

# Kapitel III Induktive Statistik

## 1. Einführung

Das Ziel der **induktiven Statistik** besteht darin, aus gemessenen Zufallsgrößen auf die zugrunde liegenden Gesetzmäßigkeiten zu schließen. Im Gegensatz dazu spricht man von **deskriptiver Statistik**, wenn man sich damit beschäftigt, große Datenmengen verständlich aufzubereiten, beispielsweise durch Berechnung des Mittelwertes oder anderer abgeleiteter Größen.

## 2. Schätzvariablen

Wir betrachten die Anzahl  $X$  von Lesezugriffen auf eine Festplatte bis zum ersten Lesefehler und nehmen an, dass  $\Pr[X = i] = (1 - p)^{i-1}p$ , setzen also für  $X$  eine geometrische Verteilung an. Dahinter verbirgt sich die Annahme, dass bei jedem Zugriff **unabhängig** und mit jeweils **derselben** Wahrscheinlichkeit  $p$  ein Lesefehler auftreten kann.

Unter diesen Annahmen ist die Verteilung der Zufallsvariablen  $X$  eindeutig festgelegt. Allerdings entzieht sich der numerische Wert des Parameters  $p$  noch unserer Kenntnis. Dieser soll daher nun empirisch geschätzt werden. Statt  $p$  können wir ebensogut  $\mathbb{E}[X]$  bestimmen, da wir daraus nach den Eigenschaften der geometrischen Verteilung  $p$  mittels  $p = \frac{1}{\mathbb{E}[X]}$  berechnen können.

Dazu betrachten wir  $n$  baugleiche Platten und die zugehörigen Zufallsvariablen  $X_i$  (für  $1 \leq i \leq n$ ), d. h. wir zählen für jede Platte die Anzahl von Zugriffen bis zum ersten Lesefehler. Die Zufallsvariablen  $X_i$  sind dann unabhängig und besitzen jeweils dieselbe Verteilung wie  $X$ . Wir führen also viele Kopien eines bestimmten Zufallsexperiments aus, um Schlüsse auf die Gesetzmäßigkeiten des einzelnen Experiments ziehen zu können. Dies ist das Grundprinzip der induktiven Statistik. Die  $n$  Messungen heißen **Stichproben**, und die Variablen  $X_i$  nennt man **Stichprobenvariablen**.

## Grundprinzip statistischer Verfahren

Wir erinnern an das Gesetz der großen Zahlen (Satz 63) bzw. den Zentralen Grenzwertsatz (Satz 108). Wenn man ein Experiment genügend oft wiederholt, so nähert sich der Durchschnitt der Versuchsergebnisse immer mehr dem Verhalten an, das man „im Mittel“ erwarten würde. Je mehr Experimente wir also durchführen, umso genauere und zuverlässigere Aussagen können wir über den zugrunde liegenden Wahrscheinlichkeitsraum ableiten. Auf diesem Grundprinzip beruhen alle statistischen Verfahren.

Um  $\mathbb{E}[X]$  empirisch zu ermitteln, bietet es sich an, aus den Zufallsvariablen  $X_i$  das arithmetische Mittel  $\bar{X}$  zu bilden, das definiert ist durch

$$\bar{X} := \frac{1}{n} \sum_{i=1}^n X_i.$$

Es gilt

$$\mathbb{E}[\bar{X}] = \frac{1}{n} \sum_{i=1}^n \mathbb{E}[X_i] = \frac{1}{n} \sum_{i=1}^n \mathbb{E}[X] = \mathbb{E}[X].$$

$\bar{X}$  liefert uns also im Mittel den gesuchten Wert  $\mathbb{E}[X]$ . Da wir  $\bar{X}$  zur Bestimmung von  $\mathbb{E}[X]$  verwenden, nennen wir  $\bar{X}$  einen **Schätzer** für den Erwartungswert  $\mathbb{E}[X]$ . Wegen der obigen Eigenschaft ist  $\bar{X}$  sogar ein so genannter **erwartungstreuer** Schätzer.

## Definition 112

Gegeben sei eine Zufallsvariable  $X$  mit der Dichte  $f(x; \theta)$ . Eine **Schätzvariable** oder kurz **Schätzer** für den Parameter  $\theta$  der Dichte von  $X$  ist eine Zufallsvariable, die aus mehreren (meist unabhängigen und identisch verteilten) Stichprobenvariablen zusammengesetzt ist. Ein Schätzer  $U$  heißt **erwartungstreu**, wenn gilt

$$\mathbb{E}[U] = \theta.$$

### **Bemerkung:**

Die Größe  $\mathbb{E}[U - \theta]$  nennt man **Bias** der Schätzvariablen  $U$ . Bei erwartungstreuen Schätzvariablen ist der Bias gleich Null.

Der Schätzer  $\bar{X}$  ist also ein erwartungstreuer Schätzer für den Erwartungswert von  $X$ . Ein wichtiges Maß für die Güte eines Schätzers ist die mittlere quadratische Abweichung, kurz **MSE** für **mean squared error** genannt. Diese berechnet sich durch  $MSE := \mathbb{E}[(U - \theta)^2]$ . Wenn  $U$  erwartungstreu ist, so folgt  $MSE = \mathbb{E}[(U - \mathbb{E}[U])^2] = \text{Var}[U]$ .

### Definition 113

Wenn die Schätzvariable  $A$  eine kleinere mittlere quadratische Abweichung besitzt als die Schätzvariable  $B$ , so sagt man, dass  $A$  **effizienter** ist als  $B$ .

Eine Schätzvariable heißt **konsistent im quadratischen Mittel**, wenn  $MSE \rightarrow 0$  für  $n \rightarrow \infty$  gilt. Hierbei bezeichne  $n$  den Umfang der Stichprobe.

Für  $\bar{X}$  erhalten wir wegen der Unabhängigkeit von  $X_1, \dots, X_n$

$$\begin{aligned}MSE = \text{Var}[\bar{X}] &= \text{Var}\left[\frac{1}{n} \sum_{i=1}^n X_i\right] \\ &= \frac{1}{n^2} \sum_{i=1}^n \text{Var}[X_i] = \frac{1}{n} \text{Var}[X].\end{aligned}$$

Bei jeder Verteilung mit endlicher Varianz folgt  $MSE = \mathcal{O}(1/n)$  und somit  $MSE \rightarrow 0$  für  $n \rightarrow \infty$ . Der Schätzer  $\bar{X}$  ist also konsistent.

Aus der Konsistenz von  $\bar{X}$  im quadratischen Mittel können wir mit Hilfe des Satzes von Chebyshev (siehe Satz 61) folgende Konsequenz ableiten. Sei  $\varepsilon > 0$  beliebig, aber fest. Dann gilt

$$\Pr[|\bar{X} - \theta| \geq \varepsilon] = \Pr[|\bar{X} - \mathbb{E}[X]| \geq \varepsilon] \leq \frac{\text{Var}[\bar{X}]}{\varepsilon^2} \rightarrow 0$$

für  $n \rightarrow \infty$ . Für genügend große  $n$  liegen also die Werte von  $\bar{X}$  beliebig nahe am gesuchten Wert  $\theta = \mathbb{E}[X]$ . Diese Eigenschaft nennt man auch **schwache Konsistenz**, da sie aus der Konsistenz im quadratischen Mittel folgt.

Als nächstes betrachten wir eine weitere von  $\bar{X}$  abgeleitete Schätzvariable:

$$S := \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2}.$$

Wir zeigen, dass  $S^2$  ein erwartungstreuer Schätzer für die Varianz von  $X$  ist. Sei  $\mu := \mathbb{E}[X] = \mathbb{E}[X_i] = \mathbb{E}[\bar{X}]$ .

$$\begin{aligned}(X_i - \bar{X})^2 &= (X_i - \mu + \mu - \bar{X})^2 \\ &= (X_i - \mu)^2 + (\mu - \bar{X})^2 + 2(X_i - \mu)(\mu - \bar{X}) \\ &= (X_i - \mu)^2 + (\mu - \bar{X})^2 - \frac{2}{n} \sum_{j=1}^n (X_i - \mu)(X_j - \mu) \\ &= \frac{n-2}{n} (X_i - \mu)^2 + (\mu - \bar{X})^2 - \frac{2}{n} \sum_{j \neq i} (X_i - \mu)(X_j - \mu).\end{aligned}$$

Für je zwei unabhängige Zufallsvariablen  $X_i, X_j$  mit  $i \neq j$  gilt

$$\begin{aligned}\mathbb{E}[(X_i - \mu)(X_j - \mu)] &= \mathbb{E}[X_i - \mu] \cdot \mathbb{E}[X_j - \mu] \\ &= (\mathbb{E}[X_i] - \mu) \cdot (\mathbb{E}[X_j] - \mu) = 0 \cdot 0 = 0.\end{aligned}$$

Daraus folgt

$$\begin{aligned}\mathbb{E}[(X_i - \bar{X})^2] &= \frac{n-2}{n} \cdot \mathbb{E}[(X_i - \mu)^2] + \mathbb{E}[(\mu - \bar{X})^2] \\ &= \frac{n-2}{n} \cdot \text{Var}[X_i] + \text{Var}[\bar{X}].\end{aligned}$$

Wegen  $\text{Var}[X_i] = \text{Var}[X]$  und  $\text{Var}[\bar{X}] = \frac{1}{n} \text{Var}[X]$  folgt nun

$$\mathbb{E}[(X_i - \bar{X})^2] = \frac{n-1}{n} \cdot \text{Var}[X],$$

und somit gilt für  $S^2$

$$\begin{aligned} \mathbb{E}[S^2] &= \frac{1}{n-1} \sum_{i=1}^n \mathbb{E}[(X_i - \bar{X})^2] \\ &= \frac{1}{n-1} \cdot n \cdot \frac{n-1}{n} \cdot \text{Var}[X] = \text{Var}[X]. \end{aligned}$$

$S^2$  ist also eine erwartungstreue Schätzvariable für die Varianz von  $X$ .

Die vorangegangene Rechnung erklärt, warum man als Schätzer nicht

$$\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \neq S^2$$

verwendet, wie man vielleicht intuitiv erwarten würde.

## Definition 114

Die Zufallsvariablen

$$\bar{X} := \frac{1}{n} \sum_{i=1}^n X_i \text{ und } S^2 := \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

heißen **Stichprobenmittel** bzw. **Stichprobenvarianz** der Stichprobe  $X_1, \dots, X_n$ .  $\bar{X}$  und  $S^2$  sind erwartungstreue Schätzer für den Erwartungswert bzw. die Varianz.

## 2.1 Maximum-Likelihood-Prinzip zur Konstruktion von Schätzvariablen

Wir betrachten nun ein Verfahren zur Konstruktion von Schätzvariablen für Parameter von Verteilungen. Sei

$$\vec{X} = (X_1, \dots, X_n).$$

Bei  $X_1, \dots, X_n$  handelt es sich um unabhängige Kopien der Zufallsvariablen  $X$  mit der Dichte  $f(x; \theta)$ . Hierbei sei  $\theta$  der gesuchte Parameter der Verteilung. Wir setzen

$$f(x; \theta) = \Pr[X = x],$$

wobei  $\theta$  ein Parameter der Verteilung ist.

Wenn wir den Parameter explizit angeben wollen, so schreiben wir dafür auch  $f(x; \theta) = \Pr_{\theta}[X = x]$ . Eine Stichprobe liefert für jede Variable  $X_i$  einen Wert  $x_i$ . Diese Werte fassen wir ebenfalls zu einem Vektor  $\vec{x} = (x_1, \dots, x_n)$  zusammen.

Der Ausdruck

$$L(\vec{x}; \theta) := \prod_{i=1}^n f(x_i; \theta) = \prod_{i=1}^n \Pr_{\theta}[X_i = x_i]$$
$$\stackrel{\text{unabh.}}{=} \Pr_{\theta}[X_1 = x_1, \dots, X_n = x_n]$$

entspricht der Wahrscheinlichkeit, dass wir die Stichprobe  $\vec{x}$  erhalten, wenn wir den Parameter mit dem Wert  $\theta$  belegen.

Wir betrachten nun eine feste Stichprobe  $\vec{x}$  und fassen  $L(\vec{x}; \theta)$  somit als Funktion von  $\theta$  auf. In diesem Fall nennen wir  $L$  die **Likelihood-Funktion** der Stichprobe.

Es erscheint sinnvoll, zu einer gegebenen Stichprobe  $\vec{x}$  den Parameter  $\theta$  so zu wählen, dass  $L(x; \theta)$  **maximal** wird.

### Definition 115

Ein Schätzwert  $\hat{\theta}$  für den Parameter einer Verteilung  $f(x; \theta)$  heißt

**Maximum-Likelihood-Schätzwert** (ML-Schätzwert) für eine Stichprobe  $\vec{x}$ , wenn gilt

$$L(\vec{x}; \theta) \leq L(\vec{x}; \hat{\theta}) \text{ für alle } \theta.$$

## Beispiel 116

Wir konstruieren mit der ML-Methode einen Schätzer für den Parameter  $p$  der Bernoulli-Verteilung. Es gilt  $\Pr_p[X_i = 1] = p$  und  $\Pr_p[X_i = 0] = 1 - p$ . Daraus schließen wir, dass  $\Pr_p[X_i = x_i] = p^{x_i}(1 - p)^{1-x_i}$ , und stellen die Likelihood-Funktion

$$L(\vec{x}; p) = \prod_{i=1}^n p^{x_i} \cdot (1 - p)^{1-x_i}$$

auf.

Wir suchen als Schätzer für  $p$  den Wert, an dem die Funktion  $L$  maximal wird. Wir erhalten

$$\begin{aligned} \ln L(\vec{x}; p) &= \sum_{i=1}^n (x_i \cdot \ln p + (1 - x_i) \cdot \ln(1 - p)) \\ &= n\bar{x} \cdot \ln p + (n - n\bar{x}) \cdot \ln(1 - p). \end{aligned}$$

Hierbei bezeichnet  $\bar{x}$  das arithmetische Mittel  $\frac{1}{n} \sum_{i=1}^n x_i$ .

## Beispiel (Forts.)

*Wir finden das Maximum durch Nullsetzen der Ableitung:*

$$\frac{d \ln L(\vec{x}; p)}{dp} = \frac{n\bar{x}}{p} - \frac{n - n\bar{x}}{1 - p} = 0.$$

*Diese Gleichung hat die Lösung  $p = \bar{x}$ .*

## Beispiel 117

Die Zufallsvariable  $X$  sei  $\mathcal{N}(\mu, \sigma^2)$ -verteilt, und wir suchen Schätzvariablen für die Parameter  $\mu$  und  $\sigma$ . Nach Definition der Likelihood-Funktion gilt

$$L(\vec{x}; \mu, \sigma^2) = \left( \frac{1}{\sqrt{2\pi}\sigma} \right)^n \cdot \prod_{i=1}^n \exp \left( -\frac{(x_i - \mu)^2}{2\sigma^2} \right).$$

Durch Logarithmieren erhalten wir

$$\ln L(\vec{x}; \mu, \sigma^2) = -n(\ln \sqrt{2\pi} + \ln \sigma) + \sum_{i=1}^n \left( -\frac{(x_i - \mu)^2}{2\sigma^2} \right).$$

## Beispiel 117

Für die Nullstellen der Ableitungen ergibt sich

$$\frac{\partial \ln L}{\partial \mu} = \sum_{i=1}^n \frac{x_i - \mu}{\sigma^2} \stackrel{!}{=} 0,$$

$$\frac{\partial \ln L}{\partial \sigma} = -\frac{n}{\sigma} + \sum_{i=1}^n \frac{(x_i - \mu)^2}{\sigma^3} \stackrel{!}{=} 0,$$

also

$$\mu = \bar{x} \quad \text{und} \quad \sigma^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2.$$

Wir haben also durch die ML-Methode „fast“ das Stichprobenmittel und die Stichprobenvarianz erhalten. Allerdings besitzt der Schätzer für die Varianz hier den Vorfaktor  $\frac{1}{n}$  statt  $\frac{1}{n-1}$ . Die ML-Schätzvariable für die Varianz ist somit nicht erwartungstreu.

### 3. Konfidenzintervalle

Bei der Verwendung von Schätzvariablen geht man davon aus, dass der erhaltene Schätzwert „nahe“ beim gesuchten Parameter  $\theta$  liegt. Die Schätzungen werden „besser“, je größer die betrachtete Stichprobe ist. Diese Angaben sind aus quantitativer Sicht natürlich unbefriedigend, da nicht erkennbar ist, wie gut man sich auf den Schätzwert verlassen kann.

Die Lösung dieses Problems besteht darin, statt einer Schätzvariablen  $U$  zwei Schätzer  $U_1$  und  $U_2$  zu betrachten.  $U_1$  und  $U_2$  werden so gewählt, dass

$$\Pr[U_1 \leq \theta \leq U_2] \geq 1 - \alpha.$$

Die Wahrscheinlichkeit  $1 - \alpha$  heißt **Konfidenzniveau** und kann dem „Sicherheitsbedürfnis“ angepasst werden.

Wenn wir für eine konkrete Stichprobe die Schätzer  $U_1$  und  $U_2$  berechnen und davon ausgehen, dass  $\theta \in [U_1, U_2]$  ist, so ziehen wir höchstens mit Wahrscheinlichkeit  $\alpha$  einen falschen Schluss.  $[U_1, U_2]$  heißt **Konfidenzintervall**.

In vielen Fällen verwendet man nur eine Schätzvariable  $U$  und konstruiert mittels  $U_1 := U - \delta$  und  $U_2 := U + \delta$  ein symmetrisches Konfidenzintervall  $[U - \delta, U + \delta]$ .

Sei  $X$  eine  $\mathcal{N}(\mu, \sigma^2)$ -verteilte Zufallsvariable, und seien  $X_1, \dots, X_n$   $n$  zugehörige Stichprobenvariablen. Gemäß der Additivität der Normalverteilung (siehe Satz 106) ist das Stichprobenmittel  $\bar{X}$  ebenfalls normalverteilt mit  $\bar{X} \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$ . Wir suchen für  $\bar{X}$  ein symmetrisches Konfidenzintervall.

Nach Satz 93 ist

$$Z := \sqrt{n} \cdot \frac{\bar{X} - \mu}{\sigma}$$

standardnormalverteilt.

Für  $Z$  betrachten wir das Konfidenzintervall  $[-c, c]$  für ein geeignetes  $c > 0$  und setzen

$$\Pr[-c \leq Z \leq c] \stackrel{!}{=} 1 - \alpha.$$

Auflösen nach  $\mu$  ergibt

$$\Pr \left[ \bar{X} - \frac{c\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + \frac{c\sigma}{\sqrt{n}} \right] \stackrel{!}{=} 1 - \alpha.$$

Das gesuchte Konfidenzintervall lautet also

$$K = \left[ \bar{X} - \frac{c\sigma}{\sqrt{n}}, \bar{X} + \frac{c\sigma}{\sqrt{n}} \right].$$

Den Parameter  $c$  wählen wir wie folgt:

$$\Pr[-c \leq Z \leq c] = \Phi(c) - \Phi(-c) \stackrel{!}{=} 1 - \alpha.$$

Wegen der Symmetrie von  $\Phi$  gilt  $\Phi(-x) = 1 - \Phi(x)$  und wir erhalten

$$\Phi(c) - \Phi(-c) = 2 \cdot \Phi(c) - 1 \stackrel{!}{=} 1 - \alpha \iff \Phi(c) = 1 - \frac{\alpha}{2},$$

also

$$c = \Phi^{-1} \left( 1 - \frac{\alpha}{2} \right).$$

### Definition 118

$X$  sei eine stetige Zufallsvariable mit Verteilung  $F_X$ . Eine Zahl  $x_\gamma$  mit

$$F_X(x_\gamma) = \gamma$$

heißt  $\gamma$ -Quantil von  $X$  bzw. der Verteilung  $F_X$ .

### Definition 119

Für die Standardnormalverteilung bezeichnet  $z_\gamma$  das  $\gamma$ -Quantil.

Damit können wir das gesuchte Konfidenzintervall angeben durch

$$K = \left[ \bar{X} - \frac{z_{(1-\frac{\alpha}{2})}\sigma}{\sqrt{n}}, \bar{X} + \frac{z_{(1-\frac{\alpha}{2})}\sigma}{\sqrt{n}} \right].$$

## 4. Testen von Hypothesen

### 4.1 Einführung

Bislang haben wir versucht, Parameter von Verteilungen zu schätzen. In der Praxis ist man jedoch oft an der eigentlichen Kenntnis dieser Parameter gar nicht interessiert, sondern man möchte gewisse, damit zusammenhängende Behauptungen überprüfen.

Im Folgenden stellen wir die Bestandteile eines statistischen Tests anhand eines abstrakten Beispiels vor. Wir betrachten dazu eine Zufallsvariable  $X$  mit  $\Pr[X = 1] = p$  und  $\Pr[X = 0] = 1 - p$ . Durch einen Test soll überprüft werden, ob  $p < 1/3$  oder  $p \geq 1/3$  gilt.

## Definition eines Tests

Wir betrachten eine Stichprobe von  $n$  unabhängigen Stichprobenvariablen  $X_1, \dots, X_n$ , die dieselbe Verteilung wie die Zufallsvariable  $X$  besitzen. Zu einem zugehörigen Stichprobenvektor  $\vec{x}$  müssen wir nun die Frage beantworten, ob wir für diesen Versuchsausgang die Hypothese „ $p \geq 1/3$ “ annehmen oder ablehnen.

Sei

$$K := \{\vec{x} \in \mathbb{R}^n; \vec{x} \text{ führt zur Ablehnung der Hypothese}\}.$$

$K$  nennen wir den **Ablehnungsbereich** oder den **kritischen Bereich** des Tests.

Gewöhnlich wird  $K$  konstruiert, indem man die Zufallsvariablen  $X_1, \dots, X_n$  zu einer neuen Variablen  $T$ , der so genannten **Testgröße**, zusammenfasst. Dann unterteilt man den Wertebereich  $\mathbb{R}$  von  $T$  in mehrere Bereiche, die entweder zur Ablehnung der Hypothese führen sollen oder nicht. Dabei betrachtet man meist ein einzelnes halboffenes oder abgeschlossenes Intervall und spricht dann von einem **einseitigen** bzw. von einem **zweiseitigen** Test.

Die Menge  $\tilde{K} \subseteq \mathbb{R}$  enthalte die Werte von  $T$ , die zur Ablehnung der Hypothese führen sollen. Da wir Tests immer über eine Testgröße definieren, werden wir der Einfachheit halber auch  $\tilde{K}$  als Ablehnungsbereich bezeichnen.  $\tilde{K} \subseteq \mathbb{R}$  entspricht direkt dem Ablehnungsbereich  $K = T^{-1}(\tilde{K}) \subseteq \mathbb{R}^n$ , wie wir ihn oben festgelegt haben.