ulm university universität uulm

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

SS 2015
22. April 2015

Stochastik I - Übungsblatt 2

Abgabe: Dienstag, 28. 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 (3 Punkte)

Seien $X_1, \dots, X_n$ unabhängige und identisch verteilte Zufallsvariablen mit $\mathbb{E}(X_1^2) < \infty$. Zeige:

$$S_n^2 := \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2 \xrightarrow{f.s.} \mathsf{Var}(X_1) \quad (n \to \infty),$$

wobei $\bar{X}_n := \frac{1}{n} \sum_{i=1}^{n} X_i$.

Aufgabe 2 (3 + 3 + 3 Punkte)

Führe die in Teil (a) und (b) beschriebenen Experimente jeweils 100 Mal durch und veranschauliche die Ergebnisse für die unterschiedlichen Versuche jeweils in einem Histogramm. Plotte die Histogramme für Teil (a) bzw. (b) dabei jeweils mit gleicher Skala und ordne sie in 2 Zeilen und 2 Spalten an mithilfe von `par(mfrow=...)`. Beschrifte die Histogramme jeweils sinnvoll (Parameter `main`).

(a) Erzeuge jeweils 1000 Realisierungen von unabhängigen auf dem Intervall $[-10, 10]$ gleichverteilten Zufallsvariablen. Berechne in jedem Durchlauf den Mittelwert und den Median. Welche Statistik nähert den Erwartungswert (bei Deiner Simulation) besser an?

(b) Erweitere die Daten um jeweils 20 Ausreißer, die normalverteilt sind mit Erwartungswert 20 und Varianz 100. Führe mit den neuen Daten noch einmal einen Vergleich der beiden Statistiken durch. Vergleiche die beiden Größen auch mit denen aus Teil (a) (achte darauf, dass Du die Werte aus der (a) übernimmst und keine neuen erzeugst[1]). Interpretiere Deine Ergebnisse kurz.

(c) Ist der Median immer eine gute Näherung für den Erwartungswert? Gib eine kurze Begründung dafür an, bzw. finde ein Gegenbeispiel.

Aufgabe 3 (2 + 6 Punkte)

Die Anzahl der Mitarbeiter bei den 15 Unternehmen von Fantasiestadt ist gegeben durch

150, 20, 3800, 405, 350, 380, 290, 5000, 2600, 25, 40, 3200, 320, 400, 10

[1] Dafür ist es sinnvoll, (a) und (b) in einer gemeinsamen for-Schleife zu implementieren.

(a) Plotte die Lorenz-Kurve der Mitarbeiterzahl[2].

(b) Berechne für die Mitarbeiterzahlen den Gini-Koeffizienten, die Konzentrationsrate der 4 größten Unternehmen in Fantasiestadt und den Herfindahl-Index. Nutze dafür nur grundlegende **R**-Funktionen wie `sum()` oder `mean()`.

Aufgabe 4 (3 + 3 + 3 Punkte)

Bestimme für die Verteilungen in (a) und (b) die Lorenz-Kurve jeweils exakt und plotte sie gemeinsam mit der 1. Winkelhalbierenden in ein Schaubild. Plotte in (c) ebenfalls die Lorenz-Kurve. Bestimme die Werte dazu mittels numerischer Integration[3].

(a) $X \sim \mathsf{U}(0, \theta)$, $\theta > 0$

(b) $X \sim \mathsf{Exp}(\lambda)$, $\lambda > 0$

(c) $X \sim \mathsf{N}(3, 1)$

Aufgabe 5 (8 + 4 Punkte)

Ein Hersteller von Füllmaschinen stellt 3 neue Geräte vor. Zu Demonstrationszwecken hat er eine Studie durchgeführt, in der jede Maschine 50 Gläser gefüllt hat. Die gemessenen Füllmengen sind in der Datei fuellstand.txt (Angaben in ml) auf der Homepage verzeichnet, wobei 100 ml die Sollfüllmenge darstellt.

(a) Modelliere die Füllmengen X_i der einzelnen Geräte jeweils als (unabhängige) Zufallsvariablen mit einer möglichst gut passenden Verteilung, wobei du dich auf folgende Möglichkeiten beschränken kannst:

- $X_i - c \sim \mathsf{Exp}(\lambda)$, $c \in \mathbb{R}$, $\lambda > 0$
- $X_i \sim \mathsf{U}(a, b)$, $a < b$
- $X_i \sim \mathsf{N}(\mu, \sigma^2)$, $\mu \in \mathbb{R}$, $\sigma^2 > 0$

Orientiere dich dabei an der Stichprobenvarianz und am Stichprobenmittel, um die Parameter der Verteilung zu schätzen und entscheide dann optisch, welche Dichte dem Histogramm der Stichprobe am besten entspricht. Plotte für jede Maschine jeweils das Histogramm und die 3 Dichten in ein gemeinsames Schaubild. Skaliere dafür die Dichten (bzw. Histogramme) zueinander passend in y-Richtung.

(b) Plotte für jedes der 3 Geräte jeweils die von dir in (a) ausgewählte am besten passende Verteilungsfunktion und die empirische Verteilungsfunktion in ein gemeinsames Schaubild.

[2] Hierfür kann das Paket "ineq" verwendet werden.

[3] Verwende dazu die Methode "integrate" in **R**. Beachte bitte, dass $F_X^{-1}(0) = -\infty$ und $F_X^{-1}(1) = \infty$. Darum sollten die Integrationsgrenzen geeignet gewählt werden.