Vorlesung Statistik II

Florian Wickelmaier

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Sommersemester 2026

Konsistenz

Die Eigenschaft der (asymptotischen) Erwartungstreue kennzeichnet die zentrale Tendenz einer Schätzfunktion

- Der Erwartungswert der Schätzfunktion wird in Beziehung gesetzt zu dem zu schätzenden Parameter

Neben dem im Mittel zu erwartenden Wert spielt aber auch die Streuung einer Schätzfunktion eine entscheidende Rolle

Im Rahmen der Konsistenz interessiert man sich vor allem für folgende Fragen

- Welche Abweichungen der Schätzwerte vom Parameter treten auf?
- Wie verändern sich diese Abweichungen bei steigendem Stichprobenumfang n?

Schwache Konsistenz

Für eine Stichprobe $X_1, \ldots, X_n$ vom Umfang n betrachten wir die absolute Abweichung des Schätzwertes vom Parameter

$$|t_n(X_1, X_2, \ldots, X_n) - \vartheta|$$

Wir interessieren uns dann für die Fälle, in denen der Schätzwert um mehr als einen Betrag $\varepsilon > 0$ vom Parameter ϑ abweicht, also für das Ereignis

$$|t_n(X_1, X_2, \ldots, X_n) - \vartheta| > \varepsilon$$

Schließlich betrachten wir die Wahrscheinlichkeit für das Auftreten dieses Ereignisses

$$P(|t_n(X_1, X_2, \ldots, X_n) - \vartheta| > \varepsilon)$$

Schwache Konsistenz

Eine Folge $T_n = t_n(X_1, X_2, \ldots, X_n)$, $n = 1, 2, \ldots$ von Schätzfunktionen für den Parameter ϑ heißt *(schwach) konsistent*, wenn für jedes $\varepsilon > 0$ gilt

$$\lim_{n \to \infty} P(|t_n(X_1, X_2, \ldots, X_n) - \vartheta| > \varepsilon) = 0$$

Die Bedingung der Konsistenz fordert daher

- Die Wahrscheinlichkeit dafür, dass die Schätzfunktion T_n Werte annimmt, die um mehr als ε vom Parameter ϑ abweichen, wird beliebig klein, wenn nur n hinreichend groß gewählt wird

Mittlere quadratische Abweichung

Die asymptotische Eigenschaft der Konsistenz hängt offensichtlich von den erwarteten Abweichungen der Schätzwerte vom Parameter ab

Entsprechend der Argumentation bei Einführung der Varianz betrachten wir daher die *mittlere quadratische Abweichung* (mean squared error, MSE)

$$MSE(T_n) = E\left[(T_n - \vartheta)^2\right]$$

Es lässt sich zeigen, dass gilt

$$MSE(T_n) = Var(T_n) + Bias(T_n)^2$$

Beweis

Es ist

$$E\left[(T_n - \vartheta)^2\right] = E\left\{\left[(T_n - E(T_n)) + (E(T_n) - \vartheta)\right]^2\right\}$$

Ausmultiplizieren der rechten Seite liefert

$$E[(T_n - E(T_n))^2] + 2\,E[(T_n - E(T_n)) \cdot (E(T_n) - \vartheta)] + E[(E(T_n) - \vartheta)^2]$$

Dabei gilt für den ersten Summanden

$$E[(T_n - E(T_n))^2] = Var(T_n)$$

Beweis

Für den zweiten Summanden

$$E[(T_n - E(T_n)) \cdot (E(T_n) - \vartheta)]$$

erhält man nach Ausmultiplizieren

$$E[T_n \cdot E(T_n) - T_n \cdot \vartheta - E(T_n)^2 + E(T_n) \cdot \vartheta]$$

und daraus

$$E(T_n)^2 - E(T_n) \cdot \vartheta - E(T_n)^2 + E(T_n) \cdot \vartheta = 0$$

Beweis

Für den dritten Summanden gilt offenbar

$$E[(E(T_n) - \vartheta)^2] = (E(T_n) - \vartheta)^2 = Bias(T_n)^2$$

Die Gleichung

$$MSE(T_n) = Var(T_n) + Bias(T_n)^2$$

macht deutlich, dass die Varianz der Schätzfunktion (und damit auch deren Standardfehler) als Kriterium für die Güte der Schätzung nur dann sinnvoll ist, wenn der Bias mit berücksichtigt wird

- Der Standardfehler ist als Vergleichsmaß für die Güte der Schätzung auf erwartungstreue Schätzfunktionen (d. h. Bias gleich Null) beschränkt

Konsistenz im quadratischen Mittel

Asymptotische Erwartungstreue beinhaltet, dass der Bias für wachsenden Stichprobenumfang verschwindet

Geht gleichzeitig auch die Varianz gegen Null, dann spricht man von Konsistenz (im quadratischen Mittel)

Wegen der Zerlegung der mittleren quadratischen Abweichung MSE in einen Varianz- und einen Bias-Anteil verschwinden für wachsenden Stichprobenumfang n Varianz und Bias, wenn MSE gegen Null geht

- Eine Schätzfunktion T_n heißt *konsistent im quadratischen Mittel*, wenn gilt

$$\lim_{n \to \infty} MSE(T_n) = 0$$

Zwei Charakterisierungen der Konsistenz

Konsistenz ist (ebenso wie die asymptotische Erwartungstreue) eine asymptotische Eigenschaft, die das Verhalten für große Stichprobenumfänge wiedergibt

- Eine konsistente Schätzfunktion kann für endlichen Stichprobenumfang n sowohl einen erheblichen Bias als auch eine große Varianz aufweisen

Im Moment haben wir zwei Charakterisierungen der Konsistenz einer Schätzfunktion, deren Zusammenhang alles andere als offensichtlich ist

- Es lässt sich zeigen, dass die Konsistenz im quadratischen Mittel die schwache Konsistenz impliziert
- Diese Beziehung wird durch die *Tschebyschowsche Ungleichung* vermittelt

Tschebyschowsche Ungleichung

Für eine Zufallsvariable X, für die Erwartungswert und Varianz existieren, gelten für beliebige $\varepsilon > 0$ folgende Ungleichungen

$$P(|X - E(X)| \geq \varepsilon) \leq \frac{Var(X)}{\varepsilon^2}$$

und

$$P(|X - E(X)| < \varepsilon) \geq 1 - \frac{Var(X)}{\varepsilon^2}$$

Angewendet auf die Schätzfunktion T_n ergibt sich damit

$$P(|T_n - E(T_n)| \geq \varepsilon) \leq \frac{Var(T_n)}{\varepsilon^2}$$

Konsistenz im quadratischen Mittel impliziert schwache Konsistenz

Ist T_n konsistent im quadratischen Mittel, dann ergibt sich wegen

$$\lim_{n\to\infty} E(T_n) = \vartheta \quad \text{und} \quad \lim_{n\to\infty} Var(T_n) = 0$$

die Abschätzung

$$\begin{aligned}\lim_{n\to\infty} P(|T_n - E(T_n)| \geq \varepsilon) &= \lim_{n\to\infty} P(|T_n - \vartheta| \geq \varepsilon) \\ &\leq \frac{1}{\varepsilon^2} \cdot \lim_{n\to\infty} Var(T_n) \\ &= 0\end{aligned}$$

Beispiel: arithmetisches Mittel

Sei $X_1, \ldots, X_n$ eine einfache Stichprobe aus einer $N(\mu, \sigma^2)$-verteilten Population

Wir haben bereits gezeigt, dass das arithmetische Mittel

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

eine erwartungstreue Schätzfunktion für den Erwartungswert μ ist und für die Varianz gilt $Var(\bar{X}) = \sigma^2/n$

Daher ist

$$\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$$

Beispiel: arithmetisches Mittel

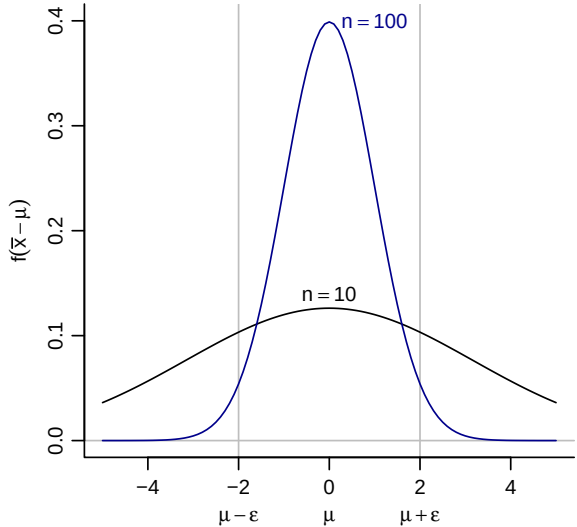
Für $n \to \infty$ verschwindet also die Varianz von $\bar{X}$, was in Verbindung mit der Erwartungstreue bedeutet, dass $\bar{X}$ konsistent im quadratischen Mittel ist

Unter den getroffenen Verteilungsannahmen erhält man darüber hinaus

$$\begin{aligned} P(|\bar{X} - \mu| \leq \varepsilon) &= P\left(\left|\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}\right| \leq \frac{\varepsilon}{\sigma/\sqrt{n}}\right) \\ &= \Phi\left(\frac{\varepsilon}{\sigma}\sqrt{n}\right) - \Phi\left(-\frac{\varepsilon}{\sigma}\sqrt{n}\right) \end{aligned}$$

Beispiel: arithmetisches Mittel

$$\mu = 0$$
$$\sigma^2 = 100$$
$$\varepsilon = 2$$
$$n \in \{10, 100\}$$



Weitere Beispiele

Folgende Schätzfunktionen sind konsistent im quadratischen Mittel und daher auch schwach konsistent

- $\hat{\mu} = \bar{X}$ (auch bei dichotomen Merkmalen mit $X(\Omega) = \{0, 1\}$)
- $\hat{\sigma}^2 = S^2$

Ein psychologisches Beispiel für fehlende Konsistenz bildet die Schätzung des Personenparameters im Rahmen des Rasch-Modells, eines testtheoretischen Ansatzes, der in der Vorlesung Psychometrie ausführlich besprochen wird

Effizienz

Die mittlere quadratische Abweichung ist ein Maß für die Güte der Schätzung, das sowohl den Bias als auch die Varianz der Schätzfunktion einbezieht

Darauf aufbauend betrachtet man daher eine Schätzfunktion als *effizienter* im Vergleich zu einer zweiten, wenn ihre mittlere quadratische Abweichung (MSE) kleiner ist

Der Vergleich zweier Schätzfunktionen hinsichtlich der Effizienz hängt im Allgemeinen von den Verteilungsannahmen in der Population ab

- Es ist also anzugeben, auf welche Verteilungsannahmen man sich bezieht (z. B. alle Verteilungen mit endlicher Varianz oder alle Normalverteilungen)

Effizientere Schätzfunktion

Eine Schätzfunktion T heißt *MSE-effizienter* als eine Schätzfunktion T', wenn für alle zugrunde gelegten Verteilungen gilt

$$MSE(T) \leq MSE(T')$$

Beschränkt man sich auf erwartungstreue Schätzfunktionen, so verschwindet der Bias und die MSE-Effizienz reduziert sich auf den Vergleich der Varianzen der Schätzfunktionen

- Von zwei erwartungstreuen Schätzfunktionen T und T' heißt T *effizienter* als T', wenn für alle zugrunde gelegten Verteilungen gilt

$$Var(T) \leq Var(T')$$

Effiziente Schätzfunktion

Für die Varianz erwartungstreuer Schätzfunktionen lässt sich eine untere Schranke angeben (Cramér-Rao-Schranke)

Die Varianz einer *effizienten* Schätzfunktion nimmt diesen kleinstmöglichen Wert an

- Eine erwartungstreue Schätzfunktion T heißt *effizient*, wenn für alle anderen erwartungstreuen Schätzfunktionen T' und alle zugrunde gelegten Verteilungen gilt

$$Var(T) \leq Var(T')$$

Effiziente Schätzfunktionen sind

- $\hat{\mu} = \bar{X}$ für alle Verteilungen mit endlicher Varianz bzw. alle Normalverteilungen
- $\hat{\lambda} = \bar{X}$ für alle Poissonverteilungen

Suffizienz

Betrachtet man die Ergebnisse eines Zufallsexperiments nicht direkt, sondern über eine Schätzfunktion $T_n = t_n(X_1, X_2, \ldots, X_n)$, so ist dies unter Umständen mit einem Informationsverlust verbunden

Beispiel

- Bei Bildung des arithmetischen Mittels

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

geht beispielsweise die Information über die Ausprägung der einzelnen X_i, aber auch über deren Reihenfolge und Streuung verloren

Suffiziente Schätzfunktion

Besonders ausgezeichnet sind daher diejenigen Schätzfunktionen, die zwar eine Reduktion der Daten erreichen, aber bezüglich der Schätzung des Parameters ϑ keinen Informationsverlust bedingen

Derartige Schätzfunktionen nennt man *suffizient* oder *erschöpfend*

Wie die bisher betrachteten Eigenschaften auch lässt sich die Suffizienz formal definieren

Da sich diese Definition aber auf den Begriff der Likelihood bezieht, über den wir an dieser Stelle noch nicht verfügen, werden wir nachfolgend nur ein Beispiel betrachten

Beispiel

Wir betrachten eine dreifache Urnenziehung nach dem Modell mit Zurücklegen, bei der die Ziehung einer roten Kugel (r) mit Wahrscheinlichkeit p und die Ziehung einer schwarzen Kugel (s) mit Wahrscheinlichkeit $1-p$ erfolgt

Die Ergebnismenge dieses Zufallsexperiments lautet

$$\begin{aligned}\Omega &= \{r,s\}^3 \\ &= \{(r,r,r),(r,r,s),(r,s,r),(s,r,r),\\ &\qquad (r,s,s),(s,r,s),(s,s,r),(s,s,s)\}\end{aligned}$$

Beispiel

Für alle $\omega = (\omega_1, \omega_2, \omega_3) \in \Omega$ und $i = 1, 2, 3$ betrachten wir die binären Zufallsvariablen X_1, X_2, X_3, definiert durch

$$X_i(\omega) = \begin{cases} 1 & \text{für } \omega_i = r \\ 0 & \text{für } \omega_i = s \end{cases}$$

Dann gilt für das Ereignis $X_1 = x_1, X_2 = x_2, X_3 = x_3$

$$\begin{aligned} P(x_1, x_2, x_3) &= p^{x_1} \cdot (1-p)^{1-x_1} \cdot p^{x_2} \cdot (1-p)^{1-x_2} \cdot p^{x_3} \cdot (1-p)^{1-x_3} \\ &= p^{x_1+x_2+x_3} \cdot (1-p)^{3-(x_1+x_2+x_3)} \end{aligned}$$

Beispiel

Für die Statistik $T = X_1 + X_2 + X_3$ erhält man über die Binomialverteilung mit Parameter p

$$P(T = k) = \binom{3}{k} p^k (1-p)^{3-k}$$

Die bedingte Wahrscheinlichkeit, die Realisierung x_1, x_2, x_3 zu beobachten, gegeben $T = k$, ist dann definiert durch

$$P(x_1, x_2, x_3 \mid T = k) = \frac{P(x_1, x_2, x_3 \text{ und } T = k)}{P(T = k)}$$

Beispiel

Somit ist

$$P(x_1, x_2, x_3 \mid T = k) = \begin{cases} 0 & \text{falls } x_1 + x_2 + x_3 \neq k \\ \frac{1}{\binom{3}{k}} & \text{falls } x_1 + x_2 + x_3 = k \end{cases}$$

Die bedingte Wahrscheinlichkeit $P(x_1, x_2, x_3 \mid T = k)$ ist also unabhängig vom Parameter p der Binomialverteilung

- Die Wahrscheinlichkeit der Daten bei gegebenem Wert von T wird durch die zusätzliche Kenntnis des Parameters p nicht mehr verändert

Beispiel

Das bedeutet, dass die Statistik T die gesamte Information der Daten über den Parameter p enthält

Die (Summen-)Statistik T der Anzahl der Fälle, in denen eine rote Kugel gezogen wird, ist also suffizient für den Parameter p einer Binomialverteilung

Die Güte der Schätzung des Parameters p kann durch Hinzunahme weiterer, in den Daten enthaltener Information nicht mehr verbessert werden

Allgemeine Charakterisierung

Eine Schätzfunktion T_n für den Parameter ϑ heißt *suffizient* für ϑ, wenn die bedingte Wahrscheinlichkeit der Daten $X_1 = x_1, \ldots, X_n = x_n$ gegeben die Schätzfunktion T_n nicht mehr vom Parameter ϑ abhängt, d. h.

$$\begin{aligned} &P(X_1 = x_1, \ldots, X_n = x_n \mid T_n = t_n, \vartheta) \\ &\quad = P(X_1 = x_1, \ldots, X_n = x_n \mid T_n = t_n) \end{aligned}$$