Übungen zur Vorlesung „Stochastik für die Informatik“

Abgabe der Lösungen zu den S-Aufgaben: Freitag, 14. Dezember 2018, vor der Vorlesung (12:10-12:15 im Magnus HS)

29 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.[1]

b) $X_1, \dots, X_n$ seien unabhängige reellwertige Zufallsvariable mit Erwartungswert μ und endlicher Varianz σ^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 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 μ und σ^2 aus.

30. a) Wir wissen schon: Die Wahrscheinlichkeit, dass eine normalverteilte Zufallsvariable Y außerhalb ihrer $2\sigma_Y$-Grenzen fällt, ist ≈ 0.05. Wie groß ist die Wahrscheinlichkeit, dass eine normalverteilte Zufallsvariable
(i) außerhalb ihrer $2.5\sigma_Y$-Grenzen fällt (gerundet auf Prozent)
(ii) außerhalb ihrer $3\sigma_Y$-Grenzen fällt (gerundet auf Promille)?
b) Berechnen Sie, gerundet auf 2 Nachkommastellen, die Zahl c jeweils so, dass eine normalverteilte Zufallsvariable Y mit Wahrscheinlichkeit
(i) 0.01 (ii) 0.001
außerhalb ihrer $c\sigma_Y$-Grenzen fällt.
(Zur Erinnerung: In R bekommt man die Verteilungsfunktion der Standardnormalverteilung als $\Phi(b) = \texttt{pnorm}(b)$ und die zugehörige *Quantilfunktion* als $\Phi^{-1}(p) = \texttt{qnorm}(p)$.)

31.S. $X_1, \dots, X_{100}$ seien unabhängige reellwertige Zufallsvariable mit Erwartungswert μ und endlicher Varianz σ^2. Wieder setzen wir $M := \frac{1}{100}(X_1 + \cdots + X_{100})$.
a) Bestimmen Sie $c > 0$ jeweils so, dass die Zufallsvariable M mit Wahrscheinlichkeit
(i) ≈ 0.95 (ii) ≈ 0.99
nicht weiter als $c\sigma_M$ von μ entfernt ausfällt.
b) Bestimmen Sie $\gamma > 0$ jeweils so, dass das zufällige Intervall $[M - \gamma\sigma, M + \gamma\sigma]$ mit Wahrscheinlichkeit
(i) ≈ 0.95 (ii) ≈ 0.99
den Parameter μ enthält.

32.S. a) Z_1 und Z_2 seien unkorreliert mit $\mathbf{E}[Z_1] = 1$, $\mathbf{E}[Z_2] = -1$, $\mathbf{E}[Z_1^2] = 3$, $\mathbf{E}[Z_2^2] = 2$. Es sei $X_1 := 5Z_1 - 2Z_2$, $X_2 := -3Z_1 + Z_2$, $X_3 = 2Z_1 + 4Z_2$. Finden Sie die (im Sinn des erwarteten quadratischen Fehlers) beste affin lineare Prognose von X_3 auf der Basis von $X_1 + X_2$.
b) Was ändert sich am Ergebnis, wenn $\mathbf{E}[Z_1^2] = \mathbf{E}[Z_2^2] = 3$ gilt?

[1] Die Geometrie hinter der Identität $(*)$ kann man so einsehen: $(x_1 - m, \dots, x_n - m)$ 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 $(x_1 - m, \dots, x_n - m)$ und $(m, \cdots, m)$.