Statistik und empirische Methoden
Wahrscheinlichkeitsrechnung

Silke Bott

Sommersemester 2023

Kovarianz und gemeinsame Verteilungen

Wir betrachten nun zwei **stochastische unabhängige** Zufallsvariablen X und Y, beide stetig oder beide diskret, mit (diskreten oder stetigen) Dichten f_X und f_Y.

Definition

$f(x,y) = f_X(x) \cdot f_Y(y)$ heißt **gemeinsame Dichte** von X und Y.

Sind X und Y diskret, so gilt

$$p(X = y, Y = y) = f(x,y)$$

Sind X und Y stetig, so gilt

$$p(a \leq X \leq b, c \leq Y \leq d) = \int_c^d \int_a^b f(x,y)\, dxdy$$

Analog kann man eine gemeinsame Dichte von drei und mehr stochastisch unabhängigen Zufallsvariablen definieren.

Kovarianz und gemeinsame Verteilungen

Definition

Sind X und Y unabhängig und diskret verteilt, mit Wertebereichen W_X und W_Y, so heißt

$$f(x,y) = \begin{cases} p(X = x, Y = y) & \text{falls} \quad x \in W_X, y \in W_Y \\ 0 & \text{sonst} \end{cases}$$

die gemeinsame (diskrete) Dichtefunktion von X und Y, und

$$F(x,y) = \sum_{\xi \leq x, \eta \leq y} f(\xi, \eta)$$

heißt die gemeinsame Verteilungsfunktion von X und Y.

Kovarianz und gemeinsame Verteilungen

Bemerkung

Die diskreten Dichten f_X von X und f_Y von Y ergeben sich aus der gemeinsamen Dichtefunktion f als Randdichten, also

$$\begin{aligned} f_X(x) &= p(X = x) = \sum_{y \in W_Y} f(x, y) \\ f_Y(y) &= p(Y = y) = \sum_{x \in W_X} f(x, y) \end{aligned}$$

Definition

Sind X und Y diskret verteilt, so ist die bedingte Wahrscheinlichkeitsfunktion von X gegeben $Y = y$ definiert als

$$F_X(x|Y = y) = \frac{f(x, y)}{f_Y(y)} \quad \text{falls } f_Y(y) > 0$$

Kovarianz und gemeinsame Verteilungen

Beispiel

Wir betrachten das zweimalige unabhängige Würfeln eines Laplacewürfels und die beiden Zufallsvariablen X: Ergebnis des ersten Wurfes und Y: Ergebnis des zweiten Wurfes.

Dann ist die gemeinsame diskrete Dichtefunktion von X und Y gegeben durch

$$f(x,y) = \begin{cases} \frac{1}{36} & \text{falls} \quad x \in 1,...,6, y \in 1,...,6 \\ 0 & \text{sonst} \end{cases}$$

Kovarianz und gemeinsame Verteilungen

Übung

Wir betrachten ein Zufallsexperiment, dass aus dem Würfeln eines Laplacewürfels und einem Laplacemünzwurf besteht und wie folgt aufgebaut ist:
Der Laplacewürfel wird geworfen und die Zufallsvariable X notiert sein Ergebnis. Ist das Ergebnis gerade, so wird die Laplacemünze geworfen und Y notiert das Ergebnis dieses Wurfes (0= für Zahl und 1 für Kopf). Ist das Würfelergebnis ungerade, wird $Y = 1$ notiert.
Bestimmen Sie die gemeinsame Dichte von X und Y.

Kovarianz und gemeinsame Verteilungen

Lösung:

Die gemeinsame Dichte von X und Y ist gegeben durch

$$f(x,y) = \begin{cases} \frac{1}{12} & \text{für} \quad x \in 2,4,6, y \in 0,1 \\ \frac{1}{6} & \text{für} \quad x \in 1,3,5, y = 1 \\ 0 & \text{sonst} \end{cases}$$

Kovarianz und gemeinsame Verteilungen

Definition

Zwei Zufallsvariablen X und Y heißen gemeinsam stetig verteilt, wenn es eine zweidimensionale (integrierbare) Dichtefunktion $f(x,y)$ gibt, so dass

$$p(a \leq X \leq b, c \leq Y \leq d) = \int_a^b \int_c^d f(x,y)\, dydx$$

für alle $a \leq b$ und $c \leq d$.

Bemerkung

Sind X und Y gemeinsam stetig verteilt mit gemeinsamer Dichte f, so sind X und Y jeweils für sich betrachtet stetig verteilt mit den Randdichten

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\, dy, \qquad f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\, dx$$

Kovarianz und gemeinsame Verteilungen

Beispiel

Das Gewicht X des Inhalts einer Pralinenschachtel ist normalverteilt mit Mittelwert $\mu_1 = 250$g und Streuung $\sigma_1 = 1.2$g und das Gewicht Y der Packung ist normalverteilt mit Mittelwert $\mu_2 = 40$g und Streuung $\sigma_2 = 0.4$g, wobei Gewicht von Inhalt und Verpackung unabhängig voneinander sind. Dann besitzen X und Y eine gemeinsame stetige Dichte $f(x, y)$, gegeben durch

$$\begin{aligned} f(x,y) &= \frac{1}{\sqrt{2\pi} \cdot 1.2} \cdot exp\left(-\frac{(x-250)^2}{2.88}\right) \cdot \frac{1}{\sqrt{2\pi} \cdot 0.4} \cdot exp\left(-\frac{(y-40)^2}{0.32}\right) \\ &= \frac{1}{2\pi \cdot 0.48} \cdot exp\left(-\frac{(x-250)^2}{2.88} - \frac{(y-40)^2}{0.32}\right) \end{aligned}$$

Kovarianz und gemeinsame Verteilungen

Übung

Von den beiden stetigen Zufallsvariablen X und Y ist bekannt, dass sie die gemeinsame stetige Dichte

$$f(x,y) = \begin{cases} x+y & \text{für} \quad 0 \leq x \leq 1, 0 \leq y \leq 1 \\ 0 & \text{sonst} \end{cases}$$

besitzen.
Berechnen Sie $p(X \leq 0.5, Y \leq 0.6)$.

Lösung:

$p(X \leq 0.5, Y \leq 0.6) = 0.165.$

Kovarianz und gemeinsame Verteilungen

Beispiel

Die beiden stetigen Zufallsvariablen X und Y mit gemeinsamer Dichte

$$f(x,y) = \begin{cases} x+y & \text{für} \quad 0 \le x \le 1, 0 \le y \le 1 \\ 0 & \text{sonst} \end{cases}$$

besitzen die Randdichten.

$$f_X(x) = \begin{cases} x+0.5 & \text{für} \quad 0 \le x \le 1 \\ 0 & \text{sonst} \end{cases}$$

Hier ist $f_X(x) \cdot f_Y(y)$ keine gemeinsame Dichte von X und Y, und daher sind X und Y nicht stochastisch unabhängig.

Kovarianz und gemeinsame Verteilungen

Sind X und Y stochastisch unabhängige diskrete oder stetige Zufallsvariablen und sind $f_X(x)$ bzw. $f_Y(y)$ Dichtefunktionen von X bzw. Y, so ist

$$f(x,y) = f_X(x) \cdot f_Y(y) \qquad \text{für } (x,y) \in \mathbb{R}^2$$

eine gemeinsame Dichte von X und Y.

Kovarianz und gemeinsame Verteilungen

Definition (Kovarianz und Korrelation)

Für zwei stetige oder diskrete Zufallsvariablen X und Y heißt

$$\mathrm{Cov}(X, Y) = E\left((X - E(X)) \cdot (Y - E(Y))\right)$$

die **Kovarianz** von X und Y, falls dieser Ausdruck existiert, und

$$\rho(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X) \cdot \mathrm{Var}(Y)}} = \frac{\mathrm{Cov}(X, Y)}{\sigma_X \cdot \sigma_Y}$$

heißt die **Korrelation** von X und Y. Dabei heißen X und Y **unkorreliert**, wenn

$$\rho(X, Y) = 0$$

und andernfalls heißen sie korreliert.

Kovarianz und gemeinsame Verteilungen

Bemerkung

Sind X und Y unabhängig und diskret mit gemeinsamer Dichte f, so gilt

$$\mathrm{Cov}(X,Y) = \sum_{x,y} (x - E(X)) \cdot (y - E(Y)) \cdot f(x,y)$$

Sind X und Y unabhängig und stetig verteilt mit gemeinsamer Dichte f, so gilt

$$\mathrm{Cov}(X,Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - E(X)) \cdot (y - E(Y)) \cdot f(x,y)\, dydx$$

Kovarianz und gemeinsame Verteilungen

Regel

Für zwei Zufallsvariablen X und Y gilt

1. *Verschiebungssatz:* $\mathrm{Cov}(X, Y) = E(X \cdot Y) - E(X) \cdot E(Y)$.
2. *Symmetrie* $\mathrm{Cov}(X, Y) = \mathrm{Cov}(Y, X)$.
3. *Transformation: Für* $X_1 = aX + b$, $Y_1 = cY + d$ *gilt*
$$\mathrm{Cov}(X_1, Y_1) = a \cdot c \cdot \mathrm{Cov}(X, Y)$$
4. *Summenregel:* $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2 \cdot \mathrm{Cov}(X, Y)$.

Kovarianz und gemeinsame Verteilungen

Bemerkung

Es ist $-1 \leq \rho(X, Y) \leq 1$.

Regel

Es gilt:

1. $\mathrm{Cov}(X, X) = \mathrm{Var}(X)$.
2. *Ist* $Y = a \cdot X + b$, *so gilt*
   - $\rho(X, Y) = 1$, *falls* $a > 0$.
   - $\rho(X, Y) = -1$, *falls* $a < 0$.

Kovarianz und gemeinsame Verteilungen

Beispiel

Wir betrachten das zweimalige unabhängige Werfen einer Laplacemünze und die Zufallsvariablen X: *Ergebnis des ersten Wurfes* und Y: *Ergebnis des zweiten Wurfes*, wobei wir 0 für *Zahl* und 1 für *Kopf* notieren.

$$E(X) = \frac{1}{2}, \qquad E(Y) = \frac{1}{2}$$

$$\begin{aligned}
\mathrm{Cov}(X, Y) &= \tfrac{1}{4} \cdot \left(0 - \tfrac{1}{2}\right) \cdot \left(0 - \tfrac{1}{2}\right) + \tfrac{1}{4} \cdot \left(0 - \tfrac{1}{2}\right) \cdot \left(1 - \tfrac{1}{2}\right) \\
&\quad + \tfrac{1}{4} \cdot \left(1 - \tfrac{1}{2}\right) \cdot \left(0 - \tfrac{1}{2}\right) + \tfrac{1}{4} \cdot \left(1 - \tfrac{1}{2}\right) \cdot \left(1 - \tfrac{1}{2}\right) \\
&= \frac{1}{16} - \frac{1}{16} - \frac{1}{16} + \frac{1}{16} \\
&= 0
\end{aligned}$$

also auch

$$\rho(X, Y) = 0$$

Kovarianz und gemeinsame Verteilungen

Beispiel

Wir betrachten das zweimalige unabhängige Werfen eines Laplacewürfels und die Zufallsvariablen X: *Ergebnis des ersten Wurfes* und Y: *Ergebnis des zweiten Wurfes*. Dann gilt

$$E(X) = \frac{7}{2}, \qquad E(Y) = \frac{7}{2}$$

und damit

$$\begin{aligned} \operatorname{Cov}(X, Y) &= \sum_{i,j=1}^{6} \tfrac{1}{36} \cdot \left(i - \tfrac{7}{2}\right) \cdot \left(j - \tfrac{7}{2}\right) \\ &= \tfrac{1}{36} \cdot \left(\sum_{i=1}^{6} \left(i - \tfrac{7}{2}\right) \right) \cdot \left(\sum_{j=1}^{6} \left(j - \tfrac{7}{2}\right) \right) \\ &= 0 \end{aligned}$$

also auch $\rho(X, Y) = 0$.

Kovarianz und gemeinsame Verteilungen

Übung

Wir betrachten das Werfen einer Laplacemünze und die Zufallsvariablen X: *Ergebnis des Wurfes*, wobei wir 0 für *Zahl* und 1 für *Kopf*. Gilt $X = 0$, so wird mit der Laplacemünze nocheinmal geworfen und Y bezeichnet des Ergebnis des zweiten Wurfes. Gilt dagegen $X = 1$, so wird für Y automatisch 0 notiert. Die drei möglichen Ergebnisse des Experiments sind $(0, 0)$, $(0, 1)$ und $(1, 1)$.
Bestimmen Sie Kovarianz und Korrelation von X und Y.

Kovarianz und gemeinsame Verteilungen

Lösung:

Es gilt

$$\mathrm{Cov}(X, Y) = -\frac{1}{8}$$

und

$$\rho(X, Y) = -\frac{1}{\sqrt{3}}$$

Parameterschätzung

Statistische Untersuchungen, bei denen Stichproben gezogen und betrachtet werden, dienen dem Zweck, ein möglichst gutes Bild der Gesamtsituation zu gewinnen und Daten zu ermitteln, die für die Parameter der Grundgesamtheit aussagekräftig sind. Schätzverfahren dienen also dazu, aus einer Zufallsstichprobe auf ein Merkmal der Grundgesamtheit zurückzuschliessen, dass durch eine Zufallsvariable beschrieben wird. Dabei konzentrieren wir uns auf bestimmte Aspekte (Parameter) dieses Merkmals, speziell

- Erwartungswert und Varianz einer Zufallsvariable.
- Median oder p–Quantile einer Zufallsvariable.
- Korrelation zweier Zufallsvariablen.

Parameterschätzung

Grundlage der Schätzung ist eine Stichprobe vom Umfang n, also n unabhängige und identisch verteilte Zufallsvariablen $X_1, \dots, X_n$, die verteilt sind wie die Zufallsvariable X (dh. n unabhängige Wiederholungen von X).
Gesucht ist ein Parameter θ der Zufallsvariable X (z.B. Erwartungswert oder Varianz).
Die Realisationen dieser Zufallsvariablen $X_1, \dots, X_n$, also die tatsächlich im Rahmen der Stichprobe gemessenen Werte, bezeichnen wir mit $x_1, \dots, x_n$.

Parameterschätzung

Definition

Eine *Schätzfunktion* oder *Schätzstatistik* für den Parameter θ der Grundgesamtheit ist eine Funktion

$$T = g(X_1, \ldots, X_n)$$

der Stichprobenvariablen $X_1, \ldots, X_n$. Der aus den Realisationen $x_1, \ldots, x_n$ gewonnene Wert

$$t = g(x_1, \ldots, x_n)$$

heißt der dazugehörige *Schätzwert*.

Beispiel

Das arithmetische Mittel ist ein Schätzer für den Erwartungswert

$$\overline{X} = \frac{1}{n} \cdot \sum_{k=1}^{n} X_k$$

Parameterschätzung

Beispiel

Die Schätzstatistik

$$S^2 = \frac{1}{n-1} \cdot \sum_{k=1}^{n} (X_k - \overline{X})^2$$

ist ein Schätzer für die Varianz.

$$s^2 = \frac{1}{n-1} \cdot \sum_{k=1}^{n} (x_k - \overline{x})^2$$

ist die korrigierte empirische Varianz der Stichprobe.

Parameterschätzung

Beispiel

Die Schätzstatistik

$$\widetilde{S}^2 = \frac{1}{n} \cdot \sum_{k=1}^{n} (X_k - \overline{X})^2$$

ist ein Schätzer für die Varianz.

$$\widetilde{s}^2 = \frac{1}{n} \cdot \sum_{k=1}^{n} (x_k - \overline{x})^2$$

ist die empirische Varianz der Stichprobe.

Parameterschätzung

Definition

Eine Schätzstatistik T heißt *erwartungstreu* oder unverzerrt für den Parameter θ, wenn

$$E(T) = \theta$$

Der *Bias* der Schätzstatistik T ist

$$\mathrm{Bias}(T) = E(T) - \theta$$

Hat eine Schätzstatistik einen Bias, so bedeutet das, dass der zu schätzende Parameter systematisch über- oder unterschätzt wird.
In der Regel ist das zu vermeiden, es kann aber erwünscht sein, wenn ein Schätzfehler in einer Richtung stärkere Auswirkungen hat als der in der anderen Richtung.

Parameterschätzung

Beispiel

Das arithmetische Mittel

$$\overline{X} = \frac{1}{n} \cdot \sum_{k=1}^{n} X_k$$

ist ein erwartungstreuer Schätzer für den Erwartungswert von X.

Es gilt nämlich

$$\begin{aligned} E(\overline{X}) &= E\left(\frac{1}{n} \cdot \sum_{k=1}^{n} X_k\right) \\ &= \frac{1}{n} \cdot \sum_{k=1}^{n} E(X_k) \\ &= \frac{1}{n} \cdot \sum_{k=1}^{n} \mu = \mu \\ &= E(X) \end{aligned}$$

Parameterschätzung

Übung

Die Schätzstatistiken

$$S^2 = \frac{1}{n-1} \cdot \sum_{k=1}^{n} (X_k - \overline{X})^2$$

und

$$\widetilde{S}^2 = \frac{1}{n} \cdot \sum_{k=1}^{n} (X_k - \overline{X})^2$$

sind Schätzstatistiken für die Varianz.
Untersuchen Sie S^2 und $\widetilde{S}^2$ auf Erwartungstreue.

Lösung:

Der Schätzer S^2 ist erwartungstreu, $\widetilde{S}^2$ nicht.

Parameterschätzung

Der entscheidende Punkt bei den Berechnungen ist die Ergänzung (mit $\mu = E(X)$):

$$\begin{aligned}(X_k - \overline{X})^2 &= (X_k - \mu + \mu - \overline{X})^2 \\ &= (X_k - \mu)^2 + 2(X_k - \mu) \cdot (\mu - \overline{X}) + (\mu - \overline{X})^2 \\ &= (X_k - \mu)^2 - 2(X_k - \mu) \cdot (\overline{X} - \mu) + (\overline{X} - \mu)^2\end{aligned}$$

und die Tatsache, dass

$$E\left((X_k - \mu)^2\right) = \mathrm{Var}(X), \quad E\left((\overline{X} - \mu)^2\right) = \frac{1}{n} \cdot \mathrm{Var}(X)$$

Parameterschätzung

Definition

Der *Standardfehler* der Schätzstatistik T ist die Standardabweichung der Schätzstatistik

$$\sigma_g = \sqrt{\mathrm{Var}(g(X_1, \ldots, X_n))}$$

und die *erwartete mittlere Abweichung* der Schätzstatistik T ist

$$\mathrm{MSE}(T) = E((T - \theta)^2)$$

Je geringer der Standardfehler ist, umso weniger werden die Schätzwerte streuen, und je geringer die erwartete mittlere Abweichung eines Schätzers ist, umso weniger werden die Schätzwerte um den tatsächlichen Parameter streuen.

Parameterschätzung

Die Schätzstatistiken $T = T_n$ sind häufig für alle $n \in \mathbb{N}$ definiert.

Definition

Eine Schätzstatistik $(T_n)_{n\in\mathbb{N}}$ heißt *asymptotisch erwartungstreu*, wenn

$$\lim_{n\to\infty} E(T_n) = \theta$$

und sie heißt *konsistent im quadratische Mittel*, wenn

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

und *schwach konsistent*, wenn für jedes $\varepsilon > 0$ gilt:

$$\lim_{n\to\infty} p(|T_n - \theta| < \varepsilon) = 1$$

Parameterschätzung

Definition

Eine Schätzstatistik T_1 heißt *MSE–wirksammer* als T_2, wenn

$$\mathrm{MSE}(T_1) \leq \mathrm{MSE}(T_2)$$

Sind T_1 und T_2 erwartungstreu, so heißt T_1 *wirksamer* als T_2, wenn

$$\mathrm{Var}(T_1) \leq \mathrm{Var}(T_2)$$

Eine Schätzstatistik T heißt *MSE–wirksamst*, wenn sie *MSE*–wirksamer ist als alle andere zugelassenen Schätzer, und eine erwartungstreue Schätzstatistik heißt *wirksamst* oder *effizient*, wenn sie wirksamer ist als aller andern zugelassenen erwartungstreuen Schätzer.

Parameterschätzung

Beispiel

Bei $n = 2$ sind die Schätzstatistiken

$$T_1 = \frac{1}{10} \cdot X_1 + \frac{9}{10} \cdot X_2$$

und

$$T_2 = \frac{1}{4} \cdot X_1 + \frac{3}{4} \cdot X_2$$

erwartungstreu für $\mu = E(X)$, aber T_2 ist wirksamer als T_1.
Es ist

$$E(T_1) = \frac{1}{10} \cdot E(X_1) + \frac{9}{10} \cdot E(X_2) = \frac{1}{10} \cdot E(X) + \frac{9}{10} \cdot E(X) = E(X)$$

und

$$E(T_2) = \frac{1}{4} \cdot E(X_1) + \frac{3}{4} \cdot E(X_2) = \frac{1}{4} \cdot E(X) + \frac{3}{4} \cdot E(X) = E(X)$$

Parameterschätzung

Beispiel

Ferner gilt (wegen der stochastischen Unabhängigkeit)

$$\begin{aligned}
\mathrm{Var}(T_1) &= \frac{1}{100}\cdot \mathrm{Var}(X_1) + \frac{81}{100}\cdot \mathrm{Var}(X_2) \\
&= \frac{1}{100}\cdot \mathrm{Var}(X) + \frac{81}{100}\cdot \mathrm{Var}(X) \\
&= \frac{82}{100}\cdot \mathrm{Var}(X)
\end{aligned}$$

und

$$\begin{aligned}
\mathrm{Var}(T_2) &= \frac{1}{16}\cdot \mathrm{Var}(X_1) + \frac{9}{16}\cdot \mathrm{Var}(X_2) \\
&= \frac{1}{16}\cdot \mathrm{Var}(X) + \frac{9}{16}\cdot \mathrm{Var}(X) \\
&= \frac{5}{8}\cdot \mathrm{Var}(X)
\end{aligned}$$

also ist T_2 effizienter als T_1.

Parameterschätzung

Übung

Wir betrachten 5 unabhängige Wiederholungen $X_1, \ldots, X_5$ von X. Für μ betrachten wir die Schätzfunktionen

$$
\begin{aligned}
T_1 &= \overline{X} = \tfrac{1}{5}(X_1 + X_2 + X_3 + X_4 + X_5) \\
T_2 &= \tfrac{1}{2}(X_1 + X_2) \\
T_3 &= \tfrac{1}{6}(X_1 + X_2 + X_3) + \tfrac{1}{4}(X_4 + X_5) \\
T_4 &= \tfrac{1}{4}(X_1 + X_2 + X_3) + \tfrac{1}{6}(X_4 + X_5) \\
T_5 &= X_1 + X_2 + X_4
\end{aligned}
$$

a) Welche Schätzer sind erwartungstreu?

b) Welcher der erwartungstreuen Schätzer ist der wirksamste unter diesen?

Parameterschätzung

Lösung:

a) Die Schätzer T_1, T_2, und T_3 sind erwartungstreue Schätzer, T_4 und T_5 nicht.

b) Der Schätzer T_1 ist der wirksamste Schätzer von T_1, T_2 und T_3.

Maximum-Likelihood–Schätzung

Wir betrachten eine Zufallsvariable X, über die wir bereits Informationen besitzen (z.B. dass sie binomialverteilt ist), bei der jedoch noch eine Parameter θ zur Bestimmung der exakten Verteilung fehlt (z.B. p bei der Binomialverteilung).
Ist θ_0 der tatsächliche Wert des Parameters, so bezeichnen wir mit $f(x|\theta_0)$ die (diskrete oder stetige) Dichte in diesem Fall. Dann gilt für die Dichtefunktion der n–fachen Wiederholung

$$f(x_1, \ldots, x_n|\theta_0) = \prod_{i=1}^{n} f(x_i|\theta_0) = f(x_1|\theta_0) \cdots f(x_n|\theta_0)$$

Definition

Bei bekannter Realisation $x_1, \ldots, x_n$ heißt $L(\theta) = f(x_1, \ldots, x_n|\theta)$ die **Likelihoodfunktion** an der Stelle θ.

Maximum-Likelihood–Schätzung

Definition

Eine Schätzfunktion $\widehat{\theta}$ für den Parameter θ ist nach dem Maximum–Likelihood–Prinzip konstruiert oder kurz **Maximum–Likelihood–Schätzer**, wenn

$$L(\widehat{\theta}) \geq L(\widetilde{\theta}) \quad \text{für alle zugelassenen } \widetilde{\theta}$$

Bemerkung

Oft wird die **log–Likelihoodfunktion** $l(\theta) = \ln(L(\theta))$ untersucht, da diese häufig einfacher zu handhaben ist. Da ln streng monoton wachsend ist, stimmen Maxima von $L(\theta)$ und $l(\theta)$ überein.

Maximum-Likelihood–Schätzung

Konstruktion eines Maximum–Likelihood–Schätzers:

1. Bestimme die Likelihood–Funktion $L(\theta) = f(x_1|\theta) \cdots f(x_n|\theta)$.
2. Bestimme die Log–Likelihood–Funktion $l(\theta) = \ln(L(\theta))$.
3. Bestimme die Ableitung $l'(\theta)$ (bzw. $L'(\theta)$, je nachdem, was einfacher wird).
4. Bestimme $\widehat{\theta}$ mit $l'(\widehat{\theta}) = 0$ (bzw. $L'(\widehat{\theta}) = 0$.
5. Überprüfe, ob $\widehat{\theta}$ ein Maximum von $l(\theta)$ (bzw. $L(\theta)$ ist) (z.B. über die zweite Ableitung).

Maximum-Likelihood–Schätzung

Beispiel

Bei n–facher Wiederholung einer Poisson–verteilten Zufallsvariable X mit unbekanntem Parameter λ erhalten wir:

$$\begin{aligned} L(\lambda) &= e^{-\lambda}\frac{\lambda^{x_1}}{x_1!} \cdot e^{-\lambda}\frac{\lambda^{x_2}}{x_2!} \cdots e^{-\lambda}\frac{\lambda^{x_n}}{x_n!} \\ &= e^{-n\cdot\lambda}\,\frac{\lambda^{x_1+x_2+\cdots x_n}}{x_1! \cdot x_2! \cdots x_n!} \end{aligned}$$

Logarithmieren der Likelihoodfunktion liefert

$$l(\lambda) = \ln(L(\lambda)) = -n \cdot \lambda + (x_1 + \cdots + x_n) \cdot \ln(\lambda) - \ln(x_1! \cdot x_2! \cdots x_n!)$$

also

$$l'(\lambda) = -n + \left(\sum_{k=1}^{n} x_k\right) \cdot \frac{1}{\lambda}$$

Maximum-Likelihood–Schätzung

Beispiel

Einen Kandidaten $\widehat{\lambda}$ für ein Maximum erhalten wir durch Nullsetzen der Ableitung, also

$$0 = -n + \left(\sum_{k=1}^{n} x_k \right) \cdot \frac{1}{\widehat{\lambda}}$$

woraus folgt

$$\widehat{\lambda} = \frac{1}{n} \cdot \left(\sum_{k=1}^{n} x_k \right) = \overline{x}$$

Dabei handelt es sich in der Tat um ein Maximum (Betrachtung der zweiten Ableitung), und damit ist das arithmetische Mittel ein Maximum–Likelihoodschätzer für λ.

Maximum-Likelihood–Schätzung

Übung

Wir betrachten eine geometrisch verteilte Zufallsvariable X, mit unbekanntem Parameter $\theta = p$, d.h.

$$p(X = k) = \theta \cdot (1 - \theta)^{k-1} \qquad (k = 1, 2, 3, \ldots)$$

mit einem unbekannten Parameter $0 < \theta < 1$. Ferner betrachten wir n unabhängige Wiederholungen $X_1, \ldots, X_n$ des Zufallsexperiments mit den Ergebnissen $x_1, \ldots, x_n$.
Bestimmen Sie einen Maximum–Likelihood–Schätzer für den Parameter θ.

Maximum-Likelihood–Schätzung

Lösung:

Es gilt für die diskrete Dichte von X, (falls θ der Parameter von X ist):

$$f(x|\,\theta) = \theta \cdot (1-\theta)^{x-1} \qquad (x = 1, 2, 3, \ldots)$$

Damit hat die Likelihoodfunktion die Gestalt

$$L(\theta) = f(x_1, \ldots, x_n|\,\theta) = \prod_{i=1}^{n} f(x_i|\,\theta) = \prod_{i=1}^{n} \theta \cdot (1-\theta)^{x_i - 1} = \theta^n \cdot (1-\theta)^{\sum\limits_{i=1}^{n} x_i - n}$$

Logarithmieren ergibt

$$l(\theta) = \ln L(\theta) = n \cdot \ln(\theta) + \left(\sum_{i=1}^{n} x_i - n \right) \cdot \ln(1-\theta)$$

Maximum-Likelihood–Schätzung

Lösung:

Die Ableitung der Log–Likelhoodfunktion ist

$$l'(\theta) = \frac{n}{\theta} - \frac{\sum_{i=1}^{n} x_i - n}{1-\theta}$$

Notwendige Bedingung für ein Maximum ist

$$0 = \frac{n}{\theta} - \frac{\sum_{i=1}^{n} x_i - n}{1-\theta}$$

Maximum-Likelihood–Schätzung

Lösung:

Äquivalent dazu ist

$$\frac{n}{\theta} = \frac{\sum_{i=1}^{n} x_i - n}{1-\theta}$$

bzw.

$$n \cdot (1-\theta) = \theta \cdot \left(\sum_{i=1}^{n} x_i - n \right)$$

bzw.

$$n = \theta \cdot \left(\sum_{i=1}^{n} x_i \right)$$

Maximum-Likelihood–Schätzung

Lösung:

Die einzigen Lösung von $l'(\theta) = 0$ ist

$$\widehat{\theta} = \frac{n}{\sum_{i=1}^{n} x_i}$$

(also das Inverse des arithmetischen Mittels).
Da

$$l''(\theta) = -\frac{n}{\theta^2} - \frac{\sum_{i=1}^{n} x_i - n}{(1-\theta)^2}$$

ist also $l''(\widehat{\theta}) < 0$, und damit ist $\widehat{\theta}$ in der Tat ein Maximum–Likelihood Schätzer für $\theta = p$.

Maximum-Likelihood–Schätzung

Beispiel

Von der stetigen Zufallsvariable X ist bekannt, dass ihre Dichtefunktion $f : \mathbb{R} \longrightarrow \mathbb{R}$ die Gestalt

$$f(x|\theta) = \begin{cases} (\theta+1)\cdot(1-x)^{\theta} & \text{für} \quad 0 < x < 1 \\ 0 & \text{sonst} \end{cases}$$

mit einem unbekannten Parameter $\theta > 0$ hat. $X_1, \ldots, X_n$ des Zufallsexperiments X werden die Ergebnissen $x_1, \ldots, x_n$ beobachtet.

Maximum-Likelihood–Schätzung

Beispiel

Die Likelihoodfunktion hat die Gestalt

$$\begin{aligned} L(\theta) &= f(x_1, \ldots, x_n \mid \theta) \\ &= \prod_{i=1}^{n} f(x_i \mid \theta) \\ &= (\theta+1) \cdot (1-x_1)^{\theta} \cdots (\theta+1) \cdot (1-x_n)^{\theta} \\ &= (\theta+1)^n \cdot (1-x_1)^{\theta} \cdots (1-x_n)^{\theta} \end{aligned}$$

$$\begin{aligned} l(\theta) &= \ln L(\theta) = \ln\left((\theta+1)^n \cdot (1-x_1)^{\theta} \cdots (1-x_n)^{\theta}\right) \\ &= n \cdot \ln(\theta+1) + \theta \cdot \ln(1-x_1) + \cdots + \theta \cdot \ln(1-x_n) \\ &= n \cdot \ln(\theta+1) + \theta \cdot \sum_{i=1}^{n} \ln(1-x_i) \end{aligned}$$

Maximum-Likelihood–Schätzung

Beispiel

Die Ableitung hiervon ist

$$l'(\theta) = \frac{n}{\theta + 1} + \sum_{i=1}^{n} \ln(1 - x_i)$$

$$0 = \frac{n}{\theta + 1} + \sum_{i=1}^{n} \ln(1 - x_i)$$

$$\widehat{\theta} = -\frac{n}{\sum\limits_{i=1}^{n} \ln(1 - x_i)} - 1$$

Maximum-Likelihood–Schätzung

Beispiel

Dabei handelt es sich in der Tat um einen Maximum–Likelhood Schätzer, da

$$l''(\widehat{\theta}) = -\frac{n}{(\widehat{\theta}+1)^2} < 0$$

Maximum-Likelihood–Schätzung

Beispiel

Wir betrachten die n–fache Wiederholung einer normalverteilten Zufallsvariable X mit Parametern μ und σ^2, die wir schätzen wollen (also $\theta(\mu, \sigma^2)$ und n Realisationen $x_1, \ldots, x_n$. Dann gilt

$$\begin{aligned} L(\mu, \sigma) &= \frac{1}{\sqrt{2\pi}\sigma} \mathrm{e}^{-\frac{(x_1-\mu)^2}{2\sigma^2}} \cdots \frac{1}{\sqrt{2\pi}\,\sigma} \mathrm{e}^{-\frac{(x_1-\mu)^2}{2\sigma^2}} \\ &= \frac{1}{(\sqrt{2\pi}\sigma)n} \mathrm{e}^{-\sum\limits_{k=1}^{n} \frac{(x_k-\mu)^2}{2\sigma^2}} \end{aligned}$$

Logarithmieren der Likelihoodfunktion liefert

$$l(\mu, \sigma) = \ln(L(\mu, \sigma)) = -n \cdot \ln(\sqrt{2\pi} \cdot \sigma) + \left(-\sum_{k=1}^{n} \frac{(x_k - \mu)^2}{2\sigma^2} \right)$$

Maximum-Likelihood–Schätzung

Beispiel

Betrachtung der partiellen Ableitungen liefert

$$\widehat{\mu} = \frac{1}{n} \cdot \sum_{k=1}^{n} x_k$$

und

$$\widehat{\sigma} = \sqrt{\frac{1}{n} \cdot \sum_{k=1}^{n} (x_k - \widehat{\mu})^2}$$

Betrachtung der Hesse–Matrix zeigt, dass das tatsächlich Maximum–Likelhood–Schätzer für μ und σ sind.

Maximum-Likelihood–Schätzung

Übung

Von einer stetigen Zufallsvariable X ist bekannt, dass sie die Wahrscheinlichkeitsdichte

$$f(x|\alpha) = \begin{cases} \alpha \cdot \mathrm{e}^{-\alpha \cdot x} & (x \geq 0) \\ 0 & (x < 0) \end{cases}$$

mit einem unbekannten Parameter $\alpha > 0$ hat. Bei n unabhängige Wiederholungen $X_1, \ldots, X_n$ des Zufallsexperiments X werden die Ergebnissen $x_1, \ldots, x_n$ beobachtet.
Bestimmen Sie einen Maximum–Likelihood–Schätzer für den Parameter α.

Maximum-Likelihood–Schätzung

Lösung:

$$L(\alpha) = \prod_{i=1}^{n} f(x_i|\,\alpha) = \alpha^n \cdot \mathrm{e}^{-\sum\limits_{i=1}^{n} \alpha \cdot x_i}$$

$$l(\alpha) = n \cdot \ln(\alpha) - \alpha \cdot (x_1 + \cdots + x_n)$$

$$l'(\alpha) = \frac{n}{\alpha} - \sum_{i=1}^{n} x_i$$

$$\widehat{\alpha} = \frac{n}{\sum\limits_{i=1}^{n} x_i}$$

$$l''(\widehat{\alpha}) = -\frac{n}{\widehat{\alpha}^2} < 0$$