Stochastik für Informatiker

Dr. rer. nat. Johannes Riesterer

Statistik - Parameterschätzung

Erraten des Bereichs von Zufallsvariablen

Angenommen man findet einen Apparat (Fluxkompensator?), der zufällig Zahlen in einem Intervall $[0, \rho]$ ausgibt. Anhand von Beobachtungen der Zahlen möchte man ρ schätzen.

Figure: Quelle: forevergeek

Erraten des Bereichs von Zufallsvariablen - Modell

Wir machen die Annahme, dass alle Zahlen in dem Intervall gleich wahrscheinlich auftreten und nehmen n Stichproben $X_1, \cdots, X_n$. Einen Schätzer für ρ bezeichnen wir mit T_n.

Erraten des Bereichs von Zufallsvariablen - Schätzer 1

Eine einfache und einleuchtende Idee ist es, ρ durch die größte beobachtete Zahl zu schätzen, also $T_n^{max} := \max(X_1, \cdots, X_n)$.

Erraten des Bereichs von Zufallsvariablen - Schätzer 2

Da das Auftreten der Zahlen gleich wahrscheinlich ist, ist der Erwartungswert des Zufallsexperiments $\rho/2$. Unter Berufung auf das schwache Gesetz der Großen Zahlen erscheint der Schätzer $T_n^E := 2 \cdot \left(\frac{1}{n} \sum_{i=1}^n X_i \right)$ sehr plausibel.

Erraten des Bereichs von Zufallsvariablen - Vergleich

Welcher Schätzer ist besser und in welchem Sinn?

Erraten des Bereichs von Zufallsvariablen - Vergleich Konvergenz

$$P(|T_n^{max} - \rho| \geq \epsilon) = P(T_n^{max} \leq \rho - \epsilon)$$
$$= P(X_1 \leq \rho - \epsilon, \cdots, X_n \leq \rho - \epsilon) = (\frac{\rho - \epsilon}{\rho})^n \underset{n \to \infty}{\longrightarrow} 0$$

Erraten des Bereichs von Zufallsvariablen - Vergleich Konvergenz

$$P(|T_n^E - \rho| \geq \epsilon) = P(|\frac{1}{n}\sum X_i - \frac{\rho}{2}| \geq \frac{\epsilon}{2}) \underset{n \to \infty}{\longrightarrow} 0$$

(Gesetz der Großen Zahlen)

Erraten des Bereichs von Zufallsvariablen - Vergleich Erwartungswert

Da $P(T_n^{max} \leq c) = (\frac{c}{\rho})^n$ und $\frac{d}{dx}(\frac{c}{\rho})^n = \frac{n}{\rho^n} x^{n-1}$ und damit

$$\mathbb{E}(T_n^{max}) = \int_0^{\rho} x \frac{n}{\rho^n} x^{n-1}\, dx = \frac{n}{\rho^n} \int_0^{\rho} x^n\, dx = \frac{n}{n+1}\rho \xrightarrow[n \to \infty]{} \rho$$

Erraten des Bereichs von Zufallsvariablen - Vergleich Erwartungswert

$$\mathbb{E}(T_n^E) = \frac{2}{n} \sum \mathbb{E}(X_i) = \rho$$

Erraten des Bereichs von Zufallsvariablen - Vergleich Varianz

$$\mathbb{V}(T_n^{max}) = \frac{4}{n}\mathbb{V}(X_1) = \frac{4}{n\rho}\int_0^\rho (x - \frac{\rho}{2})^2 \, dx = \frac{\rho^2}{3n}$$

Erraten des Bereichs von Zufallsvariablen - Vergleich Varianz

$$\mathbb{V}(T_n^E) = \mathbb{E}((T_n^E)^2) - (\mathbb{E}(T_n^E))^2$$
$$= \int_0^\rho x^2 \frac{n}{\rho^n} x^{n-1} \, dx - (\frac{n\rho}{n+1})^2 = \frac{n\rho^2}{(n+1)^2(n+2)}$$

Statistisches Modell

Ein statistisches Modell ist ein Tripel $(\mathcal{X}, \mathcal{F}, P_\rho : \rho \in \Theta)$ bestehend aus einer σ-Algebra $\mathcal{F}$ über dem Grundraum $\mathcal{X}$ und einer indizierten Menge (mit mindestens zwei Elementen) von Maßen $\{P_\rho\}_{\rho\in\Theta}$. Für $\Theta \subset \mathbb{R}^n$ bezeichnet man es auch als parametrisiertes statistisches Modell.

Statistik

Sei $(\mathcal{X}, \mathcal{F}, P_\rho : \rho \in \Theta)$ ein statistisches Modell und $(\Sigma, \mathcal{G})$ eine σ-Algebra über Σ. Eine Zufallsvariable

$$X : \mathcal{X} \to \Sigma$$

wird Statistik für $(\mathcal{X}, \mathcal{F}, P_\rho : \rho \in \Theta)$ genannt.

Schätzer

Sei $(\mathcal{X}, \mathcal{F}, P_\rho : \rho \in \Theta)$ ein statistisches Modell, $(\Sigma, \mathcal{G})$ eine σ-Algebra über Σ und

$$\begin{aligned} \tau : \Theta &\to \Sigma \\ \rho &\mapsto \tau(\rho) \end{aligned}$$

eine Abbildung. Eine Statistik $T : \mathcal{X} \to \Sigma$ wird Schätzer für τ genannt.

Konsistenz

Eine Schätzfolge $T_n : \mathcal{X} \to \Sigma$ heißt konsistent, falls

$$P(|T_n - \tau(\rho)| \geq \epsilon) \xrightarrow[n \to \infty]{} 0 \text{ für alle } \epsilon > 0 \text{ und alle } \tau \in \theta$$

also $T_n \to \tau(\rho)$ für $n \to \infty$ (stochastisch).

Konsistenz

Ein Schätzer $T : \mathcal{X} \to \Sigma$ heißt Erwartungstreu (unbiased), falls

$$\mathbb{E}(T) = \tau(\rho) \text{ für alle } \rho \in \theta.$$

Andernfalls heißt $\mathbb{E}(T) - \tau(\rho)$ der Bias oder der systematische Fehler.

Stichprobenmittel und Stichprobenvarianz

Für $n \geq 2$ sei $\left(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), P_\rho^n := \prod(P_\rho)_i\right)$ das Produktmodell, $X : \mathbb{R} \to \mathbb{R}$ eine Zufallsvariable, $X_i(x_1, \cdot, x_n) = x_i$ die Projektion auf die i-te Koordinate und $m(\rho) = \mathbb{E}(X)$ so wie $v(\rho) = \mathbb{V}(X)$. Des weiteren seien X so wie X_i unabhängig und identisch verteilt.
Dann sind das Stichprobenmittel $M = \frac{1}{n}\sum_{i=1}^n X_i$ und die Stichprobenvarianz $V = \frac{1}{n-1}\sum_{i=1}^n (X_i - M)^2$ erwartungstreue Schätzer für m bzw. v.

Beweis

Sei $\rho \in \Theta$ fest. Wegen linearität des EW und u.i.v. ist
$\mathbb{E}(V) = \frac{1}{n}\sum_{i=1}^n \mathbb{E}(X_i) = m(\rho)$.

Beweis

Aus Linearität des EW und $\mathbb{E}(X_i - M) = 0$, u.i.v, stochastische Unabhängigkeit folgt

$$(n-1)\mathbb{E}(V) = \sum_{i=1}^{n} \mathbb{V}(X_i - M)$$

$$= n\mathbb{V}(\frac{n-1}{n}X_1 - \frac{1}{n}\sum_{j=2}^{n} X_j)$$

$$= n((\frac{n-1}{n})^2 + (n-1)\frac{1}{n^2})v(\rho) = (n-1)v(\rho)$$

Durch Teilen beider Seiten durch $n-1$ folgt die Behauptung.

Likelyhood Funktion

Für eine parametrisierte Dichte $p_\rho : \mathcal{X} \to [0,1]$ heißt die Funktion

$$p : \mathcal{X} \times \Theta \to [0,1]$$
$$p(x,\rho) := p_\rho(x)$$

die zugehörige Likelyhood Funktion und

$$p_x : \Theta \to [0,1]$$
$$p_x(\rho) := p(x,\rho)$$

die Liekelyhood Funktion zum Beobachtungswert $x \in \mathcal{X}$.
Im diskreten Fall setzten wir $p(x,\rho) := P_\rho(\{x\})$

Maximum Likelyhood Schätzer

Ein Schätzer $T : \mathcal{X} \to \Sigma$ heißt Maximum-Likelyhood-Schätzer, falls

$$p(x, T(x)) = \max_{\rho \in \Theta} p(x, \rho) \text{ für alle } x \in \mathcal{X}$$

also der Schätzwert $T(x)$ eine Maximalstelle der Funktion p_x auf Θ ist.

Erraten des Bereichs von Zufallsvariablen - Maximum Likelyhood-Schätzer

Die Likelyhood Funktion ist hier gegeben durch

$$p_x(\rho) = \begin{cases} \frac{1}{\rho^n} \text{ falls } x_1, \cdots, x_n \leq \rho \\ 0 \text{ sonst} \end{cases}$$

Somit ist der Schätzer $T_n^{max}(x) := \max(x_1, \cdots, x_n)$ der Maximum-Likelyhood-Schätzer.

Physikalische Messung - Maximum Likelyhood-Schätzer

Gegeben ist ein Sensor mit einer unbekannten Messgenauigkeit. Wir nehmen deshalb an, dass die Messung des Sensors einer Normalverteilung folgt, wobei der Erwartungswert (Messwert) und die Varianz (Streuung um Messwert) unbekannt sind. Wir machen n unabhängige Messungen und erhalten damit das Normalverteilte Produktmodell $\big(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \prod_{i=1}^n N(m, v) : m \in \mathbb{R}, v > 0\big)$.

Physikalische Messung - Maximum Likelyhood-Schätzer

Die Likelyhood-Funktion ist gegeben durch

$$p_x(\rho) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi v}} e^{-\frac{(x_i - m)^2}{2v}} = \frac{1}{\sqrt{2\pi v}} e^{-\sum_{i=1}^{n}\left(\frac{(x_i - m)^2}{2v}\right)}$$

mit $x = (x_1, \cdots, x_n)$ und $\rho = (m, v)$.

Physikalische Messung - Maximum Likelyhood-Schätzer

Um diesen Ausdruck zu maximieren, muss m so gewählt werden, dass die quadratische Fehlersumme $\sum_{i=1}^{n}(x_i - m)^2$ minimal wird. Das ist der Fall für $m = M(x) = \frac{1}{n}\sum_{i=1}^{n} x_i$.

Physikalische Messung - Maximum Likelyhood-Schätzer

Des weiteren muss v so gewählt werden, dass $\frac{1}{\sqrt{2\pi v}} e^{-\sum_{i=1}^{n}\left(\frac{(x_i - M(x))^2}{2v}\right)}$ maximal wird. Differenziert man den Logarithmus dieses Ausdrucks nach v, erhalten wir

$$
\begin{aligned}
& -\frac{d}{dv}\left(\frac{n}{2}\log(v) + \frac{1}{2v}\sum_{i=1}^{n}(x_i - M(x))^2\right) \\
= & -\frac{n}{2v} + \frac{1}{2v^2}\sum_{i=1}^{n}(x_i - M(x))^2
\end{aligned}
$$

Dieser Term ist maximal, falls der letzte Term verschwindet. Dies ist er Fall für $v = V(x) = \frac{1}{n}\sum_{i=1}^{n}(x_i - M(x))^2$

Beispiel Entscheidungsfindung - Problemstellung

Es werden $N = 10.000$ Orangen geliefert. Ist mehr als 5% der Ware verfault, kann der Käufer die Lieferung an den Verkäufer zurückgehen lassen. Alle Orangen anschauen ist wegen der Menge allerdings unmöglich.

Beispiel Entscheidungsfindung - Problemstellung

Der Käufer wählt eine Stichprobenmenge von $n = 50$ Orangen und möchte anhand einer Grenze c an faulen Orangen innerhalb dieser Stichprobe entscheiden, ob er die Lieferung zurückgehen lässt oder nicht. Er wendet also die folgende Entscheidungsregel an:

höchsten c Orangen faul $\Rightarrow$ Lieferung akzeptieren

mehr als c Orangen faul $\Rightarrow$ Lieferung zurückgehen lassen.

Beispiel Entscheidungsfindung - Problemstellung

Wie sollte c gewählt werden?

Beispiel Entscheidungsfindung - Modell

Man möchte auf der einen Seite, dass die Wahrscheinlichkeit dafür, dass die Lieferung mehr als 5% faule Orangen enthält nicht erkannt wird, klein ist. Auf der anderen Seite möchte man auch, dass die Wahrscheinlichkeit für einen "peinlichen" Fehler, dass die Lieferung zurück geht, obwohl sie weniger als 5% faule Orangen enthält, ebenfalls klein ist.

Beispiel Entscheidungsfindung - Problemstellung

Wie sollte c gewählt werden?

Beispiel Entscheidungsfindung - Modell

Man möchte auf der einen Seite, dass die Wahrscheinlichkeit dafür, dass die Lieferung mehr als 5% faule Orangen enthält nicht erkannt wird, klein ist. Auf der anderen Seite möchte man auch, dass die Wahrscheinlichkeit für einen "peinlichen" Fehler, dass die Lieferung zurück geht, obwohl sie weniger als 5% faule Orangen enthält, ebenfalls klein ist.

Beispiel Entscheidungsfindung - Modell

Wir modellieren die Anzahl k der faulen Orangen einer Stichprobe n aus der Grundmenge N als "hypergeometrische" Verteilung, wobei die Zahl ω der insgesamt faulen Orangen innerhalb N unbekannt is.

$$P_\omega(k) = \frac{\binom{\omega}{k} \cdot \binom{N-\omega}{n-k}}{\binom{N}{n}}$$

Beispiel Entscheidungsfindung - Modell

Ist x die Anzahl der faulen Orangen in einer Stichprobe $n = 50$, so lauten unsere vorigen Überlegung mit Hilfe des Modells: Wähle Schranke c für faule Orangen in Stichprobe so, dass für $\alpha = 0.025$

$$P_\omega(x : x > c) < \alpha \text{ für } \omega < 500 (= 5\% \text{ von } N = 10.000)$$
$$P_\omega(x : x > c) \text{ möglichst groß für } \omega > 500$$

gilt.

Beispiel Entscheidungsfindung - Modell

Die genaue Bestimmung von c ist auf den ersten Blick nicht so einfach. Wir werden das Problem auf eine einfachere Situation, einen sogenannten Neyman-Pearson-Test, zurückführen.