Beurteilende Statistik

- In der Wahrscheinlichkeitstheorie ist der erste Schritt jeglicher Rechnung die Modellierung, also die Wahl des Wahrscheinlichkeitsraums $(\Omega, \mathcal{A}, \mathbb{P})$.
- Danach versuchen wir zu berechnen, was man mit welcher Wahrscheinlichkeit von einer Stichprobe, die gemäß $\mathbb{P}$ aus Ω gewählt wird, erwarten sollte.
- In der beurteilenden Statistik fragt man nun andersherum:
- Angenommen, ich kenne das $\mathbb{P}$ nicht, aber ich kenne eine Stichprobe, die gemäß $\mathbb{P}$ gewählt wurde: Was kann ich aufgrund dieser Stichprobe über $\mathbb{P}$ aussagen?

- Dabei gibt es verschiedene Schwierigkeitsgrade für die Problemstellungen der beurteilenden Statistik.
- Das schwierigste und leider das realistischste Problem hiervon, dass wir nämlich *à priori* nichts oder so gut wie nichts über die Verteilung der Daten wissen, können wir hier noch nicht einmal ansatzweise behandeln.
- Dies fällt in den Bereich der sogenannten *nicht–parametrischen Statistik*.
- In der *parametrischen Statistik* gehen wir davon aus, dass wir schon von vornherein wissen, dass die Verteilung der beobachteten Daten ein Element einer gegebenen Familie $(\mathrm{P}_\theta)_{\theta \in \Theta}$ von Verteilungen auf Ω ist.
- Wir müssen also nur Kriterien finden, nach denen wir unsere Wahl treffen sollten.
- Hierbei soll archetypisch eine Situation genauer studiert werden, nämlich der *n*–fache Münzwurf.

- Seien also n Beobachtungen $x_1, \dots, x_n$ gegeben, die wir durch n–faches Werfen einer Münze erhalten haben, von der wir nicht wissen, mit welcher Wahrscheinlichkeit sie eine 0 bzw. eine 1 zeigt.
- Dementsprechend müssen wir uns für eine der Verteilungen $(P_p)_{p \in [0,1]}$ entscheiden, die durch

$$P_p = (p(0), p(1)) = (1 - p, p) \text{ für } p \in [0, 1]$$

definiert sind.
- Dabei sind alle Entscheidungen mit Unsicherheit behaftet, da die Stichprobe zufällig ist.

- Dabei kann unser Ziel nur sein, auf Basis der Beobachtungen folgende Aufgaben anzugehen:
  1. eine gute Schätzung abzugeben, welche Verteilung P_p dem Münzwurf zugrunde liegt.
  2. vorgegebene Hypothesen über P_p zu testen und sie anzunehmen oder zu verwerfen.
  3. eine Auswahl $\{P_p \mid p \in K \subset [0,1]\}$ anzugeben, unter der sich mit großer Wahrscheinlichkeit die dem Münzwurf zugrunde liegende Verteilung befindet.
- Wir werden uns den ersten beiden Punkten widmen.
- Im Folgenden seien also unabhängige Zufallsvariablen $X_1, \dots, X_n$ gegeben, die alle dieselbe Verteilung P_p für ein uns unbekanntes $p \in [0,1]$ besitzen.
- Die Verteilung des Vektors $(X_1, \dots, X_n)$ bezeichnen wir mit $\mathbb{P}_p$.

Eine Schätzung der unbekannten Erfolgswahrscheinlichkeit p unserer Münze sollte natürlich von den gemachten Beobachtungen abhängen. Dies führt uns auf die Definition einer Schätzfunktion.

Definition

Eine Schätzfunktion für p ist eine Funktion $\hat{p} : \{0, 1\}^n \to [0, 1]$.

Bilden wir $\hat{p}(x_1, \ldots, x_n)$ mit Beobachtungsdaten $x_1, \ldots, x_n$, so erhalten wir einen konkreten Schätzwert für p. Setzen wir die Zufallsvariablen ein und bilden

$$\hat{p}(X_1, \ldots, X_n),$$

so erhalten wir eine Zufallsvariable, den Schätzer.

Definition

Die Zufallsvariable $\hat{p}(X_1, \ldots, X_n)$ heißt *Schätzer für p*.

Das grundlegende Prinzip, das sogenannte *Maximum–Likelihood–Prinzip*, ist es, einen Schätzer für p zu konstruieren, der unter allen zur Verfügung stehenden Verteilungen diejenige auswählt, die die gemachte Beobachtung mit der größten Wahrscheinlichkeit misst.

Definition (Maximum–Likelihood–Schätzer)

Jeder Schätzer $\hat{p}$ für p mit

$$\mathbb{P}_{\hat{p}}(\{(x_1, \ldots, x_n)\}) = \max_{p \in [0,1]} \mathbb{P}_p(\{(x_1, \ldots, x_n)\})$$

für alle $x_1, \ldots, x_n \in \{0, 1\}$ heißt Maximum–Likelihood–Schätzer für p.

Der Maximum-Likelihood-Schätzer

Satz

Für die Erfolgswahrscheinlichkeit beim Münzwurf ist der naive Schätzer der einzige Maximum–Likelihood–Schätzer.

Der Maximum-Likelihood-Schätzer

Definition

Ein Schätzer $\hat{p}$ für p heißt *erwartungstreu*, wenn

$$\mathbb{E}_p\left[\hat{p}(X_1, \ldots, X_n) - p\right] = 0 \text{ für alle } p \in [0, 1].$$

Hierbei bezeichnet $\mathbb{E}_p$ den gemäß der Verteilung $\mathbb{P}_p$ berechneten Erwartungswert. Häufig schreibt man die Gleichung in der Definition der Erwartungstreue auch

$$\mathbb{E}_p[\hat{p}(X_1, \ldots, X_n)] = p \text{ für alle } p \in [0, 1].$$

Satz

Der naive Schätzer für p ist erwartungstreu.

Der Maximum-Likelihood-Schätzer

Definition

Ein Schätzer $\hat{p}$ heißt *konsistent*, wenn für alle $\varepsilon > 0$

$$\lim_{n\to\infty} \mathbb{P}_p(|\hat{p}(X_1, \ldots, X_n) - p| \geq \varepsilon) = 0 \quad \text{für alle } p \in [0, 1].$$

In dieser Definition der Konsistenz eines Schätzers erkennen wir den Konvergenzbegriff aus dem Schwachen Gesetz der großen Zahlen wieder. Für konsistente Schätzer wird es also immer wahrscheinlicher, dass der Schätzfehler klein ist.

Satz

Der naive Schätzer für p ist konsistent.

Erwartungstreue Schätzer

Satz

Seien $0 \leq \alpha_1, \ldots, \alpha_n \leq 1$ *reelle Zahlen mit* $\sum_{i=1}^n \alpha_i = 1$, *dann ist* $\hat{p} : \{0,1\}^n \to [0,1]$ *mit*

$$\hat{p}(x_1, \ldots, x_n) = \sum_{i=1}^n \alpha_i x_i$$

die Schätzfunktion eines erwartungstreuen Schätzers. Insbesondere ist jeder Schätzer $\hat{p}(X_1, \ldots, X_n) = X_i$, $1 \leq i \leq n$, *erwartungstreu.*

Quadratisches Risiko

Definition

Das *quadratische Risiko* eines erwartungstreuen Schätzers $\hat{p}$ für p ist seine Varianz

$$R(p, \hat{p}) = \mathbb{V}_p[\hat{p}(X_1, \ldots, X_n)].$$

Satz (Satz von Cramér und Rao)

Unter allen erwartungstreuen Schätzern für p besitzt der naive Schätzer das kleinste quadratische Risiko.

Das Testproblem

- Die grundsätzliche Frage beim Testproblem ist zu beurteilen, welche von zwei sich widersprechenden Hypothesen wahr ist.
- Etwa
  $H : p = p_0$ gegen $K : p = p_1$
  oder
  $H_0 : p \leq p_0$ gegen $H_1 : p > p_0$
- Problem:Egal, wie wir uns entscheiden, können wir zwei Fehler begehen, nämlich
- den Fehler 1. Art: H_0 ist wahr und wird verworfen
- den Fehler 2. Art: H_0 ist falsch und wird angenommen.

Das Testproblem

- Offensichtlich ist es unmöglich, beide Fehler gleichzeitig in den Griff zu bekommen.
- Versucht man beispielsweise den Fehler erster Art kleinzuhalten, indem man das Ablehnen von H_0 erschwert, so vergrößert dies unausweichlich die Chancen auf einen Fehler zweiter Art.
- In der Statistik hat sich nun eingebürgert, nach einem Test zu suchen, bei dem der Fehler 1. Art kleiner ist als ein vorgegebenes *Signifikanzniveau* α.
- Typische Werte für α sind 5%, 2,5%, 1% oder 0,5%.
- Die Wahl des Werts für α hängt natürlich von der jeweiligen Situation ab: je gravierendere Folgen ein Irrtum hätte, desto kleiner sollten wir α wählen.
- In unserem Beispiel hat ein Fehler 1. Art keine desaströsen Auswirkungen und wir wählen $\alpha = 5\%$.

- Das folgende Testverfahren liegt nahe:
- Wir schätzen p aus den Daten mithilfe des naiven Schätzers $\hat{p}$.
- Wenn $\hat{p}$ so aussieht, wie wir es unter H_0 vermuten würden (also z.B. nicht zu weit entfernt von 1/2 liegt, wenn $H : p = 1/2$), so nehmen wir H_0 an, sonst H_1.
- Hierbei tolerieren wir eine Wahrscheinlichkeit von α für einen Fehler 1. Art. Genauer sieht das Testverfahren so aus:
- Schätze p durch $\hat{p}$
- Bestimme eine Schranke Γ für p
- Akzeptiere H_0, falls $\hat{p} < \Gamma$ und verwerfe H_0 andernfalls.
- Der Wert für die Schranke $\Gamma = \Gamma(\alpha)$ berechnet sich dabei so, dass die Wahrscheinlichkeit für einen Fehler 1. Art höchstens gleich α ist.

Das Testproblem

Formal:

Definition

Zu testen sei die Hypothese $H \subset \Theta$ gegen die Alternative $K \neq \emptyset$. Ein Test is eine Abbildung

$$\phi : \{0,1\}^n \to \{0,1\}.$$

$\phi(x) = 0$ soll bedeuten, dass wir uns für H entscheiden, während $\phi(x) = 1$ bedeutet, wir entscheinden uns für K (wir lehnen die Hypothese ab). Ein Test ist vollständig festgelegt durch das Gebiet $R \subseteq \Theta$, auf dem wir die Hypothese verwerfen (R ist das Verwerfungsgebiet von ϕ), d.h. $\phi(x) = 1 \Leftrightarrow x \in R$.

Das Testproblem

Neben gewöhnlichen Tests betrachten wir auch randomisierte Tests:

Definition

Ein randomisierter Test ist eine Abbildung.

$$\phi : \{0, 1\}^n \longrightarrow [0, 1].$$

$\phi(x)$ ist die Wahrscheinlichkeit H abzulehnen.