Dr. H. Lanzinger
C. Hering

Winter 2010/11
03.02.2011
Blatt 11

Stochastik III

(Abgabe: Do., 10.02.2011, 14:15 Uhr, vor den Übungen)

1. Gegeben sei eine i.i.d. Stichprobe $X_1, \ldots, X_n$ zum Merkmal $X \sim F$. Wir betrachten nun den Schätzer aus Blatt 10, Aufgabe 1:

$$\hat{\theta} := \overline{X_n}^2 := \Big(\tfrac{1}{n}\sum_{i=1}^{n} X_i\Big)^2.$$

Zeige, dass die Bootstrap-Varianz dieses Schätzers gegeben ist durch

$$\widehat{\mathrm{Var}}(\hat{\theta})_{\mathrm{boot}} = \frac{\hat{\mu}_4 - \hat{\mu}_2^2}{n^3} + 4\frac{\overline{X_n}\hat{\mu}_3}{n^2} + 4\frac{\overline{X_n}^2\hat{\mu}_2}{n},$$

wobei für $p = 2, \ldots, 4$ jeweils gilt

$$\hat{\mu}_p := \tfrac{1}{n}\sum_{i=1}^{n} \big(X_i - \overline{X_n}\big)^p.$$

(5 Punkte)

2. Ein Hersteller von Speziallampen, z.B. für Projektoren, gibt für ein bestimmtes Leuchtmittel eine Farbtemperatur von 5800 Kelvin an. Ist die tatsächliche Farbtemperatur höher, enthält das Licht der Lampe einen höheren Blau-Anteil und wirkt zu "kühl". Ist die Farbtemperatur niedriger, so wirkt das Licht gelblich. Zum Vergleich: Tageslicht hat etwa ein Farbtemperatur von 5600 Kelvin, eine Halogenlampe ca. 3400 Kelvin. Um die Einhaltung der Spezifikation zu prüfen, werden 15 Leuchtmittel zufällig ausgewählt und getestet. Die Ergebnisse finden sich im Datensatz `lampen.dat`.

   (a) Untersuche die Daten auf Normalverteilung.

   (b) Berechne basierende auf der Statistik

   $$T = \sqrt{n}\frac{\overline{X_n} - \mu}{\hat{\sigma}}$$

   ein 95%-Konfidenzintervall für den Mittelwert μ. Es soll hier von normalverteilten Daten ausgegangen werden.

   (c) Verzichtet man auf die Normalverteilungsannahme, so ist es bei der vorigen Teilaufgabe nicht mehr ohne Weiteres möglich die Verteilung der Statistik T anzugeben, lediglich asymptotisch können wir dann noch eine Verteilungsaussage treffen. Die Quantile, die für das Konfidenzintervall benötigt werden können deshalb nicht mehr angegeben werden. Verwende einen Bootstrap mit 1000 Bootstrapreplikationen, um ein Konfidenzintervall anzugeben, d.h. verwende den Bootstrap um näherungsweise Quantile $\hat{t}_{\alpha/2}$ und $\hat{t}_{1-\alpha/2}$ der wahren Verteilung von T zu finden. α soll hier wieder 95% sein.

   (d) Vergleiche die beiden Konfidenzintervalle aus b) und c)

(2 + 2 + 3 + 2 Punkte)

3. In dieser Aufgabe soll der Datensatz `norm.dat` auf Normalität untersucht werden. Dazu verwenden wir im Folgenden die sog. Kolmogorov-Smirnow Test-Statistik:

$$\sqrt{n} \sup_{x \in \mathbb{R}} \left| \hat{F}_n(x) - F_0(x) \right|,$$

wobei F_0 die Verteilung unter der Nullhypothese ist. Ein Test basierend auf dieser Test-Statistik ist in `R` als `ks.test` implementiert.

(a) Erstelle Diagnoseplots um die Standardnormalverteilungsannahme der Daten zu überprüfen und kommentiere Deine Ergbnisse.

(b) Verwende einen Kolmogorov-Smirnow-Test, um die Standardnormalverteilungsannahme der Daten zu überprüfen.

(c) Schätze den Mittelwert und die Standardabweichung der Daten und standardisiere die Daten entsprechend. Führe anschließend wieder einen Kolmogorov-Smirnow-Test mit den standardisierten Daten durch. Was fällt auf? Wie ist dies zu erklären?

(d) Verwende einen sog. parametrischen Bootstrap um einen ungefähren p-Wert des Tests aus der letzten Teilaufgabe zu erhalten. Erzeuge dazu 5000 Bootstrapreplikationen der Länge des ursprünglichen Datensatzes, wobei jede Stichprobe normalverteilt mit dem geschätzten Mittelwert und der geschätzten Standardabweichung aus der letzten Teilaufgabe sein soll. Verwende diese simulierten Daten, um den Bootstrap durchzuführen.

(2 + 1 + 3 + 4 Punkte)