# Übungen zur Vorlesung „Stochastik für die Informatik“

Abgabe der Lösungen zu den S-Aufgaben: Freitag, 13. Dezember 2019, vor der Vorlesung (12:10-12:15 im Magnus HS) oder bis zu diesem Termin direkt an Ihre Tutorin/Ihren Tutor, z. B. über die Kästen in der RM6, 3. Stock

**Geben Sie kurze und treffende Begründungen für Ihre Ergebnisse!**

**29 S.** a) $x_1, \ldots, x_n$ 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.[1]
b) $X_1, \ldots, X_n$ seien unabhängige reellwertige Zufallsvariable mit Erwartungswert $\mu$ und endlicher Varianz $\sigma^2$. Wie in Vorlesung 7b setzen wir $M := \frac{1}{n}(X_1 + \cdots + X_n)$ und $\hat{\sigma}^2 := \frac{1}{n}\sum_{i=1}^{n}(X_i - M)^2$. Verwenden Sie (unter dem Motto “Was einer Gleichheit von Termen für Variablen recht ist, ist derseben Gleichheit für Zufallsvariablen billig”) die Identität $(*)$ 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 $\mu$ und $\sigma^2$ aus.

**30.** Wir betrachten die Situation von Aufgabe 20. Es sei $\mu$ der in A20a) berechnete Populationsmittelwert, $\sigma^2$ die in A20a) berechnete Populationsvarianz, und $X_1, \ldots, X_{30}$ die zufälligen Werte, die bei einem 30-maligen Ziehen ohne Zurücklegen entstehen (vgl. A20b)).
a) Warum sind die $X_i$ nicht unabhängig?
b) Wir setzen $M := \frac{1}{30}(X_1 + \cdots + X_{30})$. Es sei Ihnen verraten, dass trotz der fehlenden Unabhängigkeit der $X_i$ auch hier die asymptotische Normalität greift, vgl. dazu die Ausblicke in VL 7a Folie 89 und VL 7b Folie 16. Deshalb dürfen Sie im Rest der Aufgabe mit der Normalapproximation rechnen, Sie sollten dabei aber die Standardabweichung von $M$ verwenden, die wir in Aufgabe 20 berechnet haben.
(i) Für welche Zahl $\delta$ ist die Wahrscheinlichkeit, dass $M$ um mehr als $\delta$ von $\mu$ abweicht, ungefähr gleich 0.05?
(ii) Geben Sie ein um $M$ zentriertes Intervall $I$ an, sodass $\mathbf{P}(\mu \in I) \approx 0.95$.[2]
c) Überprüfen Sie, mit welcher Wahrscheinlichkeit das in b) ii) berechnete Intervall dem Populationsmittelwert $\mu$ überdeckt, indem Sie den Rechner viele Stichproben der Größe 30 ziehen lassen. Verwenden Sie dazu das unter dem Link A30.R auf der Lehrveranstaltungs-Webseite bereitgestellte R-Programm.

**31.** $X_1, \ldots, X_k$ seien unabhängig und Exp(1)-verteilt (mit $k \in \mathbb{N}$), und $Y := X_1 + \ldots + X_k$. Berechnen Sie für $-\infty < t < 1$ (i) $\mathbf{E}[e^{tX_1}]$, (ii) $\mathbf{E}[e^{tY}]$.

**32 S.** $X_1$, $X_2$ und $Y$ seien reellwertige Zufallsvariable. Die Standardabweichungen von $X_1$ und $X_2$ seien gleich.
a) Das Wievielfache von $\sigma_{X_1}$ ist $\sigma_{X_1+X_2}$, wenn
(i) $X_1$ und $X_2$ unkorreliert sind,
(ii) $X_1$ und $X_2$ Korrelationskoeffizient $-0.5$ haben?
b) Die Standardabweichung von $Y$ sei doppelt so groß wie die von $X_1 + X_2$ und die Standardabweichungen von $X_1$ und $X_2$ seien gleich. Der Korrelationskoeffizient von $X_1$ und $X_2$ sei -0.5. Außerdem sei der Korrelationskoeffizient von $Y$ und $X_1$ gleich dem Korrelationskoeffizienten von $Y$ und $X_2$. Die beste affin lineare Prognose von $Y$ auf der Basis von $X_1+X_2$ sei $\hat{Y} = (X_1+X_2)+5$. Berechnen Sie den Korrelationskoeffizienten von $Y$ und $X_1$.

---

[1] Die Geometrie hinter der Identität $(*)$ kann man so einsehen: $(x_1 - m, \ldots, x_n - m)$ und $(m, \cdots, m)$ sind zwei Vektoren im $\mathbb{R}^n$, deren Skalarprodukt Null ist (überprüfen Sie das!). Multipliziert man $(*)$ mit $n$, dann wird es zur Formel von Pythagoras für die (aufeinander ortogonal stehenden) Vektoren $(x_1 - m, \ldots, x_n - m)$ und $(m, \cdots, m)$.

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