3. Schließende Statistik

3.1 Statistische Modelle

Hajo Holzmann

Philipps-Universität Marburg

Grundidee der induktiven Statistik

- Interpretiere Beobachtungen $x_1, \dots, x_n$ als
  $\rightarrow$ Realisierungen von u.i.v. Z.V. $X_1, \dots, X_n$
- Verteilung P der X_i quantifiziert Unsicherheit/Schwankungen in den Daten x_i.

Grundidee der induktiven Statistik

- Interpretiere Beobachtungen $x_1, \ldots, x_n$ als
  $\rightarrow$ Realisierungen von u.i.v. Z.V. $X_1, \ldots, X_n$
- Verteilung P der X_i quantifiziert Unsicherheit/Schwankungen in den Daten x_i.

Aber: Verteilung P unbekannt, ziehe aus Daten $x_1, \ldots, x_n$ Rückschlüsse auf P.

Dazu: Statistisches Modell für P.

Sei

$$\mathcal{M}_1(\mathbb{R}) = \{P \text{ Wahrscheinlichkeitsmaß auf } \mathbb{R}\}.$$

Definition: Eine Teilmenge $\mathcal{P} \subset \mathcal{M}_1(\mathbb{R})$ heißt ein statistisches Modell.

Statistische Modelle

Sei

$$\mathcal{M}_1(\mathbb{R}) = \{P \text{ Wahrscheinlichkeitsmaß auf } \mathbb{R}\}.$$

Definition: Eine Teilmenge $\mathcal{P} \subset \mathcal{M}_1(\mathbb{R})$ heißt ein statistisches Modell.

Modell für Beobachtungen $x_1, \ldots, x_n$:

Realisierungen von *u.i.v. Z.V.* $X_1, \ldots, X_n$ mit

$X_i \sim P$ für *ein (unbekanntes)* $P \in \mathcal{P}$.

Definition: Ein statistisches Modell $\mathcal{P} = (P_\theta)_{\theta \in \Theta}$, in welchem die Elemente P_θ durch endlichdim. Parameter $\theta \in \Theta \subset \mathbb{R}^p$ indiziert sind (in natürlicher Weise), heißt ein parametrisches Modell.

Definition: Ein statistisches Modell $\mathcal{P} = (P_\theta)_{\theta \in \Theta}$, in welchem die Elemente P_θ durch endlichdim. Parameter $\theta \in \Theta \subset \mathbb{R}^p$ indiziert sind (in natürlicher Weise), heißt ein parametrisches Modell.

- θ heißt der Parameter von $\mathcal{P}$,
- Θ heißt der Parameterraum

Bernoulli - Modell:

$$\mathcal{P} = \{P = \text{Ber}(p) \text{ für ein } p \in [0,1].\}.$$

$\rightarrow \theta = p$, $\Theta = [0,1]$.

Je Versuchdurchgang zwei mögliche Kategorien beobachten.

Gesamt: $X_1, \ldots, X_n$ u.i.v., $X_i \sim Ber(p)$, $p \in [0,1]$.

Bernoulli - Modell:

$$\mathcal{P} = \{P = \mathrm{Ber}(p) \text{ für ein } p \in [0,1].\}.$$

$\rightarrow \theta = p$, $\Theta = [0,1]$.

Je Versuchdurchgang zwei mögliche Kategorien beobachten.

Gesamt: $X_1, \ldots, X_n$ u.i.v., $X_i \sim Ber(p)$, $p \in [0,1]$.

Bsp.:

- Werfen einer Reizzwecke
- Tea-tasting lady: Lady pro Durchgang zwei Tassen Tee, eine erst Milch, dann Tee, die andere erst Tee, dann Milch. Muss richtig bestimmen.

Multinomial - Modell: Für festes $k \in \mathbb{N}$

$$\mathcal{P} = \{P = \mathrm{Mult}(1; p_1, \ldots, p_k), \; p_i > 0, \; p_1 + \ldots + p_k = 1.\}.$$

$\rightarrow \theta = (p_1, \ldots, p_k)$, *Theta*: k-Simplex.

Je Versuchdurchgang k mögliche Kategorien beobachten.

Also: $X_1, \ldots, X_n$ u.i.v., $X_i \sim \mathrm{Mult}(1; p_1, \ldots, p_k)$.

Multinomial Modell

Multinomial - Modell: Für festes $k \in \mathbb{N}$

$$\mathcal{P} = \{P = \text{Mult}(1; p_1, \ldots, p_k), \; p_i > 0, \; p_1 + \ldots + p_k = 1.\}.$$

$\rightarrow \theta = (p_1, \ldots, p_k)$, *Theta*: k-Simplex.

Je Versuchdurchgang k mögliche Kategorien beobachten.

Also: $X_1, \ldots, X_n$ u.i.v., $X_i \sim \text{Mult}(1; p_1, \ldots, p_k)$.

Bsp.:

- Beobachten Autotyp
- Augenfarbe Person

Hypergeometrisches Modell

Hypgergeometrisches Modell: Für ein festes $m \in \mathbb{N}$

$$\mathcal{P} = \{P = \text{Hyper}(; m, R, M) \text{ für } R, M \in \mathbb{N}, R \leq M.\}$$

Beobachte meist nur ein $X \sim \text{Hyper}(; m, R, M)$ $(n = 1)$.

Hypgergeometrisches Modell: Für ein festes $m \in \mathbb{N}$

$$\mathcal{P} = \{P = \text{Hyper}(; m, R, M) \text{ für } R, M \in \mathbb{N}, R \leq M.\}$$

Beobachte meist nur ein $X \sim \text{Hyper}(; m, R, M)$ $(n = 1)$.

Bsp.:

- M Größe Population, R Personen mit hohem Blutdruck.
- Es werden m verschiedene Personen beobachtet und jeweils der Blutdruck festgestellt.
- Bei X wird hoher Blutdruck beobachtet.

Poisson Modell

$$\mathcal{P} = \{P = \mathrm{Poi}(\lambda) \text{ für } \lambda > 0\}.$$

$\rightarrow \theta = \lambda$, $\Theta = (0, \infty)$.

Modell für Zähldaten.

addiere viele mögliche, einzeln relativ unwahrscheinliche Ereignisse.

Poisson Modell

$$\mathcal{P} = \{P = \text{Poi}(\lambda) \text{ für } \lambda > 0\}.$$

$\rightarrow \theta = \lambda$, $\Theta = (0, \infty)$.

Modell für Zähldaten.

addiere viele mögliche, einzeln relativ unwahrscheinliche Ereignisse.

Bsp.: Rutherford-Geiger-Experiment.

- Anzahl der Zerfälle in radioaktiven Präparat
- über 2608 Zeitintervalle von je 7.5 Sekunden.

- **Normalverteilungs-Modell**

$$\mathcal{P} = \{P = \mathrm{N}(\mu, \sigma^2) \text{ für } \mu \in \mathbb{R}, \sigma^2 > 0\}.$$

$\rightarrow \theta = (\mu, \sigma^2)$, $\Theta = \mathbb{R} \times (0, \infty)$.

Modell für stetige Daten.

ZGWS: Summe vieler kleiner, unabhängiger Einflüsse

Bsp.: Körpergröße, Messfehler.

- **Normalverteilungs-Modell**

$$\mathcal{P} = \{P = \mathrm{N}(\mu, \sigma^2) \text{ für } \mu \in \mathbb{R}, \sigma^2 > 0\}.$$

$\rightarrow \theta = (\mu, \sigma^2)$, $\Theta = \mathbb{R} \times (0, \infty)$.

Modell für stetige Daten.

ZGWS: Summe vieler kleiner, unabhängiger Einflüsse

Bsp.: Körpergröße, Messfehler.

- **Exponential-Modell**

$$\mathcal{P} = \{P = \mathrm{Ex}(\lambda) \text{ für } \lambda > 0\}.$$

$\rightarrow$ Modell für positive, stetige Daten.

Gedächtnislosigkeit: Überlebensdauern falls gedächtnislos.

Modell-Vergleich

- Modell direkt aus Experiment:

  Bernoulli-Modell, Multinomial-Modell, Hypergeometrisches Modell.

- Modell aus Datentyp:

  Poisson-Modell, Exponential-Modell, Normalverteilungsmodell.

3.1.3 Nichtparametrische Modelle

Modelle, die nicht durch einen endlichdimensionalen Parameter indiziert werden.
$\rightarrow$ modellfreie Statistik.

Beispiele

1. Alle Wahrscheinlichkeitsmaße auf $\mathbb{R}$: $\mathcal{P} = \mathcal{M}_1(\mathbb{R})$.
2. Alle Wahrscheinlichkeitsmaße mit stetig-differenzierbarer Dichte

$$\mathcal{P} = \{P \in \mathcal{M}_1(\mathbb{R}), \quad P \text{ absolut stetig, Dichte } f \in C^1(\mathbb{R})\}.$$

3. Alle Wahrscheinlichkeitsmaße mit exitierendem Erwartungswert und Varianz:

$$\mathcal{P} = \{P \in \mathcal{M}_1(\mathbb{R}), \quad \text{Ist } X \text{ Z.V. }, X \sim P, \text{ so ist } E_P X^2 < \infty\}.$$