

9. Übungsblatt zur Vorlesung Datenanalyse mit R

Aufgabe 1: Zufallszahlen $x_1, x_2, \dots$ heissen exponential-verteilt mit Parameter $\lambda > 0$, wenn für alle $x \geq 0$

$$\text{Prob}[x_i \in [x, x + dx)] = \lambda e^{-\lambda x} dx$$

gilt und die Wahrscheinlichkeit für negative Zahlen 0 ist, also $\text{Prob}[x_i \in [x, x + dx)] = 0$ falls $x < 0$.

- a) Es sei x eine mit Parameter $\lambda > 0$ exponential-verteilte Zufallszahl. Zeigen Sie, dass für den Erwartungswert von x gilt: $E[x] = \frac{1}{\lambda}$.

Nehmen wir jetzt an, dass die Zufallszahlen $x_1, x_2, \dots, x_n$ gegeben sind, und dass wir wissen, dass sie durch eine Exponentialverteilung mit Parameter λ erzeugt worden sind, wir kennen jedoch den konkreten Wert für λ nicht. Wir möchten das λ mit Hilfe der Maximum Likelihood Methode schätzen:

- b) Geben Sie die Likelihood Funktion $L = L(\lambda)$ an. Die Funktion hängt neben λ auch von den gegebenen Zufallszahlen $x_1, \dots, x_n$ ab, was wir aber nicht im Argument explizit machen, da wir die Funktion durch Variation von λ maximieren möchten, bei festem $x_1, \dots, x_n$. Berechnen Sie ebenfalls den Logarithmus $\log L(\lambda)$ und vereinfachen Sie das Resultat soweit wie möglich. Da der Logarithmus eine streng monoton steigende Funktion ist, ist das Problem

$$L(\lambda) \xrightarrow{!} \max$$

äquivalent zu

$$\log L(\lambda) \xrightarrow{!} \max,$$

aber rechentechnisch und numerisch ist es einfacher, den Logarithmus der Likelihood Funktion $\log L(\lambda)$ zu maximieren.

- c) Zeigen Sie, dass der Maximum Likelihood Schätzer für λ gegeben ist durch

$$\hat{\lambda}_{\text{ML}}(x_1, \dots, x_n) = \frac{1}{\frac{1}{n} \sum_{i=1}^n x_i}.$$

Bestimmen Sie dazu das Maximum von $\log L(\lambda)$.

Aufgabe 2: Wir wollen die Resultate aus Aufgabe 1 mit einer geeigneten R-Simulation überprüfen:

- a) Die Exponential-Verteilung gehört zu den in R eingebauten Verteilungen. Erzeugen Sie $n = 1000$ mit Parameter $\lambda = 4$ exponential-verteilte Zufallszahlen und speichern Sie sie in dem Vektor **x**. Erstellen Sie einen Plot und ein Histogramm von **x**. Die Syntax für die Exponential-Verteilung können Sie in dem file **W'keitsverteilungen-in-R.pdf** finden, es befindet sich auf der Vorlesungshomepage.
- b) Berechnen Sie den Wert des Schätzers $\hat{\lambda}_{ML}$. Diese Zahl sollte dann in der Nähe von 4 liegen.
- c) Legen Sie den Vektor

$$\lambda = (0.1, 0.2, 0.3, \dots, 9.8, 9.9, 10.0) =: (\lambda_1, \dots, \lambda_{100})$$

an. Berechnen Sie dann numerisch den Wert der log-Likelihood-Funktion für alle $\lambda_1, \dots, \lambda_{100}$ und speichern Sie diese Zahlen in dem Vektor

$$\log L = (\log L[1], \log L[2], \dots, \log L[100]).$$

Plotten Sie schliesslich $\log L$ als Funktion von λ und verifizieren Sie, dass das Maximum von $\log L$ an der Stelle $\hat{\lambda}_{ML}$ angenommen wird.

- d) Wiederholen Sie die Schritte (a)-(c) noch einmal für verschiedene Werte von n , etwa $n = 2000, 4000, 8000$ und $n = 16000$.