Übungen zur Vorlesung „Elementare Stochastik“

Abgabe der Lösungen zu den S-Aufgaben per E-Mail bis Freitag, 19. Juni 2020, 10 Uhr
Anleitung zur Form der Abgabe: siehe
https://www.math.uni-frankfurt.de/~ismi/wakolbinger/teaching/EleSto20/Abgaben.html

29. *Die Stichprobenvarianz und ihre erwartungstreue Schwester.* a) $x_1, \dots, x_n$ seien reelle Zahlen, und $m := \frac{1}{n}(x_1 + \cdots + x_n)$. Begründen Sie die Identität

$$(*) \qquad \frac{1}{n}\sum_{i=1}^{n}(x_i - m)^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - m^2$$

ohne weitere Rechung aus unserer "hilfreichen Formel für die Varianz" durch Angabe eines passenden Zufallsexperiments.
b) $X_1, \dots, X_n$ seien unabhängige reellwertige Zufallsvariable mit Erwartungswert μ und endlicher Varianz σ^2. Wir setzen $M := \frac{1}{n}(X_1 + \cdots + X_n)$ und $\hat\sigma^2 := \frac{1}{n}\sum_{i=1}^{n}(X_i - M)^2$.
Verwenden Sie $(*)$ jetzt für X_i anstelle von x_i, um zu zeigen:

$$\mathbf{E}[\hat\sigma^2] = \frac{n-1}{n}\sigma^2.$$

Drücken Sie dazu $\mathbf{E}[X_i^2]$ und $\mathbf{E}[M^2]$ jeweils unter Verwendung der "hilfreichen Formel für die Varianz" durch μ und σ^2 aus.

30.S. *Ein Konfidenzintervall für den Populationsmittelwert.* Wir betrachten die Situation von Aufgabe 24b), mit $M_{40} := \frac{1}{40}(Y_1 + \cdots + Y_{40})$, wobei Y_i das Gewicht des im i-ten Zug gewählten Elefanten ist. Es sei Ihnen verraten, dass M_{40} trotz der Abhängigkeit zwischen den Y_i approximativ nornalverteilt ist (dahinter steckt die gegen Ende von V7a1 angesprochene Weiterung des Zentralen Grenzwertsatzes). Im Rest der Aufgabe dürfen Sie deshalb mit der Normalapproximation rechnen. Es sei $\mu = \frac{1}{100}(h(a_1) + \cdots + h(a_{100}))$ der in A14a) berechnete Populationsmittelwert.
a) Für welche Zahl δ ist die Wahrscheinlichkeit, dass M_{40} um mehr als δ von μ abweicht, ungefähr gleich 0.05? Verwenden Sie das Ergebnis aus A24a) und die Normalapproximation.
b) Geben Sie ein um M_{40} zentriertes Intervall I an, welches nur auf den zufälligen Beobachtungen $Y_1, \dots, Y_{40}$ basiert, sodass $\mathbf{P}(\mu \in I) \approx 0.95$.[1] Dabei dürfen Sie davon ausgehen, dass hier die in A29 definierte *Stichprobenvarianz* $\hat\sigma^2$ nahe bei der Populationsvarianz σ^2 liegt.
c) (optional) Überprüfen Sie empirisch, mit welcher Wahrscheinlichkeit das von Ihnen in b) angegebene Intervall den Populationsmittelwert μ überdeckt, indem Sie den Rechner viele (sagen wir 1000) Stichproben der Größe 40 ziehen lassen. Verwenden Sie dazu das unter dem Link zu diesem Übungsblatt auf der Lehrveranstaltungs-Webseite A30.R . verlinkte R-Programm.

31. *Regression towards the mean: Sir Francis Galtons Paradoxon.* Ein aus einer großen Population rein zufällig herausgegriffenes Vater-Sohn-Paar habe die Körpergrößen (X, Y), wobei (X, Y) so verteilt sei wie $(175 + 8Z_0 + 8Z_1, 175 + 8Z_0 + 8Z_2)$, mit Z_0, Z_1, Z_2 unabhängig und standard-normalverteilt.
a) Bestimmen Sie die beste affin lineare Vorhersage von Y auf der Basis von X.
b) Bestimmen Sie die beste affin lineare Vorhersage von X auf der Basis von Y.
c) Ist der Sohn eines überdurchnittlich großen Vaters eher größer oder eher kleiner als sein Vater?
d) Ist der Vater eines überdurchnittlich großen Sohnes eher größer oder eher kleiner als sein Sohn?

32.S. *Zweistufig hin und zurück.* Das zufällige Paar (X_1, X_2) mit Werten in $\{1, 2\} \times \{b, c, d\}$ komme durch ein zweistufiges Experiment zustande, wobei $\mathbf{P}(X_1 = 1) = 2\mathbf{P}(X_1 = 2)$ gelte und die Übergangswahrscheinlichkeiten $P(a, .)$, $a \in \{1, 2\}$, durch die rechts angegebene Matrix bestimmt sind.

	b	c	d
1	0.4	0.4	0.2
2	0.1	0.2	0.7

(i) Finden Sie die Matrix der gemeinsamen Verteilungsgewichte von (X_1, X_2) sowie die Verteilungsgewichte von X_2.
(ii) Finden Sie Übergangswahrscheinlichkeiten $Q(a, .)$, $a \in \{b, c, d\}$, sodass (X_2, X_1) als zweistufiges Zufallsexperiment (mit X_2 als erster und X_1 als zweiter Stufe) entsteht.

[1] Ein zufälliges Intervall I mit der Eigenschaft $\mathbf{P}(\mu \in I) \approx 0.95$ heißt *Konfidenzintervall für μ, approximativ zum Niveau* 0.95.