Prof. Dr. Evgeny Spodarev
Dipl.-Math. Stefan Roth

SS 2015
14. April 2015

Stochastik I - Übungsblatt 1

Abgabe: Dienstag, 21. April vor Beginn der Übung.

Hinweise zu den R-Aufgaben:

- Den Namen beider abgebenden Studenten auf jedes Blatt der Ausgabe drucken!
- Immer Quelltext und Ausgabe zusammen abgeben (nicht auf getrennten Blättern). Bei Aufgaben mit Grafikausgabe Quelltext und Plots abgeben.

Aufgabe 1 (4 Punkte)

Es seien $X_1, \ldots, X_n$ i.i.d. Zufallsvariablen auf einem gemeinsamen Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P)$. Die Funktion $\hat{F}_n : \Omega \times \mathbb{R} \to [0,1]$ definiert durch

$$\hat{F}_n(x, \omega) = \hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{(-\infty, x]}(X_i), \quad x \in \mathbb{R},$$

heißt empirische Verteilungsfunktion der Zufallsstichprobe $(X_1, \ldots, X_n)$. Zeige, dass $\hat{F}_n(x)$ für jedes $\omega \in \Omega$ eine Verteilungsfunktion ist.

Aufgabe 2 (8 Punkte)

Seien $X_1, \ldots, X_n$ unabhängig und identisch gleichverteilt auf dem Intervall $(0,1)$. Für $\lambda > 0$ definiere die Zufallsvariablen $Y_1, \ldots, Y_n$ durch

$$Y_k = -\frac{1}{\lambda} \log(1 - X_k), \quad k = 1, \ldots, n.$$

(a) Zeige, dass $Y_1, \ldots, Y_n$ unabhängig und identisch exponentialverteilt sind mit Paramter λ.

(b) Schreibe eine Funktion `MyExp(n, lambda)` in **R**, mit Eingabeparametern $n \in \mathbb{N}$ und lambda > 0, die n unabhängige Realisierungen einer exponentialverteilten Zufallsvariable mit Parameter lambda erzeugt. Verwende dabei nicht die **R**-Funktion `rexp()`!

(c) Verwende die Funktion aus (b) um 10000 unabhängige Realisierungen einer Exp(1)-Verteilung zu erzeugen und plotte ein Histogramm zusammen mit der Dichte der Exp(1)-Verteilung in ein Schaubild. Erzeuge außerdem einen Boxplot der Realisierungen. Speichere die Schaubilder jeweils in einer PDF-Datei.

(d) Plotte die empirische Verteilungsfunktion[1] der Stichprobe aus (c) zusammen mit der Verteilungsfunktion der Exp(1)-Verteilung in ein Schaubild.

[1] Hinweis: Verwende die Funktion ecdf. Beachte dabei auch die Beispiele in der **R**-Manual.

Aufgabe 3 (10 Punkte)

Für diese Aufgabe soll der Datensatz "microtus.data" über Feldmäuse mit **R** untersucht werden. Er beschreibt unterschiedliche Merkmale der Feldmausarten Microtus Multiplex und Microtus Subterraneus. Die Gruppe `unknown` enthält Tiere, die den beiden schwer zu unterscheidenden Feldmausarten nicht eindeutig zugeordnet werden konnten.

(a) Lies zunächst den Datensatz mit `read.table()` ein. Die Spalte `Foramen` enthält die Längen (in 1/1000 mm) des sogenannten fossa incisiva, einem kleinen Spalt im Unterkiefer hinter den vorderen Zähnen. Gib die summary-Statistiken dieser Spalte aus und plotte ein Histogramm mit den relativen Häufigkeiten der Werte.

(b) Extrahiere nun die Spalten `Length` und `Height` für den Microtus Subterraneus und erstelle für beide Merkmale jeweils einen Boxplot. Gib dabei beide Boxplots in einem gemeinsamen Fenster aus und speichere die Grafik als pdf.

(c) Zeichne einen Plot, der das Merkmal `Length` aus "microtus.data" in Abhängigkeit von `Height` darstellt. Zeichne eine Gerade als möglichst gute Näherung ein. Wie kann man die Gerade interpretieren?

(d) Schreibe eine Funktion, die einen Data-Frame und eine Zahl b als Eingabe hat und zurückgibt, wieviele der Feldmäuse länger als b sind, d.h. bei denen `Length`$> b$ gilt. Wende die Funktion auf den Microtus Multiplex und $b = 2400$ an.

Aufgabe 4 (6 Punkte)

Es bezeichne $\bar{x}_n$ das Stichprobenmittel und x_{med} den Median einer Stichprobe $(x_1, \ldots, x_n)$. Zeige folgende Identitäten:

(a) $\bar{x}_n = \underset{x\in\mathbb{R}}{\operatorname{argmin}} \sum_{i=1}^{n} (x_i - x)^2$

(b) $x_{\text{med}} = \underset{x\in\mathbb{R}}{\operatorname{argmin}} \sum_{i=1}^{n} |x_i - x|,$

wobei in (b) davon ausgegangen werden kann, dass die argmin-Funktion immer einen Wert liefert, auch wenn das Minimum nicht eindeutig ist.