungsannahme täuscht oft nicht vorhandene Genauigkeit vor.
$\wp$ sollte das wahre P enthalten. (realistisch?)

Bei diskreten Zufallsvariablen ergibt sich $\wp$ manchmal zwangsläufig; bei stetigen Zufallsvariablen ist $\wp$ häufig nicht vorgezeichnet.

1.9 Typische Fragestellungen der Statistik

a) Punktschätzung
Schätze aufgrund von $x \in \mathfrak{X}$ den Wert $\gamma(\vartheta) \in \mathbb{R}$ möglichst „gut“.

b) Konfidenzbereiche
Konstruiere „möglichst kleinen“, von x abhängigen Bereich, der $\gamma(\vartheta)$ mit „großer Wahrscheinlichkeit“ enthält.

c) Testprobleme
Es sei $\Theta = \Theta_0 + \Theta_1$ eine Zerlegung von Θ. Teste die Hypothese $H_0 :\ \vartheta \in \Theta_0$ gegen die Alternative $H_1 :\ \vartheta \in \Theta_1$.

1.10 Asymptotische Betrachtungen

Häufig liegt Folge $(X_j)_{j\in\mathbb{N}}$ unabhängiger Zufallsvariablen zugrunde (alle auf nicht interessierenden Wahrscheinlichkeitsräume $(\Omega, \mathcal{A}, \mathbb{P})$ definiert) mit Werten in einem Messraum $(\mathfrak{X}_0, \mathcal{B}_0)$.
Häufig: $P^{X_j} = P\ \forall j$ (identische Verteilung)

Unter der Verteilungsannahme $P \in \wp_0 \subset \mathcal{M}^1(\mathfrak{X}_0, \mathcal{B}_0)$ nimmt dann die Folge $(X_j)_{j\in\mathbb{N}}$ Werte im statistischen Raum

$$(\mathfrak{X}, \mathcal{B}, \wp) := (\times_{j=1}^{\infty}\mathfrak{X}_0, \bigotimes_{j=1}^{\infty}\mathcal{B}_0, \{\bigotimes_{j=1}^{\infty}P :\ P \in \wp_0\})$$

an. Also: $X_1, X_2, \ldots$ unabhängig, $\mathfrak{X}_0$-wertig mit gleicher Verteilung $P \in \wp_0$

1.11 Statistiken

Es seien $(\mathfrak{X}, \mathcal{B})$ Stichprobenraum und $(\mathcal{T}, \mathcal{D})$ Messraum. Eine messbare Abbildung $T :\ \mathfrak{X} \to \mathcal{T}$ heißt Statistik (Stichprobenfunktion).
Häufig: $(\mathcal{T}, \mathcal{D}) = (\mathbb{R}, \mathcal{B}^1)$.

Wichtigstes Beispiel:
$\mathfrak{X} = \mathbb{R}^n, \mathcal{T} = \mathbb{R}$

$$T(x_1, \ldots, x_n) = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Stichproben-Funktionen bewirken eine **Datenkompression**.
Statistische Entscheidungen wie Ablehnung von Hypothesen hängen von $x \in \mathfrak{X}$ im Allgemeinen durch den Wert $T(x)$ einer geeigneten Statistik ab.
Bei Tests: Statistik $\hat{=}$ Testgröße $\hat{=}$ Prüfgröße

Sind $X_1, \ldots, X_n \overset{uiv}{\sim} P$, so nennt man $X_1, \ldots, X_n$ eine Stichprobe vom Umfang n aus der Verteilung P.
Ist $T(X_1, \ldots, X_n)$ eine mit $X_1, \ldots, X_n$ operierende Statistik, so schreib man auch $T_n := T_n(X_1, \ldots, X_n) := T(X_1, \ldots, X_n)$.
Insbesondere bei bei asymptotischen Betrachtungen ist $(T_n)_{n \geq 1}$ dann eine Folge von Statistiken.