Institut für Angewandte Mathematik
Sommersemester 2024
Andreas Eberle, Stefan Oberdörster

8. Übungsblatt „Einführung in die Statistik"

Abgabe bis Dienstag 11.6.

---

1. (Informationsungleichung) Seien $X_1, \ldots, X_n$ unabhängig und Beta$(r, 1)$-verteilt mit unbekanntem Parameter $r > 0$, d.h. die Verteilung ist absolutstetig mit Dichte

$$f_r(x) \;=\; r\, x^{r-1}\, 1_{(0,1)}(x)\,.$$

a) Bestimmen Sie den Maximum-Likelihood-Schätzer für $1/r$. Ist der Schätzer erwartungstreu? Realisiert er die untere Schranke aus der Informationsungleichung für festes n beziehungsweise asymptotisch?

b) Zeigen Sie, dass $\bar{X}_n$ ein erwartungstreuer Schätzer für $r/(r+1)$ ist. Realisiert dieser die untere Schranke aus der Informationsungleichung?

2. (Test auf Normalverteilung) Die *Kurtose* einer Stichprobe $\boldsymbol{X} = (X_1, \ldots, X_n)$ mit Werten in $\mathbb{R}$ ist definiert als

$$K_n \;=\frac{1}{n}\sum_{i=1}^{n}\left(\frac{X_i-\bar{X}_n}{\hat{\sigma}_n}\right)^4 - 3 \qquad \text{mit} \qquad \hat{\sigma}_n \;=\; \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X}_n)^2}\,.$$

Diese Kenngröße wird als Teststatistik verwendet, um Abweichungen von Normalverteilungen zu entdecken.

a) Für welches statistische Funktional $g(\mu)$ ist K_n ein plug-in Schätzer?

b) Zeigen Sie, dass $g(\mu) = 0$ für alle Normalverteilungen gilt.

c) Konstruieren Sie einen Test auf Normalverteilung, der auf der Kurtose basiert.

d) Sei $n = 9$ und $K_9 = 1,27$. Berechnen Sie approximativ den p-Wert des Tests (z.B. mithilfe eines Monte-Carlo-Verfahrens).

3. (Box-Plots, Mittelwerte und Quantile) Angenommen, Sie kennen von einem Beobachtungsvektor $\boldsymbol{X} = (X_1, \ldots, X_n)$ nur die fünf Kenngrößen

$$Q_0 \;:=\; \min(\boldsymbol{X}), \quad Q_j \;:=\; \hat{q}_{j/4}(\boldsymbol{X}) \;\; \text{für } j = 1,2,3 \quad \text{und} \quad Q_4 \;:= \max(\boldsymbol{X}).$$

Nicht einmal der Stichprobenumfang n sei Ihnen bekannt.

a) Zeigen Sie, dass

$$\frac{Q_0 + Q_1 + Q_2 + Q_3}{4} \;\le\; \overline{X} \;\le\; \frac{Q_1 + Q_2 + Q_3 + Q_4}{4}.$$

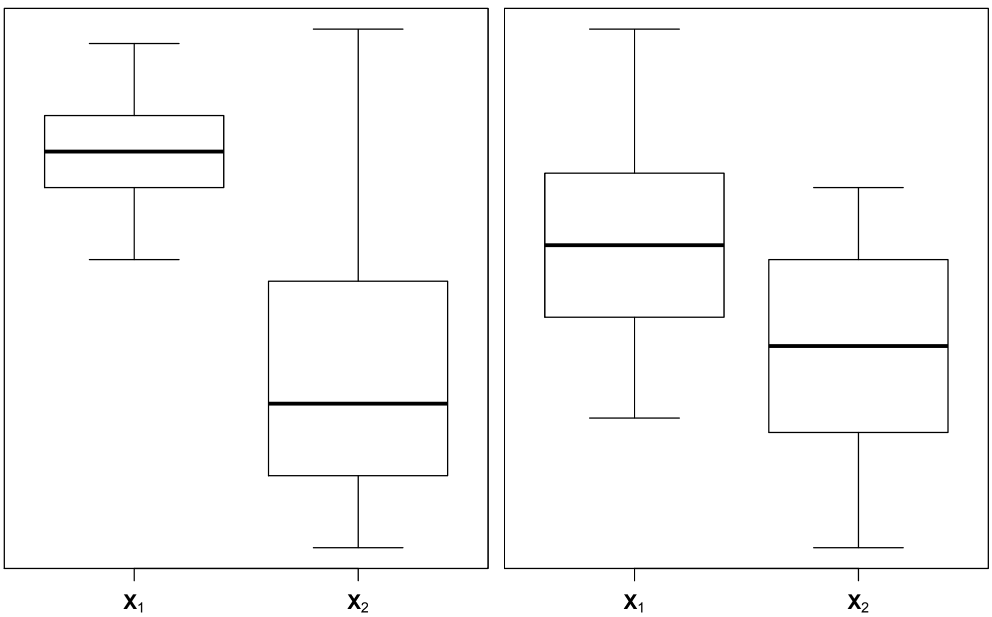
Hinweis: Die Größen $Q_0, \ldots, Q_4$ bleiben unverändert, wenn man jede Komponente von $\boldsymbol{X}$ durch k Kopien ersetzt und somit eine Stichprobe vom Umfang kn erhält. Sie dürfen also davon ausgehen, dass n ein beliebig großes Vielfaches von 4 ist.

b) Angenommen, $X_1, \ldots, X_n$ sind unabhängig und identisch verteilt mit stetiger Verteilungsfunktion. Mit welcher Wahrscheinlichkeit ist $[q_{1/4}, q_{3/4}] \subseteq [Q_0, Q_4]$? Wie groß muss n sein, damit diese Wahrscheinlichkeit mindestens 99% beträgt? Mit welcher Wahrscheinlichkeit ist $q_{1/2} \in \big[Q_1, Q_3\big]$, wenn n kein Vielfaches von 4 ist?

4. (Mann-Whitney-U-Statistik und Boxplots) Die Abbildungen unten zeigen Box-Plots zweier Stichproben $\boldsymbol{X}_1$ und $\boldsymbol{X}_2$ mit unbekannten Stichprobenumfängen n_1 bzw. n_2.

a) Bestimmen Sie aufgrund dieser Box-Plots untere und obere Schranken für die normierte sogenannte Mann-Whitney-U-Statistik

$$\hat{u} \;=\; \frac{1}{n_1 n_2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} h(X_{1i}, X_{2j}) \quad \text{mit} \quad h(x,y) := 1_{[x>y]} + 1_{[x=y]}/2.$$



b) Welche anschauliche Interpretation hat $\hat{u}$? Für welche Kenngröße ist $\hat{u}$ ein Schätzwert, und für welche Testprobleme kann man $\hat{u}$ als Teststatistik verwenden?