Übungen zur Vorlesung „Stochastik für die Informatik“

Abgabe der Lösungen zu den S-Aufgaben: Dienstag, 12. Januar 2016, vor der Vorlesung (10:05-10:15 im Magnus HS)

37. a) X und Y seien unabhängig und Pois(α) bzw. Pois(β)-verteilt. Begründen Sie, dass $X+Y$ Pois($\alpha+\beta$) verteilt ist. *Hinweis: Für alle* $k \in \mathbb{N}_0$ *ist* $\sum_{i+j=k} \frac{\alpha^i}{i!}e^{-\alpha}\frac{\beta^j}{j!}e^{-\beta} = \frac{(\alpha+\beta)^k}{k!}e^{-(\alpha+\beta)}$ - *warum gilt das?*
b) Für jedes n sei X_n Poissonverteilt zum Parameter n. Begründen Sie, warum für $n \to \infty$
(i) $\frac{X_n}{n}$ stochastisch gegen 1 konvergiert (vgl. VL 7b Folie 28),
(ii) $\frac{X_n - n}{\sqrt{n}}$ in Verteilung gegen N(0,1) konvergiert.

38. a) Aus einer großen Population von Werten auf $\mathbb{R}$ wird eine Stichprobe des Umfangs n (mit nicht all zu kleinem n) entnommen. Der Stichprobenmittelwert ist m und die modifizierte Stichprobenvarianz ist s^2. Geben Sie unter Verwendung der Normalapproximation ein aus M und S konstruiertes (möglichst kurzes) zufälliges Intervall I an, welches den Mittelwert μ der Population mit Wahrscheinlichkeit 0.95 überdeckt.

b) Machen Sie die Probe aufs Exempel, wie gut dieses Verfahren bei unserer fiktiven Population der HiWi-Stundenverdienste aus Aufgabe 17 und 38 funktioniert. Verwenden Sie dabei das von Benjamin Straub zur Verfügung gestellte R-Programm *A38.R* (Link auf der LV-Seite unterhalb der Links zu den Übungsaufgaben.) Mit diesem Programm wird für einen von Ihnen einstellbaren Stichprobenunfang n ein Histogramm der Verteilung des Stichprobenmittelwertes M geplottet und (für eine ebenfalls von Ihnen wählbare Anzahl von Wiederholungen) die empirische Überdeckungswahrscheinlichkeit von Konfidenzintervallen für μ bestimmt, und zwar (i) mit der modifizierte Stichprobenvarianz s^2 wie in a), sowie (ii) mit als bekannt vorausgesetzter Populationsvarianz σ^2. (Die Population, aus der gezogen wird, wird dabei als so groß vorausgesetzt, dass die Effekte der Kollisionen beim Ziehen ohne Zurücklegen vernachlässigt werden können.) Diskutieren Sie, was für $n = 10$, $n = 20$ und $n = 30$ zu beobachten ist.

39. S. Z_i ist N(i, i^2)-verteilt, $i = 1, 2, 3$, und Z_1, Z_2, Z_3 sind unabhängig. Es sei $X := Z_1 - 2Z_2 + 3Z_3$, $Y := 3Z_1 - Z_2 - 3Z_3$. Finden Sie die (im Sinn des erwarteten quadratischen Abstandes) beste affin lineare Prognose von Y auf der Basis von X, d.h. dasjenige h von der Form $h(x) = \beta_0 + \beta_1 x$, für welches $\mathbf{E}[(Y - h(X))^2]$ minimal wird.

40. S. Das zufällige Paar (X_1, X_2) mit Werten in $\{1,2,3\} \times \{b,c,d\}$ komme durch ein zweistufiges Experiment zustande, wobei die Verteilung von X_1 uniform sei und die Übergangswahrscheinlichkeiten $P(a_1,.)$, $a_1 \in \{1,2,3\}$, durch die rechts angegebene Matrix bestimmt sind.

	b	c	d
1	0	0.6	0.4
2	0.3	0.2	0.5
3	0.6	0.3	0.1

(i) Finden Sie die Matrix der gemeinsamen Verteilungsgewichte von (X_1, X_2) und die Verteilung von X_2.

(ii) Finden Sie Übergangswahrscheinlichkeiten $Q(a_2,.)$, $a_2 \in \{b,c,d\}$ so, dass das zufällige Paar (X_2, X_1) als zweistufiges Zufallsexperiment (jetzt mit X_2 als erster Stufe) entsteht.

(iii) Berechnen Sie jeweils für $a_2 = b$, $a_2 = c$ bzw $a_2 = d$ den bedingten Erwartungswert von X_1 gegeben $X_2 = a_2$.

(iv) Finden Sie die *(im Sinn des erwarteten quadratischen Abstandes) beste Prognose von* X_1 auf der Basis von X_2, d.h. diejenige Funktion h, für die $\mathbf{E}[(X_1 - h(X_2))^2]$ minimal wird.