05. Übung am 27. April 2026

UV Angewandte Statistik (405.170)
Link Ankreuzliste: siehe www.trutschnig.net/courses
Mit 'F' versehene Aufgaben sind freiwillig, mit * versehene Aufgaben haben einen erhöhten Schwierigkeitsgrad.
Alle Verweise beziehen sich auf das Statistik oder das Angewandte Statistik Skriptum.

Übungsaufgabe 22. Sei X geometrisch verteilt mit Parameter $p \in (0,1)$, für jedes $k \in \mathbb{N}_0$ gelte also

$$\mathbb{P}(X = k) = (1-p)^k \, p.$$

Beweisen Sie, dass dann $\mathbb{E}(X) = \frac{1-p}{p}$ gilt.
Hinweis: Zeigen Sie dafür beispielsweise (mittels elementarer Analysis) zuerst die Gleichheit

$$\sum_{k=1}^{\infty} k(1-x)^{k-1} = \frac{1}{x^2}, \qquad x \in (0,1)$$

und schließen Sie direkt draus $\mathbb{E}(X) = \frac{1-p}{p}$.

Übungsaufgabe 23 (Flaschenrückgabe ein vorletztes Mal, R)**.** Gehen Sie den R-Code Link auf R-Snippet genau durch und finden Sie heraus, was hier gemacht wird. Wie kann es sein, dass schon bei Verwendung von nur 2-3 Zeilen des Outgoing/Incoming Datensatzes der (unbekannte) Parameter p recht genau geschätzt werden kann?

Übungsaufgabe 24 (Flaschenrückgabe ein letztes Mal, R, *)**.** Überlegen Sie sich, wie der R-Code in der letzten Aufgabe so erweitert werden kann, dass ein wesentlich größerer Teil des Datensatzes[iii] für die Schätzung verwendet wird. Wird die Güte des Schätzers für p damit nochmals höher?

Übungsaufgabe 25 (Zu Frage 3.13, R)**.** Die Unabhängigkeit von $\overline{X}_n$ und S_n^2 kann auch illustriert werden, indem überprüft wird, ob $H_n(x,y) \approx F_n(x)G_n(y)$ für alle $x, y \in \mathbb{R}$ gilt. Dabei bezeichnet H_n die zweidimensionale Verteilungsfunktion der Stichprobe

$$(\overline{x}_{n,1}, s_{n,1}^2), (\overline{x}_{n,2}, s_{n,2}^2), \ldots, (\overline{x}_{n,R}, s_{n,R}^2)$$

von $(\overline{X}_n, S_n^2)$ und F_n, G_n die entsprechenden eindimensionalen empirischen Verteilungsfunktionen.
Gehen Sie dazu wiefolgt vor:

1. Wählen Sie $n = 20$ und erzeugen Sie eine Stichprobe $X_1, \ldots, X_n$ von $X \sim \mathcal{N}(0,1)$ und berechnen Sie $(\overline{x}_{n,1}, s_{n,1}^2) := (\overline{X}_n, S_n^2)$.
2. Wiederholen Sie den obigen Vorgang $R = 10.000$ Mal, $(\overline{x}_{n,1}, s_{n,1}^2), (\overline{x}_{n,2}, s_{n,2}^2), \ldots, (\overline{x}_{n,R}, s_{n,R}^2)$ bezeichnen die erhaltenen Paare.
3. Berechnen Sie das Maximum der Funktion $(x,y) \mapsto \Delta_n(x,y) := |H_n(x,y) - F_n(x)G_n(y)|$ auf einem feinen, äqudistanten Gitter G in $[-5,5]^2$. Zusätzlich kann die Funktion auch als heatmap geplottet werden (ggplot command 'geom_tile').

[iii] oder, noch besser, der gesamte Datensatz

4. Wiederholen Sie obige Schritte für $n \in \{50, 100, 1000\}$.

Übungsaufgabe 26 (R)**.** Es gibt mittlerweile zahlreiche Hypothesentests auf Unabhängigkeit von Zufallsvariablen X und Y. Ein relativ neuer Zugang arbeitet mit Copulas und testet nicht nur auf Unabhängigkeit, sondern schätzt auch die Stärke der Abhängigkeit. Gehen Sie die Beschreibung auf Link auf hp bis inkl. Example 2 durch (es reicht, die Grundidee zu verstehen), installieren Sie 'qad' und schätzen Sie dann mittels qad die Abhängigkeit von $\overline{X}_n$ und S_n^2 basierend auf den in Schritt 2 von Aufgabe 25 generierten Daten.