Kapitel 9

Schätzverfahren und Konfidenzintervalle

9.1 Grundlagen zu Schätzverfahren

Für eine Messreihe $x_1, \ldots, x_n$ wird im Folgenden angenommen, dass sie durch n gleiche Zufallsexperimente unabhängig voneinander ermittelt werden. Jeden Messwert sehen wir als unabhängige Realisierung einer Zufallsvariable X an. Als mathematisches Modell für das Entstehen von Messreihen werden im folgenden unabhängige, identisch wie X verteilte Zufallsvariablen $X_1, \ldots, X_n$ verwendet. Eine Messreihe $x_1, \ldots, x_n$ wird als Realisierung der Zufallsvariablen $X_1, \ldots, X_n$ angesehen, wir nehmen also an, dass ein Ergebnis $\omega \in \Omega$ existiert mit

$$x_1 = X_1(\omega), \ldots, x_n = X_n(\omega).$$

Es wird nun angenommen, dass die Verteilungsfunktion F von X, die auch die Verteilungsfunktion der unabhängigen, identisch verteilten Zufallsvariablen X_i, $1 \leq i \leq n$, ist, einer durch einen Parameter $\theta \in \Theta \subset \mathbb{R}^k$ parametrisierten Familie

$$F_\theta, \quad \theta \in \Theta,$$

von Verteilungsfunktionen angehört. Dieser Parameter oder ein durch ihn bestimmter Zahlenwert $\tau(\theta)$ mit einer Abbildung $\tau : \Theta \to \mathbb{R}$ sei unbekannt und soll aufgrund der Messreihe näherungsweise geschätzt werden.

Beispiel: X und alle $X_1, \ldots, X_n$ seien normalverteilt. F_θ mit

$$\theta = (\mu, \sigma^2) \in \Theta = \mathbb{R}\times]0, \infty[$$

ist dann die Verteilungsfunktion einer $N(\mu, \sigma^2)$-Verteilung. Soll der Erwartungswert geschätzt werden, so ist $\tau(\theta) = \mu$. Will man die Varianz schätzen, dann ist $\tau(\theta) = \sigma^2$.

Definition 9.1.1 *Ein* Schätzverfahren *oder eine* Schätzfunktion *oder kurz ein* Schätzer *ist eine Abbildung*

$$T_n: \ \mathbb{R}^n \to \mathbb{R}.$$

Sie ordnet einer Messreihe $x_1, \ldots, x_n$ einen Schätzwert $T_n(x_1, \ldots, x_n)$ *für den unbekannten Wert $\tau(\theta)$ zu.*

Die Zufallsvariable $T_n(X_1, \ldots, X_n)$ heißt Schätzvariable.

Erwartungswert und Varianz der Schätzvariablen $T_n(X_1, \ldots, X_n)$ sowie aller X_i hängen von der Verteilungsfunktion F_θ ab, die seiner Berechnung zugrundegelegt wird. Um dies zu verdeutlichen, schreiben wir

$$E_\theta(T_n(X_1, \ldots, X_n)), \quad E_\theta(X_1), \ldots$$

sowie

$$\mathrm{Var}_\theta(T_n(X_1, \ldots, X_n)), \quad \mathrm{Var}_\theta(X_1), \ldots$$

Außerdem schreiben wir für durch F_θ berechnete Wahrscheinlichkeiten

$$P_\theta(a \le T_n(X_1, \ldots, X_n) \le b), \quad P_\theta(a \le X_1 \le b), \ldots$$

Definition 9.1.2 *Ein Schätzer $T_n: \ \mathbb{R}^n \to \mathbb{R}$ heißt* erwartungstreu *für $\tau: \Theta \to \mathbb{R}$, falls gilt*

$$E_\theta(T_n(X_1, \ldots, X_n)) = \tau(\theta) \quad \textit{für alle } \theta \in \Theta.$$

Beispiele:

1. τ sei gegeben durch $\tau(\theta) = E_\theta(X) = \mu$. Das arithmetische Mittel $\bar{X}_{(n)} = \frac{1}{n}(X_1 + \ldots + X_n)$ ist ein erwartungstreuer Schätzer für $\tau(\theta)$. Tatsächlich gilt

$$E_\theta(\bar{X}_{(n)}) = E_\theta(\frac{1}{n}(X_1 + \ldots + X_n)) = \frac{1}{n}(E_\theta(X_1) + \ldots + E_\theta(X_n)) = \frac{1}{n} n\mu = \mu.$$

2. τ sei gegeben durch $\tau(\theta) = \mathrm{Var}_\theta(X)$. Die Stichprobenvarianz $S^2_{(n)} := \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X}_{(n)})^2$ ist ein erwartungstreuer Schätzer für $\tau(\theta)$. Denn es gilt

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

Nun gilt $E_\theta((\bar{X}_{(n)}-\mu)^2) = \text{Var}_\theta(\bar{X}_{(n)}) = \frac{1}{n^2} n \text{Var}_\theta(X)$, also

$$E_\theta\left(\sum_{i=1}^n (X_i - \bar{X}_{(n)})^2\right) = \sum_{i=1}^n E_\theta((X_i-\mu)^2) - nE_\theta((\bar{X}_{(n)}-\mu)^2)$$
$$= n\text{Var}_\theta(X) - n\frac{1}{n}\text{Var}_\theta(X) = (n-1)\text{Var}_\theta(X).$$

Als Abschwächung der Erwartungstreue betrachtet man asymptotische Erwartungstreue bei wachsender Stichprobenlänge.

Definition 9.1.3 *Ein Folge von Schätzer $T_n: \ \mathbb{R}^n \to \mathbb{R}$, $n = 1, 2, \ldots$ heißt* asymptotisch erwartungstreu *für $\tau: \Theta \to \mathbb{R}$, falls gilt*

$$\lim_{n\to\infty} E_\theta(T_n(X_1, \ldots, X_n)) = \tau(\theta) \quad \textit{für alle } \theta \in \Theta.$$

Zur Beurteilung der Güte eines Schätzers dient der

Mittlere quadratische Fehler (mean squared error):

$$MSE_\theta(T) := E_\theta((T - \tau(\theta))^2).$$

Offensichtlich gilt

$$T \text{ erwartungstreu} \quad \Longrightarrow \quad MSE_\theta(T) = \text{Var}_\theta(T).$$

Sind T_1 und T_2 zwei Schätzer für τ, dann heißt T_1 *effizienter* als T_2, wenn gilt

$$MSE_\theta(T_1) \leq MSE_\theta(T_2) \quad \forall\, \theta \in \Theta.$$

Sind T_1, T_2 erwartungstreu, dann bedeutet dies

$$\text{Var}_\theta(T_1) \leq \text{Var}_\theta(T_2) \quad \forall\, \theta \in \Theta.$$

Definition 9.1.4 *Ein Folge von Schätzern $T_1, T_2, \ldots$ heißt* konsistent *für τ, wenn für alle $\varepsilon > 0$ und alle $\theta \in \Theta$ gilt*

$$\lim_{n\to\infty} P_\theta\left(|T_n(X_1, \ldots, X_n) - \tau(\theta)| > \varepsilon\right) = 0.$$

Sie heißt konsistent im quadratischen Mittel *für τ, wenn für alle $\theta \in \Theta$ gilt*

$$\lim_{n\to\infty} MSE_\theta(T_n) = 0.$$

Es gilt folgender

Satz 9.1.5 *Ist $T_1, T_2, \ldots$ eine Folge von Schätzern, die erwartungstreu für τ sind und gilt*

$$\lim_{n\to\infty} \mathit{Var}_\theta(T_n(X_1, \ldots, X_n)) = 0 \quad \textit{für alle } \theta \in \Theta,$$

dann ist die Folge von Schätzern konsistent für τ.

Beweis: Wegen $E_\theta(T_n(X_1, \ldots, X_n)) = \tau(\theta)$ gilt nach der Ungleichung von Tschebyschev

$$P_\theta(|T_n(X_1, \ldots, X_n) - \tau(\theta)| > \varepsilon) \leq \frac{\mathrm{Var}_\theta(T_n(X_1, \ldots, X_n))}{\varepsilon^2} \to 0.$$

□

Allgemeiner haben wir mit ganz ähnlichem Beweis

Satz 9.1.6 *Ist $T_1, T_2, \ldots$ eine Folge von Schätzern, die konsistent im quadratischen Mittel für τ ist, dann ist die Folge von Schätzern konsistent für τ.*

Beispiel: Es sei X $N(\mu, \sigma^2)$-verteilt, $\theta = (\mu, \sigma^2) \in \Theta = \mathbb{R} \times]0, \infty[$ und $\tau(\theta) = \mu$. Der Schätzer

$$T_n(X_1, \ldots, X_n) = \bar{X}_{(n)} = \frac{1}{n}(X_1 + \ldots + X_n)$$

ist nach Satz 8.8.1 $N(\mu, \sigma^2/n)$-verteilt, also erwartungstreu mit Varianz

$$\mathrm{Var}_\theta(T_n(X_1, \ldots, X_n)) = \sigma^2/n \to 0 \quad \text{für } n \to \infty.$$

Daher ist die Schätzerfolge nach Satz 9.1.5 auch konsistent.

9.2 Maximum-Likelihood-Schätzer

Bei gegebener Verteilungsklasse F_θ, $\theta \in \Theta$, lassen sich Schätzer für den Parameter θ oft mit der Maximum-Likelihood-Methode gewinnen.

Sind die zugrundeliegenden Zufallsvariablen $X_1, \ldots, X_n$ stetig verteilt, so ist die Verteilungsfunktion F_θ durch eine Dichte

$$f_\theta(x), \quad x \in \mathbb{R},$$

bestimmt. Im Fall diskreter Zufallsvariablen X, bzw. $X_1, \ldots, X_n$ definieren wir

$$f_\theta(x) = P_\theta(X = x) \quad \text{für alle } x \text{ aus dem Wertevorrat } \mathbb{X} \text{ von } X.$$

Definition 9.2.1 *Für eine Messreihe* $x_1, \ldots, x_n$ *heißt die Funktion* $L(\cdot; x_1, \ldots, x_n)$ *mit*

$$L(\theta; x_1, \ldots, x_n) = f_\theta(x_1) \cdot f_\theta(x_2) \cdot \ldots \cdot f_\theta(x_n)$$

die zu $x_1, \ldots, x_n$ *gehörige* Likelihood-Funktion.

Ein Parameterwert

$$\hat{\theta} = \hat{\theta}(x_1, \ldots, x_n)$$

mit

$$L(\hat{\theta}; x_1, \ldots, x_n) \geq L(\theta; x_1, \ldots, x_n) \quad \text{für alle } \theta \in \Theta$$

heißt Maximum-Likelihood-Schätzwert *für* θ*. Existiert zu jeder möglichen Messreihe* $x_1, \ldots, x_n$ *ein Maximum-Likelihood-Schätzwert* $\hat{\theta}(x_1, \ldots, x_n)$*, dann heißt*

$$T_n : \mathbb{X}^n \to \mathbb{R}, \quad T_n(x_1, \ldots, x_n) = \hat{\theta}(x_1, \ldots, x_n)$$

Maximum-Likelihood-Schätzer.

Beispiel: Die Zufallsvariablen seien Poisson-verteilt mit Parameter $\theta > 0$, also

$$f_\theta(x) = \frac{\theta^x}{x!} e^{-\theta}, \quad x \in \mathbb{N} \cup \{0\}.$$

Dies ergibt

$$L(\theta; x_1, \ldots, x_n) = \frac{1}{x_1! \cdots x_n!} \cdot \theta^{x_1 + \ldots + x_n} \cdot e^{-n\theta}, \quad x_i \in \mathbb{N} \cup \{0\}.$$

L wird genau dann maximal, wenn die Log-Likelihood-Funktion $\ln(L)$, also

$$\ln L(\theta; x_1, \ldots, x_n) = -n\theta - \ln(x_1! \cdots x_n!) + (x_1 + \ldots + x_n) \ln \theta,$$

maximal wird. Die erste Ableitung dieser Funktion nach θ ist

$$\frac{d \ln L}{d\theta} = -n + \frac{x_1 + \ldots + x_n}{\theta}$$

mit der eindeutigen Nullstelle

$$\hat{\theta}(x_1, \ldots, x_n) = \frac{1}{n}(x_1 + \ldots + x_n).$$

Da die zweite Ableitung negativ ist, ist $\hat{\theta}$ der Maximum-Likelihood-Schätzer für θ und ist nichts anderes als das arithmetische Mittel.

9.3 Konfidenzintervalle

Die Situation sei wie beim Schätzen. Es wird eine Messreihe $x_1, \ldots, x_n$ beobachtet und es sollen diesmal Ober- und Unterschranken für den Wert $\tau(\theta)$ aus der Messreihe ermittelt werden. Durch ein Paar

$$U : \mathbb{R}^n \to \mathbb{R}, \quad O : \mathbb{R}^n \to \mathbb{R}$$

von Schätzern mit

$$U(x_1, \ldots, x_n) \leq O(x_1, \ldots, x_n)$$

wird ein ”zufälliges Intervall”

$$I(X_1, \ldots, X_n) = [U(X_1, \ldots, X_n), O(X_1, \ldots, X_n)]$$

definiert.

Definition 9.3.1 *Sei* $0 < \alpha < 1$. *Das zufällige Intervall* $I(X_1, \ldots, X_n)$ *heißt* Konfidenzintervall *für* $\tau(\theta)$ *zum* Konfidenzniveau $1 - \alpha$, *falls gilt*

$$P_\theta(U(X_1, \ldots, X_n) \leq \tau(\theta) \leq O(X_1, \ldots, X_n)) \geq 1 - \alpha \quad \textit{für alle } \theta \in \Theta.$$

Das zu einer bestimmten Messreihe $x_1, \ldots, x_n$ *gehörige Intervall*

$$I(x_1, \ldots, x_n) = [U(x_1, \ldots, x_n), O(x_1, \ldots, x_n)]$$

heißt konkretes Schätzintervall *für* $\tau(\theta)$.

Die Forderung stellt sicher, dass mit Wahrscheinlichkeit $1 - \alpha$ ein konkretes Schätzintervall den Wert $\tau(\theta)$ enthält.

9.3.1 Konstruktion von Konfidenzintervallen

Wir nehmen an, dass $X_1, \ldots, X_n$ unabhängig, identisch normalverteilt sind. Die Verteilungsfunktion F_θ ist dann durch den zweidimensionalen Parameter $\theta = (\mu, \sigma^2)$ bestimmt durch

$$F_\theta(x) = F_{(\mu,\sigma^2)}(x) = \Phi\left(\frac{x - \mu}{\sigma}\right).$$

Mit den bereits eingeführten Bezeichnungen

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

erhält man folgende Konfidenzintervalle zum Niveau $1 - \alpha$:

Konfidenzintervall für μ bei bekannter Varianz $\sigma^2 = \sigma_0^2$:

Hier ist $\Theta = \{(\mu, \sigma_0^2) \,:\, \mu \in \mathbb{R}\}$ und $\tau(\theta) = \mu$. Das Konfidenzintervall für μ lautet

$$I(X_1, \ldots, X_n) = \left[\bar{X}_{(n)} - u_{1-\alpha/2}\frac{\sigma_0}{\sqrt{n}}, \bar{X}_{(n)} + u_{1-\alpha/2}\frac{\sigma_0}{\sqrt{n}}\right].$$

mit dem $(1-\alpha/2)$-Quantil $u_{1-\alpha/2}$ der $N(0,1)$-Verteilung, also

$$\Phi(u_{1-\alpha/2}) = 1 - \alpha/2.$$

Begründung: $\bar{X}_{(n)}$ ist nach Satz 8.8.1 $N(\mu, \sigma_0^2/n)$-verteilt. Also gilt:

$$Y_n := \frac{\bar{X}_{(n)} - \mu}{\sqrt{\sigma_0^2/n}} \quad \text{ist } N(0,1)\text{-verteilt.}$$

Wegen $\Phi(-u_{1-\alpha/2}) = \alpha/2$ gilt

$$P_\theta(-u_{1-\alpha/2} \le Y_n \le u_{1-\alpha/2}) = 1 - \alpha.$$

Einsetzen und Umformen ergibt

$$P_\theta(-u_{1-\alpha/2} \le \frac{\bar{X}_{(n)} - \mu}{\sigma_0/\sqrt{n}} \le u_{1-\alpha/2}) = P_\theta\left(\bar{X}_{(n)} - u_{1-\alpha/2}\frac{\sigma_0}{\sqrt{n}} \le \mu \le \bar{X}_{(n)} + u_{1-\alpha/2}\frac{\sigma_0}{\sqrt{n}}\right) = 1-\alpha.$$

Konfidenzintervall für μ bei unbekannter Varianz σ^2:

Hier ist $\Theta = \{(\mu, \sigma^2) \,:\, \mu \in \mathbb{R}, \sigma^2 > 0\}$ und $\tau(\theta) = \mu$. Das Konfidenzintervall für μ lautet

$$I(X_1, \ldots, X_n) = \left[\bar{X}_{(n)} - t_{n-1;1-\alpha/2}\sqrt{\frac{S_{(n)}^2}{n}}, \bar{X}_{(n)} + t_{n-1;1-\alpha/2}\sqrt{\frac{S_{(n)}^2}{n}}\right]$$

mit dem $(1-\alpha/2)$-Quantil $t_{n-1;1-\alpha/2}$ der t_{n-1}-Verteilung.

Begründung: Nach Satz 8.8.1 ist

$$Y_n := \frac{\bar{X}_{(n)} - \mu}{\sqrt{S_{(n)}^2/n}} \quad \text{ist } t_{n-1}\text{-verteilt.}$$

Eine Rechnung völlig analog wie eben liefert das Konfidenzintervall.

Konfidenzintervall für σ^2 bei bekanntem Erwartungswert $\mu = \mu_0$:

Hier ist $\Theta = \{(\mu_0, \sigma^2) \;:\; \sigma^2 > 0\}$ und $\tau(\theta) = \sigma^2$. Das Konfidenzintervall für σ^2 lautet

$$I(X_1, \dots, X_n) = \left[\frac{\sum_{i=1}^n (X_i - \mu_0)^2}{\chi^2_{n;1-\alpha/2}}, \frac{\sum_{i=1}^n (X_i - \mu_0)^2}{\chi^2_{n;\alpha/2}} \right].$$

Begründung: Jedes $\frac{X_i - \mu_0}{\sigma}$ ist $N(0,1)$-verteilt. Wegen der Unabhängigkeit ist also nach 8.8 $\frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \mu_0)^2 \quad \chi^2_n$-verteilt. Dies ergibt

$$P_\theta \left(\chi^2_{n;\alpha/2} \le \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \mu_0)^2 \le \chi^2_{n;1-\alpha/2} \right) = 1 - \alpha$$

und Auflösen nach σ^2 liefert das Konfidenzintervall.

Konfidenzintervall für σ^2 bei unbekanntem Erwartungswert:

Hier ist $\Theta = \{(\mu, \sigma^2) \;:\; \mu \in \mathbb{R}, \sigma^2 > 0\}$ und $\tau(\theta) = \sigma^2$. Das Konfidenzintervall für σ^2 lautet

$$I(X_1, \dots, X_n) = \left[\frac{(n-1)S^2_{(n)}}{\chi^2_{n-1;1-\alpha/2}}, \frac{(n-1)S^2_{(n)}}{\chi^2_{n-1;\alpha/2}} \right].$$

Begründung: Nach Satz 8.8.1 ist $\frac{n-1}{\sigma^2} S^2_{(n)} \quad \chi^2_{n-1}$-verteilt. Dies ergibt

$$P_\theta \left(\chi^2_{n-1;\alpha/2} \le \frac{n-1}{\sigma^2} S^2_{(n)} \le \chi^2_{n-1;1-\alpha/2} \right) = 1 - \alpha$$

und Auflösen nach σ^2 liefert das Konfidenzintervall.