Wahrscheinlichkeitsrechnung und Statistik für Studierende der Informatik

Vorlesung 13

Schätzprobleme

Allgemeines Setting:

Wir haben einen Wahrscheinlichkeitsraum (Ω, P). Die Menge Ω ist die Grundpopulation.

Auf diesem Wahrscheinlichkeitsraum haben wir eine Zufallsvariable X, die jedem $\omega \in \Omega$ einen Wert zuweist. Die Verteilung, Erwartungswert, Varianz, …. von X ist oft nicht bekannt.

Bsp.: Ω sind die Studierenden an der UDE, und X weist jedem $\omega \in \Omega$ die Körpergröße in cm zu.

Der/die fehlende/n Parameter soll anhand einer konkreten Stichprobe geschätzt werden. (Natürlich möglichst gut!).

Stichprobe von X: Vektor $(X_1, \ldots, X_n)$;
$X_1, \ldots, X_n$ unabhängig und gleichverteilt wie X (n-malige Auswertung von X)

Realisierung von $(X_1, \ldots, X_n)$/konkrete Stichprobe von X:
n reelle Zahlen $(x_1, \ldots, x_n) = (X_1(\omega), \ldots, X_n(\omega))$

Schätzprobleme

Beispiel S1: Wie groß ist die Wahrscheinlichkeit, dass eine gegebene Münze „Kopf“ zeigt?
Wir werfen die Münze 100-mal und erhalten dabei 60-mal „Kopf“.

In jedem Fall handelt es sich beim Münzwurf um ein Bernoulli-Experiment.

Betrachte die Zufallsvariable $X = \begin{cases} 1 & \text{"Kopf"} \\ 0 & \text{"Zahl"} \end{cases}$**, mit** $P(X = 1) = p$ **unbekannt.**

Mit obigen Angaben vermuten wir natürlich $p = \frac{60}{100} = \frac{3}{5}$**, da dies die (gemessene) relative Häufigkeit des Ereignisses „Kopf“ ist.**

Alternativ: Betrachte 100 **Kopien** $X_1, \ldots, X_{100}$ **von** X**, mit**

$X_i = 1 \iff$ **"Kopf" im** i**-ten Wurf**

Dann ist $Y = \sum_{i=1}^{100} X_i$ **binomialverteilt, mit** $P(Y = k) = \binom{100}{k} p^k (1-p)^{100-k}$

Finde nun $p \in [0,1]$ **so, dass** $P(Y = 60)$ **maximal wird.**

Das ist natürlich das erwartete $p = \frac{60}{100}$**.**

Schätzprobleme

Beispiel S2: Es soll geschätzt werden wie viele Fische in einem Teich leben. Was können wir tun?

An einem Tag werden 30 Fische gefangen und mit einem Punkt markiert.
Am nächsten Tag werden 40 Fische gefangen. Von diesen sind 5 mit einem Punkt markiert.

Wir vermuten, dass alle Fänge unabhängig voneinander sind. Dann sollte das Verhältnis von markierten Fischen im Teich dem Verhältnis von markierten Fischen im zweiten Fang entsprechen. D.h.:

$$\frac{30}{\text{Anzahl von Fischen im Teich}} = \frac{\text{Anzahl markierter Fische im Teich}}{\text{Anzahl von Fischen im Teich}} = \frac{5}{40}$$

Wir schätzen also

$$\text{Anzahl Fische im Teich} = 30 \cdot \frac{40}{5} = 240.$$

Formal: ...

Maximum-Likelihood-Methode

Definition 5.11. Es sei $(x_1, \ldots, x_n) \in \mathbb{R}^n$ eine konkrete Stichprobe vom Umfang n aus der Grundgesamtheit X, die von einem Parameter ϑ abhängt.

(i) Ist X stetig verteilt mit der Verteilungsdichte f, so heißt

$$L_x(\vartheta) = L(x_1, \ldots, x_n; \vartheta) = \prod_{k=1}^{n} f(x_k; \vartheta), \tag{5.2}$$

Likelihood-Funktion der konkreten Stichprobe $x = (x_1, \ldots, x_n)$.

(ii) Ist X diskret verteilt mit der Verteilung $P(X = x_i, \vartheta)$, $i = 1, 2 \ldots, n$, so heißt

$$L_x(\vartheta) = L(x_1, \ldots, x_n; \vartheta) = \prod_{k=1}^{n} P(X_k = x_k; \vartheta) \tag{5.3}$$

Likelihood-Funktion der konkreten Stichprobe $x = (x_1, \ldots, x_n)$.

Ziel: Maximiere $L_X(\theta)$.

Maximum-Likelihood-Methode

Definition 5.12. Nimmt die Likelihood-Funktion L_x mit $L_x(\vartheta) := L(x_1, \ldots, x_n; \vartheta)$ in $\hat{\vartheta}(x)$ ein Maximum an, d.h. gilt

$$L_x(\hat{\vartheta}(x)) := \sup\{L_x(\vartheta) \mid \vartheta \in \Theta\} ,$$

so heißt $\hat{\vartheta}(x)$ eine *Maximum-Likelihood-Schätzung* von ϑ.

Schätzprobleme

Stichprobe von X: Vektor $(X_1, \ldots, X_n)$;
$X_1, \ldots, X_n$ unabhängig und gleichverteilt wie X (n-malige Auswertung von X)
Realisierung von $(X_1, \ldots, X_n)$/konkrete Stichprobe von X:
n reelle Zahlen $(x_1, \ldots, x_n) = (X_1(\omega), \ldots, X_n(\omega))$

Sei θ ein noch unbekannter Parameter von X. Ziel ist es, eine geeignete Schätzfunktion $\hat{\theta}(X_1, \ldots, X_n)$ anhand einer Stichprobe von X herzuleiten.

Ist dann $(x_1, \ldots, x_n)$ eine Realisierung von $(X_1, \ldots, X_n)$, so nennen wir $\hat{\theta}(x_1, \ldots, x_n)$ eine Realisierung von $\hat{\theta}(X_1, \ldots, X_n)$ und dieser Wert ist unsere Schätzung für θ.

Definition 5.15. Sei $\hat{\theta}_n := \hat{\theta}(X_1, \ldots, X_n)$ eine Schätzfunktion für den Parameter θ.

(a) Die Schätzfunktion $\hat{\theta}_n$ heißt erwartungstreue Schätzfunktion für θ (*engl: unbiased*), wenn gilt
$$E_\theta\big(\hat{\theta}(X_1, \ldots, X_n)\big) = \theta \text{ für alle } \vartheta \in \Theta\,.$$

(b) Man nennt
$$B_\theta(\hat{\theta}_n) = E_\theta\big(\hat{\theta}(X_1, \ldots, X_n)\big) - \theta$$
den systematischen Fehler (*engl.* Bias).

Das ist uns schonmal begegnet:

Darstellung von Daten

Definition 5.1. a) Die Zahl

$$\overline{x} := \overline{x}_n := \frac{1}{n}\sum_{i=1}^{n} x_i$$

heißt *(Stichproben-)* oder *arithmetisches Mittel* bzw. kurz *Mittelwert* der Daten $x_1, \ldots, x_n$.

b) Die Zahl

$$s^2 := s_x^2 := \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - n\overline{x}^2\right)$$

heißt *(Stichproben-)* oder *empirische Varianz* der Daten $x_1, \ldots, x_n$. Die Zahl $s_x = \sqrt{s_x^2}$ heißt *(Stichproben-)* oder *empirische Standardabweichung* von $x_1, \ldots, x_n$.

Betrachten wir die Zufallsvariablen $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ und $S^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} X_i^2 - n\overline{X}^2\right)$, so ist $E(\overline{X}) = E(X)$ und $E(S^2) = Var(X)$.

Die Zahlen $\overline{x}$ und s^2 dienen also als „Schätzung" für $E(X)$ und $Var(X)$.

Schätzprobleme

Satz 5.17. *Es seien* $X_1, \ldots, X_n$ *unabhängig mit* $E_\vartheta(X_i) = \mu$ *und* $Var_\vartheta(X_i) = \sigma^2$ *für* $1 \leq i \leq n$ *und alle* $\vartheta \in \Theta$*; dann gilt:*

a) *Der Mittelwert* $\overline{X}(= \overline{X}(X_1, \ldots, X_n)) = \frac{1}{n}\sum_{i=1}^{n} X_i$ *ist ein erwartungstreuer Schätzer für* μ.

b) *Die empirische Varianz* $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ *ist ein erwartungstreuer Schätzer für* σ^2.

c) $\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$ *ist kein erwartungstreuer Schätzer für* σ^2.

Konfidenzintervalle

Bisher haben wir versucht einen unbekannten Parameter θ der Verteilung X durch eine konkrete Stichprobe punktgenau zu schätzen.

Mehr Erfolg haben wir natürlich, wenn wir „nur“ schätzen in welchem Intervall θ liegt.

Idee: Wir möchten anhand einer konkreten Stichprobe $(x_1, \ldots, x_2)$ von X ein Intervall $[\hat{\theta}_u, \hat{\theta}_o]$ bestimmen, so dass θ mit einer von uns vorgegebenen Wahrscheinlichkeit $1-\alpha$ in diesem Intervall liegt.

ABER: θ ist fest, nur unbekannt!

GENAUER: Wir suchen zwei Schätzfunktionen $\hat{\theta}_u(X_1, \ldots, X_n)$ und $\hat{\theta}_o(X_1, \ldots, X_n)$, so dass $P(\hat{\theta}_u \leq \theta \leq \hat{\theta}_o) = 1-\alpha$, für vorgegebenes $\alpha \in (0,1)$.

NOCH GENAUER: Wir suchen zwei Schätzfunktionen $\hat{\theta}_u(X_1, \ldots, X_n)$ und $\hat{\theta}_o(X_1, \ldots, X_n)$,

so dass $P_\theta(\hat{\theta}_u \leq \theta \leq \hat{\theta}_o) = 1-\alpha$, für vorgegebenes $\alpha \in (0,1)$ und alle möglichen θ!

Konfidenzintervalle

Definition 5.19. Ist θ ein zu schätzender Parameter und sind $\hat{\theta}_u$ und $\hat{\theta}_o$ Schätzfunktionen mit der Eigenschaft

$$P_\theta(\hat{\theta}_u \leq \theta \leq \hat{\theta}_o) = 1 - \alpha \text{ (oder auch } \geq 1 - \alpha)$$

- so liefert jede konkrete Stichprobe $x = (x_1, \ldots, x_n)$ ein *Konfidenzintervall* $[\hat{\theta}_u(x), \hat{\theta}_o(x)]$.
- Die Zahl $1 - \alpha$ heißt *Konfidenzniveau*. Die Zahl α heißt *Irrtumswahrscheinlichkeit.*

Häufig ist man in der Situation, die *unteren Konfidenz- oder Vertrauensgrenze* und die *obere Konfidenz- oder Vertrauensgrenze* für einen Parameter θ einzeln zu bestimmmen:

- $\hat{\theta}_u$ heißt *untere Konfidenzgrenze* für den Parameter θ zum Konfidenzniveau $1 - \beta$, wenn
$$P_\theta(\hat{\theta}_u \leq \theta) \geq 1 - \beta.$$
- $\hat{\theta}_o$ heißt *obere Konfidenzgrenze* für den Parameter θ zum Konfidenzniveau $1-\beta$, wenn
$$P_\theta(\theta \leq \hat{\theta}_o) \geq 1 - \beta.$$

Konfidenzintervalle

Häufig ist man in der Situation, die *unteren Konfidenz- oder Vertrauensgrenze* und die *obere Konfidenz- oder Vertrauensgrenze* für einen Parameter θ einzeln zu bestimmmen:

- $\hat{\theta}_u$ heißt *untere Konfidenzgrenze* für den Parameter θ zum Konfidenzniveau $1-\beta$, wenn
$$P_\theta(\hat{\theta}_u \leq \theta) \geq 1-\beta.$$
- $\hat{\theta}_o$ heißt *obere Konfidenzgrenze* für den Parameter θ zum Konfidenzniveau $1-\beta$, wenn
$$P_\theta(\theta \leq \hat{\theta}_o) \geq 1-\beta.$$

Bemerkung: Aus $P_\theta(\hat{\theta}_u \leq \theta) \geq 1-\frac{\alpha}{2}$ und $P_\theta(\theta \leq \hat{\theta}_o) \geq 1-\frac{\alpha}{2}$ folgt

$$P_\theta(\hat{\theta}_u \leq \theta \leq \hat{\theta}_o) \geq 1-\alpha.$$

Die Notation P_θ bedeutet, dass wir die Wahrscheinlichkeit unter der Annahme, dass der Parameter θ ist, berechnen.