

Induktive Statistik

Martin Kolb

Universität Paderborn

8. Dezember 2021



In der Wahrscheinlichkeitstheorie ist das stochastische Modell typischerweise gegeben und die Aufgabe besteht in der Berechnung von Wahrscheinlichkeiten für das Auftreten gewisser Ereignisse. In gewissem Sinne befasst sich die Statistik mit einem dazu inversen Problem.



In der Wahrscheinlichkeitstheorie ist das stochastische Modell typischerweise gegeben und die Aufgabe besteht in der Berechnung von Wahrscheinlichkeiten für das Auftreten gewisser Ereignisse. In gewissem Sinne befasst sich die Statistik mit einem dazu inversen Problem.

Die zugrundeliegende Wahrscheinlichkeitsverteilung ist in der Statistik nicht explizit bekannt. Um Rückschlüsse auf die zugrundeliegende Verteilung zu ziehen, erhebt man dafür z.B. durch Experimente Daten. Die Frage, auf welche Weise man aufgrund von Daten Charakteristika der zugrundeliegenden Verteilung ermitteln kann, ist ein zentraler Gegenstand der Statistik.



Beispiel:

$X \triangleq$ Anzahl von Lesezugriffen auf eine Festplatte bis zum ersten Lesefehler. Wir spezifizieren eine geeignete Klasse von Wahrscheinlichkeitsverteilungen, die die gegebene Problemstellung geeignet modellieren.



Beispiel:

$X \triangleq$ Anzahl von Lesezugriffen auf eine Festplatte bis zum ersten Lesefehler. Wir spezifizieren eine geeignete Klasse von Wahrscheinlichkeitsverteilungen, die die gegebene Problemstellung geeignet modellieren.

Hierbei ergibt sich zum Beispiel der folgende erste sinnvolle Ansatz:

$$\mathbb{P}_p(X = i) = (1 - p)^{i-1} \cdot p, p \in [0, 1].$$

Man nimmt hierbei an, dass bei jedem Zugriff unabhängig und mit derselben Wahrscheinlichkeit p ein Lesefehler auftreten kann. Unsere Aufgabenstellung lässt sich nun wie folgt präziser formulieren: Man schätze auf eine sinnvolle Art und Weise den Wert p aus vorhandenen Daten.



Erinnerung:

$$\mathbb{P}_p(X = i) = (1 - p)^{i-1} \cdot p, p \in [0, 1].$$

Wir wissen, dass

$$\mathbb{E}_p[X] = \frac{1}{p},$$

d.h. wir können auch ebenso $\mathbb{E}_p[X]$ ermitteln.



Erinnerung:

$$\mathbb{P}_p(X = i) = (1 - p)^{i-1} \cdot p, p \in [0, 1].$$

Wir wissen, dass

$$\mathbb{E}_p[X] = \frac{1}{p},$$

d.h. wir können auch ebenso $\mathbb{E}_p[X]$ ermitteln.

Das Gesetz der großen Zahlen liefert Lösungsansatz.

Betrachte n baugleiche Platten mit zugehörigen unabhängigen Zufallsvariablen X_1, \dots, X_n , alle verteilt wie X .



Um $\mathbb{E}_p[X]$ empirisch zu ermitteln, bilde

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

Das arithmetische Mittel schätzt den Erwartungswert und liefert damit eine Schätzung für $\frac{1}{p}$.



Das typische Setting der induktiven Statistik ist also ein

- 1) Ergebnisraum Ω , eine σ -Algebra \mathcal{A} über Ω
- 2) **Familie** $(\mathbb{P}_\theta)_{\theta \in \Theta}$ von Verteilungen auf \mathcal{A}



Das typische Setting der induktiven Statistik ist also ein

- 1) Ergebnisraum Ω , eine σ -Algebra \mathcal{A} über Ω
- 2) **Familie** $(\mathbb{P}_\theta)_{\theta \in \Theta}$ von Verteilungen auf \mathcal{A}

Man versucht dann den 'richtigen' Parameter θ zu schätzen.



Das typische Setting der induktiven Statistik ist also ein

- 1) Ergebnisraum Ω , eine σ -Algebra \mathcal{A} über Ω
- 2) **Familie** $(\mathbb{P}_\theta)_{\theta \in \Theta}$ von Verteilungen auf \mathcal{A}

Man versucht dann den 'richtigen' Parameter θ zu schätzen.
Statistisches Modell = Spezifikation von Ω , \mathcal{A} und $(\mathbb{P}_\theta)_{\theta \in \Theta}$.



math. Stichprobe

Definition:

Gegeben sei eine Zufallsvariable X (Grundgesamtheit). Als **mathematische Stichprobe** vom Umfang n dieser Grundgesamtheit bezeichnet man eine Menge an Zufallsvariablen X_1, \dots, X_n mit folgenden Eigenschaften:

- Jedes X_i besitze dieselbe Verteilung wie X
- Die Zufallsvariablen X_1, \dots, X_n sind unabhängig

Ermittelt man n konkrete Werte x_1, \dots, x_n

($x_1 \in W_{X_1}, \dots, x_n \in W_{X_n}$), so nennen wir (x_1, \dots, x_n) eine **konkrete Stichprobe**.



Schätzer

Definition:

Gegeben sei eine mathematische Stichprobe X_1, \dots, X_n und eine (messbare) Funktion

$$T : W_{X_1} \times \dots \times W_{X_n} \rightarrow \mathbb{R}.$$

Dann nennt man die Zufallsvariable

$$T(X_1, \dots, X_n) : \Omega \rightarrow \mathbb{R}$$

eine mathematische Schätzfunktion bzw. Schätzer.



Beispiel:

Sei $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$. Die folgenden Funktionen sind Schätzer.

- i) $T(X_1, \dots, X_n) = \frac{1}{n} \sum_{i=1}^n X_i$
- ii) $\tilde{S}^2 = T(X_1, \dots, X_n) = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$
- iii) $S^2 = T(X_1, \dots, X_n) = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$

Beispiel:

Wir nehmen an, dass X_1^*, \dots, X_n^* aus X_1, \dots, X_n gebildet werden, indem wir die Werte der Größe nach ordnen, d.h.

$$X_1^* = \min(X_1, \dots, X_n), \dots, X_n^* = \max(X_1, \dots, X_n).$$



Beispiel:

Wir nehmen an, dass X_1^*, \dots, X_n^* aus X_1, \dots, X_n gebildet werden, indem wir die Werte der Größe nach ordnen, d.h.

$$X_1^* = \min(X_1, \dots, X_n), \dots, X_n^* = \max(X_1, \dots, X_n).$$

Dann ist

$$X_{0,5} = T(X_1, \dots, X_n) = \begin{cases} X_{\frac{n+1}{2}}^* & n \text{ ungerade} \\ \frac{1}{2} \left(X_{\frac{n}{2}}^* + X_{\frac{n+1}{2}}^* \right) & n \text{ gerade} \end{cases}$$

eine mathematische Schätzfunktion und heißt Stichprobenmedian.



Der Begriff des Schätzers ist sehr allgemein. Es wird also **gute und schlechte** Schätzer für einen Parameter geben und wir werden uns mit einigen **Gütekriterien** befassen.



Erwartungstreue

Definition:

Es sei eine mathematische Stichprobe X_1, \dots, X_n für die Grundgesamtheit X auf Ω gegeben. Ein Schätzer $T(X_1, \dots, X_n)$ heißt **erwartungstreu** für Parameter $p \in \mathbb{R}$, wenn

$$\mathbb{E}[T(X_1, \dots, X_n)] = p.$$

Die Folge von Schätzern $(T_n(X_1, \dots, X_n))_{n \in \mathbb{N}}$ heißt **asymptotisch erwartungstreu**, wenn

$$\lim_{n \rightarrow \infty} \mathbb{E}[T_n(X_1, \dots, X_n)] = p.$$



Beispiel:

Sei X die Grundgesamtheit und X_1, \dots, X_n eine mathematische Stichprobe für X . Dann ist

$$T(X_1, \dots, X_n) = \frac{1}{n} \sum_{i=1}^n X_i$$

ein erwartungstreuer Schätzer für $\mathbb{E}[X]$, denn es gilt

$$\begin{aligned} \mathbb{E}[T(X_1, \dots, X_n)] &= \mathbb{E}\left[\frac{1}{n} \sum_{i=1}^n X_i\right] = \frac{1}{n} \sum_{i=1}^n \mathbb{E}[X_i] \\ &= \frac{1}{n} \sum_{i=1}^n \mathbb{E}[X] = \mathbb{E}[X]. \end{aligned}$$

Beispiel:

Betrachte aus Beispiel nochmals den Schätzer

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

Dann ist \tilde{S}^2 nicht erwartungstreu für $p = \text{Var}(X)$.



Untersuche $\mathbb{E}[\tilde{S}^2]$:

$$\begin{aligned}(X_i - \bar{X}) &= (X_i - \mathbb{E}[X]) - (\bar{X} - \mathbb{E}[X]) \\ &= (X_i - \mathbb{E}[X]) - \left(\frac{1}{n} \sum_{j=1}^n (X_j - \mathbb{E}[X]) \right)\end{aligned}$$



Untersuche $\mathbb{E}[\tilde{S}^2]$:

$$\begin{aligned}(X_i - \bar{X}) &= (X_i - \mathbb{E}[X]) - (\bar{X} - \mathbb{E}[X]) \\ &= (X_i - \mathbb{E}[X]) - \left(\frac{1}{n} \sum_{j=1}^n (X_j - \mathbb{E}[X]) \right)\end{aligned}$$

Mit binom. Formel folgt

$$\begin{aligned}\mathbb{E}[(X_i - \bar{X})^2] &= \mathbb{E}[(X_i - \mathbb{E}[X])^2] + \mathbb{E}[(\bar{X} - \mathbb{E}[X])^2] \\ &\quad - \frac{2}{n} \cdot \mathbb{E} \left[\sum_{j=1}^n (X_j - \mathbb{E}[X])(X_i - \mathbb{E}[X]) \right]\end{aligned}$$



Weiter ist für den letzten Term

$$\begin{aligned}\mathbb{E}\left[\sum_{j=1}^n (X_j - \mathbb{E}[X])(X_i - \mathbb{E}[X])\right] &= \sum_{j=1}^n \mathbb{E}[(X_j - \mathbb{E}[X])(X_i - \mathbb{E}[X])] \\ &= \mathbb{E}[(X_i - \mathbb{E}[X])^2] = \text{Var}(X_i) \\ &= \text{Var}(X)\end{aligned}$$



Weiter ist für den letzten Term

$$\begin{aligned}\mathbb{E}\left[\sum_{j=1}^n (X_j - \mathbb{E}[X])(X_i - \mathbb{E}[X])\right] &= \sum_{j=1}^n \mathbb{E}[(X_j - \mathbb{E}[X])(X_i - \mathbb{E}[X])] \\ &= \mathbb{E}[(X_i - \mathbb{E}[X])^2] = \text{Var}(X_i) \\ &= \text{Var}(X)\end{aligned}$$

und somit

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



Satz:

Die mathematische Schätzfunktion

$$S^2 = T(X_1, \dots, X_n) = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

ist erwartungstreu für die Varianz der Grundgesamtheit X .



Weitere wünschenswerte Eigenschaften:



Weitere wünschenswerte Eigenschaften:

Definition:

Für jedes $n \in \mathbb{N}$ sei $T_n(X_1, \dots, X_n)$ eine mathematische Schätzfunktion für die mathematische Stichprobe (X_1, \dots, X_n) der Grundgesamtheit X .

Die Folge $(T_n(X_1, \dots, X_n))_{n \in \mathbb{N}}$ von Schätzern heißt **konsistent**, wenn für Parameter p und jedes $\epsilon > 0$ gilt

$$\lim_{n \rightarrow \infty} \mathbb{P}(|T_n(X_1, \dots, X_n) - p| > \epsilon) = 0.$$

Beispiel:

Sei (X_1, \dots, X_n) eine mathematische Stichprobe für die Grundgesamtheit X . Weiter besitze X einen Erwartungswert und es gelte $\text{Var}(X) < \infty$. Dann ist

$$T_n(X_1, \dots, X_n) = \frac{1}{n} \sum_{i=1}^n X_i$$

eine konsistente Schätzfolge für $\mathbb{E}[X]$, denn nach dem Gesetz der großen Zahlen gilt für jedes $\epsilon > 0$

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(\left| \frac{1}{n} \sum_{i=1}^n X_i - \mathbb{E}[X] \right| > \epsilon \right) = 0.$$



Oft hat man zwei Schätzer für den selben Parameter und möchte die Güte der beiden sinnvoll vergleichen.



Oft hat man zwei Schätzer für den selben Parameter und möchte die Güte der beiden sinnvoll vergleichen.

Definition:

Es sei (X_1, \dots, X_n) eine mathematische Stichprobe für die Grundgesamtheit X und seien $T_1(X_1, \dots, X_n)$ und $T_2(X_1, \dots, X_n)$ zwei erwartungstreue Schätzer für einen Parameter μ der Grundgesamtheit X . Der Schätzer $T_1(X_1, \dots, X_n)$ heißt **effizienter** als $T_2(X_1, \dots, X_n)$, wenn

$$\text{Var}\left(T_1(X_1, \dots, X_n)\right) \leq \text{Var}\left(T_2(X_1, \dots, X_n)\right).$$

Bemerkung:

Die Frage, welche wünschenswerten Eigenschaften wie z.B. Erwartungstreue ein Schätzer erfüllen soll, hängt von der Situation ab. In verschiedenen Kontexten ist es durchaus denkbar, dass man bewusst nicht erwartungstreue Schätzer wählt. Man denke hier zum Beispiel an Messungen des Blutzuckerwertes. Niedrige Blutzuckerwerte sind kurzfristig gefährlicher als höhere.



Beispiel:

Wir betrachten das Schätzproblem der Erfolgswahrscheinlichkeit bei einer Bernoullikette, i.e. $\mathcal{X} = \{0, \dots, n\}$, $\Theta = [0, 1]$ und

$$\mathbb{P}_\theta(\{x\}) = \binom{n}{k} \theta^k (1 - \theta)^{n-k}.$$

Ein natürlicher Schätzer für den Parameter θ ist gegeben durch

$$T(x) = \frac{x}{n}.$$

Dieser Schätzer ist sogar derjenige erwartungstreue Schätzer mit minimaler Varianz.



Der **Mittlere Quadratische Fehler** des Schätzers T ist gegeben durch

$$\text{MQF}(T) = \mathbb{E}_{\theta} [|T - \theta|^2] = \frac{\theta(1 - \theta)}{n}.$$

Der Schätzer

$$S(x) = \frac{x + 1}{n + 2}$$

für den Parameter θ hat allerdings einen kleineren Mittleren Quadratischen Fehler für den Fall, dass

$$|\theta - 1/2| \leq 1/\sqrt{8}.$$



Konstruktion von Schätzern

Definition:

Es sei X eine Grundgesamtheit und seien X_1, \dots, X_n eine mathematische Stichprobe vom Umfang n . Sei weiter $f(\cdot, p)$ die Dichte der Zufallsvariable X_i .

Dann ist die **Likelihood-Funktion** $L(x, p)$ für die konkrete Stichprobe $x = (x_1, \dots, x_n)$ wie folgt definiert

$$L(x, p) = L_x(p) = \prod_{k=1}^n f(x_k, p).$$

Man beachte, dass die Likelihood-Funktion als eine Funktion in den Parametern ist.

Konstruktion von Schätzern

Definition:

Der Schätzer $T(X_1, \dots, X_n)$ heißt **Maximum-Likelihood-Schätzer** für einen Parameter p und eine konkrete Stichprobe x_1, \dots, x_n , falls für alle möglichen Werte des Parameters gilt

$$L(x_1, \dots, x_n, p) \leq L(x_1, \dots, x_n, T(x_1, \dots, x_n)).$$

Wir maximieren also bei gegebenen Daten die Likelihood-Funktion im Parameter.



Zusammenfassend sei also nochmals explizit angemerkt, dass man eine **Maximum-Likelihood-Schätzung** in zwei Schritten vollzieht:

- Aufstellen einer zum statistischen Modell gehörenden Likelihood-Funktion.
- Maximierung dieser Funktion.

Der Maximum-Likelihood-Schätzer ist also derjenige Parameterwert, der zur Verteilung gehört, bzgl. der die gegebenen Daten so wahrscheinlich wie möglich sind.



Beispiel:

Wir untersuchen die Lebensdauer eines technischen Produkts, wobei zum Beispiel wegen der experimentell untersuchten Gedächtnislosigkeit angenommen wird, dass die Lebensdauer exponentialverteilt ist mit unbekanntem Parameter $\lambda > 0$.

Das Experiment liefert Ihnen eine konkrete Stichprobe vom Umfang n : $(x_1, \dots, x_n) \in (0, \infty)^n$.

Wir bestimmen einen Schätzer über die

Maximum-Likelihood-Methode: Für $x = (x_1, \dots, x_n)$ ist die Likelihood-Funktion gegeben durch

$$L(x, \lambda) = \mathbf{1}_{[0, \infty)}(x_1) \cdots \mathbf{1}_{[0, \infty)}(x_n) \cdot \lambda^n e^{-\lambda x_1} \cdots e^{-\lambda x_n}$$

Die Maximum-Likelihood-Methode besagt, dass wir die Likelihood-Funktion im Parameter λ maximieren. Betrachte bei $x_1, \dots, x_n > 0$ stattdessen die Log-Likelihood-Funktion

$$\log L(x, \lambda) = n \cdot \log(\lambda) - \lambda \cdot \sum_{i=1}^n x_i.$$

Wir bestimmen einen Kandidaten durch Bestimmen der Ableitung

$$\frac{d}{d\lambda} \log(L(x, \lambda)) = n \cdot \frac{1}{\lambda} - \sum_{i=1}^n x_i$$

Diese wird 0 bei

$$\lambda_{ML} = \left(\frac{1}{n} \sum_{x_i} \right)^{-1}$$

Leicht ist zu sehen, dass λ_{ML} ein Maximum ist und damit erhalten wir

$$T_{ML}(X_1, \dots, X_n) = \frac{1}{\frac{1}{n} \sum_{i=1}^n X_i}$$