

# Wie funktioniert Maximum-Likelihood-Schätzung?

HENRY HAUSTEIN

## Vom Speziellen...

Aus einem Zufallszahlengenerator  $X$  erhalten wir die folgenden 5 Zufallszahlen: 3, 4, 4, 4, 5. Wir wissen außerdem, dass die Zufallszahlen exponentialverteilt und unabhängig sind, aber wir kennen den Parameter der Verteilung nicht. Wie können wir diesen bestimmen?

Schauen wir uns zuerst an, wie *wahrscheinlich*<sup>1</sup> es ist, eine der obigen Zufallszahlen erhalten zu haben, abhängig vom Parameter der Exponentialverteilung

$$\mathbb{P}(X = 3) = \lambda \cdot \exp(-3 \cdot \lambda)$$

$$\mathbb{P}(X = 4) = \lambda \cdot \exp(-4 \cdot \lambda)$$

$$\mathbb{P}(X = 4) = \lambda \cdot \exp(-4 \cdot \lambda)$$

$$\mathbb{P}(X = 4) = \lambda \cdot \exp(-4 \cdot \lambda)$$

$$\mathbb{P}(X = 5) = \lambda \cdot \exp(-5 \cdot \lambda)$$

Wir sollten also den Parameter  $\lambda$  so wählen, dass die *Wahrscheinlichkeiten* für alle diese Ereignisse möglichst groß werden.

Wir könnten uns anschauen, wie *wahrscheinlich* es ist, dass wir alle Zufallszahlen sehen: Dazu nutzen wir die Eigenschaft, dass die Zufallszahlen unabhängig sind und erinnern uns an

$$\mathbb{P}(A \text{ und } B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$$

Wir erweitern diese Idee auf unsere 5 Zufallszahlen und erhalten:

$$\mathbb{P}(\text{alle obigen Zufallszahlen gesehen}) = \mathbb{P}(X = 3) \cdot \mathbb{P}(X = 4) \cdot \mathbb{P}(X = 4) \cdot \mathbb{P}(X = 4) \cdot \mathbb{P}(X = 5)$$

Wir nennen diesen Term die Likelihood-Funktion  $L(\lambda)$ . Diese gibt an, wie *wahrscheinlich* (bzw. plausibel) es ist, dass wir die obigen Zufallszahlen bekommen haben, in Abhängigkeit des Parameters  $\lambda$ .

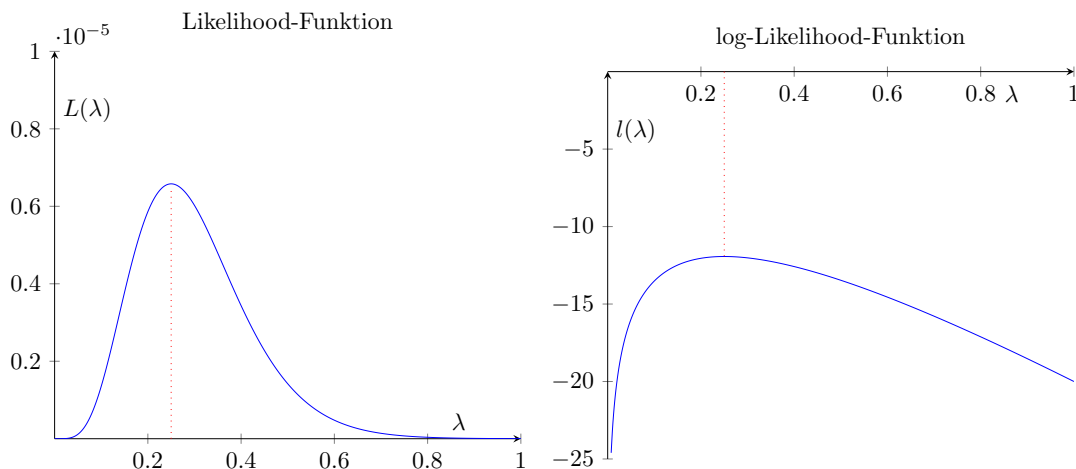
$$\begin{aligned} L(\lambda) &= (\lambda \cdot \exp(-3\lambda)) \cdot (\lambda \cdot \exp(-4\lambda)) \cdot (\lambda \cdot \exp(-4\lambda)) \cdot (\lambda \cdot \exp(-4\lambda)) \cdot (\lambda \cdot \exp(-5\lambda)) \\ &= \lambda^5 \cdot \exp(-3\lambda) \exp(-4\lambda) \exp(-4\lambda) \exp(-4\lambda) \exp(-5\lambda) \\ &= \lambda^5 \cdot \exp(-3\lambda + -4\lambda + -4\lambda + -4\lambda + -5\lambda) \\ &= \lambda^5 \cdot \exp(-20\lambda) \end{aligned}$$

Wir wollen die Likelihood-Funktion maximieren, dafür könnten wir  $L(\lambda)$  nach  $\lambda$  ableiten und nullsetzen. Aber für einige Verteilungen ist das ziemlich aufwendig, da wir die Produktregel für Ableitungen benutzen

---

<sup>1</sup>Die Exponentialverteilung ist eine stetige Verteilung, das heißt die Wahrscheinlichkeit, eine bestimmte Zahl zu erhalten, ist per Definition 0. Statt von Wahrscheinlichkeiten wäre es besser von Plausibilitäten zu sprechen, aber das macht den Text weniger gut lesbar.

müssten. Wir greifen deswegen auf einen kleinen Trick zurück und bilden  $l(\lambda) = \log(L(\lambda))$ . Diese Funktion heißt log-Likelihood-Funktion und hat die Eigenschaft, dass das Maximum dieser genau beim selben  $\lambda$  wie bei  $L(\lambda)$  ist.



Die log-Likelihood-Funktion sieht so aus:

$$\begin{aligned} l(\lambda) &= \log(\lambda^5 \cdot \exp(-20\lambda)) \\ &= \log(\lambda^5) + \log(\exp(-20\lambda)) \\ &= 5 \cdot \log(\lambda) - 20\lambda \end{aligned}$$

Diese Funktion ist viel leichter abzuleiten, weil der Logarithmus alle Produkte in Summen verwandelt.

Maximieren von  $l(\lambda)$  liefert:

$$\begin{aligned} \frac{\partial l}{\partial \lambda} &= 5 \cdot \frac{1}{\lambda} - 20 \stackrel{!}{=} 0 \\ 20 &= \frac{5}{\lambda} \\ \lambda &= \frac{5}{20} = \frac{1}{4} \end{aligned}$$

Damit haben wir eine Schätzung für den Parameter  $\lambda$  gefunden. Man sollte aber noch bedenken, dass dies der plausibelste Parameter ist - wir wissen nicht mit Sicherheit, welchen wahren Wert der Parameter in unserem Zufallsgenerator hat.

## ... ins Allgemeine

Wir betrachten nun den Fall, dass wir nicht mehr nur 5 Zufallszahlen, sondern  $n$  Zufallszahlen  $x_1, \dots, x_n$  haben. Die *Wahrscheinlichkeit*, dass unser Zufallszahlengenerator nun eine dieser Zahlen ausgibt, ist

$$\begin{aligned} \mathbb{P}(X = x_1) &= \lambda \cdot \exp(-\lambda x_1) \\ \mathbb{P}(X = x_2) &= \lambda \cdot \exp(-\lambda x_2) \\ &\vdots \\ \mathbb{P}(X = x_n) &= \lambda \cdot \exp(-\lambda x_n) \end{aligned}$$

Die *Wahrscheinlichkeit*, dass wir alle  $n$  Zufallszahlen sehen, ist

$$\begin{aligned}\mathbb{P}(\text{alle } n \text{ Zufallszahlen gesehen}) &= \mathbb{P}(X = x_1) \cdot \mathbb{P}(X = x_2) \cdot \dots \cdot \mathbb{P}(X = x_n) \\ &= \prod_{i=1}^n \mathbb{P}(X = x_i)\end{aligned}$$

Das ist die Likelihood-Funktion  $L(\lambda)$ ; ausgeschrieben ist sie

$$\begin{aligned}L(\lambda) &= \prod_{i=1}^n \mathbb{P}(X = x_i) \\ &= \prod_{i=1}^n \lambda \cdot \exp(-\lambda x_i) \\ &= \lambda^n \cdot \prod_{i=1}^n \exp(-\lambda x_i) \\ &= \lambda^n \cdot \exp\left(\sum_{i=1}^n -\lambda x_i\right) \\ &= \lambda^n \cdot \exp\left(-\lambda \sum_{i=1}^n x_i\right)\end{aligned}$$

Diese Funktion wollen wir wieder maximieren, aber die Ableitung erweist sich hier als schwierig. Deswegen nutzen wir den Trick mit dem logarithmieren und erhalten die log-Likelihood-Funktion  $l(\lambda)$ :

$$\begin{aligned}l(\lambda) &= \log(L(\lambda)) \\ &= \log\left(\lambda^n \cdot \exp\left(-\lambda \sum_{i=1}^n x_i\right)\right) \\ &= \log(\lambda^n) + \log\left(\exp\left(-\lambda \sum_{i=1}^n x_i\right)\right) \\ &= n \log(\lambda) - \lambda \sum_{i=1}^n x_i\end{aligned}$$

Ableiten und nullsetzen liefert

$$\begin{aligned}\frac{\partial l}{\partial \lambda} &= \frac{n}{\lambda} - \sum_{i=1}^n x_i \stackrel{!}{=} 0 \\ \sum_{i=1}^n x_i &= \frac{n}{\lambda} \\ \lambda &= \frac{n}{\sum_{i=1}^n x_i} = \frac{1}{\bar{x}}\end{aligned}$$

Damit haben wir den allgemeinen Schätzer der Exponentialverteilung gefunden.