Prof. Dr. Volker Schmidt
Björn Kriesche, M.Sc.

SS 2014
20. 05. 2014

Stochastik I - Übungsblatt 4

Abgabe am 27. 05. **vor** Beginn der Übung

Aufgabe 1 (2+2+2+4 Punkte)

(a) Es sei $(X_1, \dots, X_n)$ eine Zufallsstichprobe. Berechne die benötigten Momente und darauf basierende Punktschätzer für den Parameter θ mit der Momentenmethode, falls

(i) $X_i \sim \text{Geo}(\theta)$ mit $\theta \in (0, 1)$,

(ii) $X_i \sim \text{Exp}(\theta)$ mit $\theta > 0$,

(iii) $X_i \sim \text{U}(-\theta, \theta)$ mit $\theta > 0$.

(b) Ein Herrsteller von Füllmaschinen hat ein neues Gerät entwickelt um Gläser mit einer Sollfüllmenge von 100ml zu befüllen. Die Maschine arbeitet jedoch etwas ungenau, sodass tatsächlich immer ein paar ml zu viel abgefüllt werden. Die zusätzlich abgefüllte Menge kann als eine Zufallsvariable X betrachtet werden, deren Verteilung man gern ermitteln möchte. Aus Erfahrung weiß man, dass nur die folgenden drei Familien von Verteilungen in Frage kommen:

- $X \sim \text{U}(a - 2, a + 2)$ mit $a \in \mathbb{R}$
- $X \sim \text{N}(\mu, \sigma^2)$ mit $\mu \in \mathbb{R}$ und $\sigma^2 > 0$
- $X \sim \Gamma(b, p)$ mit $b, p > 0$

Zu Testzwecken wurden 50 Gläser befüllt und die tatsächlich zu viel abgefüllten Mengen in der Datei **fuellstand.txt** (siehe Homepage) gespeichert. Plotte ein Histogramm der Daten zusammen mit den Dichten der drei möglichen Verteilungen (in verschiedenen Farben) in eine Abbildung. Verwende die Momentenmethode um die benötigten Parameter aus den gemessenen Mengen zu schätzen. Entscheide welche Verteilung am besten passt.

Aufgabe 2 (2+5 Punkte)

(a) Betrachte die Betafunktion $B : \mathbb{R}^2 \to \mathbb{R}$, die definiert ist als

$$B(\alpha, \beta) = \int_0^1 t^{\alpha-1}(1 - t)^{\beta-1} dt.$$

Zeige, dass

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)} \quad \forall \alpha, \beta > 0,$$

wobei $\Gamma(\cdot)$ die in der Vorlesung eingeführte Gammafunktion bezeichnet.

(b) Es seien X, Y unabhängige Zufallsvariablen mit $X \sim \Gamma(\lambda, \alpha)$ und $Y \sim \Gamma(\lambda, \beta)$, wobei $\alpha, \beta, \lambda > 0$. Dann nennt man die Zufallsvariable $Z = \dfrac{X}{X + Y}$ betaverteilt mit Parametern α und β (Schreibweise $Z \sim B(\alpha, \beta)$). Man kann zeigen, dass die Dichte von Z folgende Form hat:

$$f_Z(t) = \frac{1}{B(\alpha, \beta)} t^{\alpha-1}(1 - t)^{\beta-1} \mathbb{1}_{[0,1]}(t).$$

Bestimme den Momentenschätzer $(\hat{\alpha}, \hat{\beta})$ für den Parametervektor (α, β).

Aufgabe 3 (2+1,5+3+1,5 Punkte)

Sei $(X_1, \dots, X_n)$ eine Zufallsstichprobe mit Verteilungsfunktion F, d.h. $X_i \sim F$ für $i = 1, \dots, n$. Die Abbildung $\hat{F}_n : \mathbb{R} \times \Omega \to [0, 1]$ mit

$$\hat{F}_n(x, \omega) = \frac{\#\{i : 1 \leq i \leq n, X_i(\omega) \leq x\}}{n}$$

heißt empirische Verteilungsfunktion der Zufallsstichprobe $(X_1, \dots, X_n)$. Zeige, dass für alle $x, y \in \mathbb{R}$ mit $x < y$ folgende Aussagen gelten:

(a) $n\hat{F}_n(x)$ ist binomialverteilt mit Parametern n und $p = F(x)$.

(b) $\mathbb{E}(\hat{F}_n(x)) = F(x)$ und $\mathrm{Var}(\hat{F}_n(x)) = \dfrac{F(x)(1 - F(x))}{n}$.

(c) $Cov(\hat{F}_n(x), \hat{F}_n(y)) = \dfrac{F(x)(1 - F(y))}{n}$.

(d) $\hat{F}_n(x) \xrightarrow{f.s.} F(x)$ für $n \to \infty$.

Aufgabe 4 (3 Punkte + 2 Bonuspunkte)

Schreibe eine eigene Funktion (ohne den Befehl ecdf zu verwenden), die die empirische Verteilungsfunktion einer übergebenen Stichprobe bestimmt. Generiere 20 Realsiserungen einer $\Gamma(1, 2.5)$-verteilten Zufallsvariable und plotte die daraus geschätzte empirische Verteilungsfunktion gemeinsam mit der theoretischen Verteilungsfunktion in eine Abbildung. Verwende verschiedene Linientypen um die Funktionen zu unterscheiden.

Hinweis: Es reicht aus, wenn die empirische Verteilungsfunktion korrekt berechnet wird, **R** die entstehenden Sprungstellen beim Plotten jedoch verbindet.

Bonus: Versuche die empirische Verteilungsfunktion so zu implementieren, dass **R** sie tatsächlich als Treppenfunktion mit Sprungstellen zeichnet, wobei für jedes Lininesegment ein Punkt am linken Ende die Rechtsstetigkeit der Funktion symbolisiert (so wie bei Benutzung der **R**-Funktion ecdf).