10. Übung am 30. Mai 2022

UV Angewandte Statistik (405.170)

Übungsaufgabe 55. Denken Sie den Beweis von Satz 5.11 sowie die darauffolgende Bemerkung Zeile für Zeile durch.

Übungsaufgabe 56. Verwenden Sie Satz 4.20 um ausgehend von einer Stichprobe $X_1, \dots, X_n \sim F$ (mit F stetig) einen Test für $H_0 : F = F_0$ versus $H_1 : F \neq F_0$ herzuleiten. Wie kann die Power dieses Tests mittels Simulationen überprüft werden?

Übungsaufgabe 57. $X_1, \dots, X_n$ sei eine Stichprobe von $X \sim \mathcal{N}(\mu, \sigma^2)$, die Zufallsvariable T_n sei definiert durch $T_n = \sqrt{n}\,\frac{\overline{X}_n - \mu}{\sqrt{S_n^2}}$. Weiters gelte $Y \sim t_{n-1}$ und F_Y bezeichne die Verteilungsfunktion von Y. Berechnen Sie die Verteilung der Zufallsvariable $\hat{p}$, definiert durch

$$\hat{p}(\omega) = \mathbb{P}\big(|Y| \geq |T_n(\omega)|\big) = 2(1 - F_Y(|T_n(\omega)|)).$$

NB: $\hat{p}(\omega)$ ist die Wahrscheinlichkeit dafür, dass eine t_{n-1}-verteilte Zufallsvariable einen (vom Absolutbetrag her) mindestens so großen Wert wie $T_n(\omega)$ annimmt. Welche in Abschnitt 5 im Skriptum getätigte Aussage verifiziert diese Übungsaufgabe?

Übungsaufgabe 58. Das R-Snippet R-Codes_permtest.R implementiert einen auf den ersten Blick etwas ungewöhnlichen Hypothesentest. Finden Sie heraus, was der R-Code macht, und berechnen Sie approximativ die power-Funktion des Tests. Vergleichen Sie die Powerfunktion mit jener des entsprechenden t-tests. Adaptieren Sie weiters den Code so, dass im Falle zweier Exponentialverteilung auf Gleichheit der Parameter getestet wird, und berechnen Sie approximativ den Fehler erster Art.

Übungsaufgabe 59. Erika Musterfrau schlägt den folgenden Test auf Normalverteilung vor: Ausgehend von einer Stichprobe $X_1, \dots X_n$ von $X \sim F$ setzen wir $\mu_0 = \overline{X}_n$ und $\sigma_0 = \sqrt{S_n^2}$ und testen dann (mittels Aufgabe 44) $H_0 : F = F_{\mu_0, \sigma_0^2}$, wobei F_{μ_0, σ_0^2} die Verteilungsfunktion von $\mathcal{N}(\mu_0, \sigma_0^2)$ bezeichnet. Überprüfen Sie mittels Simulationen, dass der resultierende Test für kleine Sample Sizes ($n = 10, 20, 30$) sehr geringe Power hat und überlegen Sie sich, warum dem so ist.

Übungsaufgabe 60. Wir nehmen an, dass die Zufallsvariablen X, Y, ε die Gleichheit

$$Y = 2X + 1 + \varepsilon$$

erfüllen, wobei $X \sim \mathcal{U}(-3, 3), \varepsilon \sim \mathcal{N}(0, 0.5)$ gilt, X und ε unabhängig sind, und Y dann via $Y = 2X + 1 + \varepsilon$ berechnet wird. Erzeugen Sie in R eine Stichprobe $(X_1, Y_1), \dots, (X_n, Y_n)$ der Größe $n = 100$ von $(X, Y)^\dagger$, berechnen Sie dann in R jene Werte für $a, b \in \mathbb{R}$, die

$$\sum_{i=1}^{n} (Y_i - aX_i - b)^2$$

minimieren (Sie können dafür bspweise die schon bekannte Funktion *nls* oder irgendeinen anderen Minimierer verwenden), und nennen Sie die Werte $\hat{a}_n$ und $\hat{b}_n$. Was ist für großes n zu beobachten?

---

†d.h., erzeugen Sie $X_1, \dots, X_n$ sowie $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ und setzen Sie $Y_i = 2X_i + 1 + \varepsilon_i$