Übungen zur Vorlesung „Stochastik für die Informatik“

Abgabe der Lösungen zu den S-Aufgaben: Freitag, 14. Dezember 2012

25. a) Die Chebychev–Ungleichung besagt, dass für eine Zufallsvariable Y mit Erwartungswert μ und Standardabweichung σ gilt: $\mathbf{P}(|Y-\mu| \geq c\sigma) \leq \frac{1}{c^2}$ für jedes $c > 0$.
Begründen Sie mit der Chebychev–Ungleichung: Für jedes $p \in [0,1]$ gilt: Falls $X_1, X_2, \ldots$ eine Münzwurffolge zum Parameter p ist, so ist die Wahrscheinlichkeit, p durch das Intervall $I := \left[\overline{X}_n - \frac{\sqrt{5}}{\sqrt{n}}\ ,\quad \overline{X}_n + \frac{\sqrt{5}}{\sqrt{n}}\right]$ zu überdecken, nicht kleiner als 0.95. (Dabei ist $\overline{X}_n := \frac{1}{n}\sum\limits_{i=1}^{n} X_i$.) (Man sagt: I ist ein *Konfidenzintervall* für den Parameter p mit einer Überdeckungswahrscheinlichkeit von mindestens 95%.)

b) Es sei Y eine $\mathrm{N}(p, pq/n)$-verteilte Zufallsvariable. Finden Sie ein möglichst kleines $\delta = \delta(n)$ so, dass die Wahrscheinlichkeit, die Zahl p durch das Intervall $J = [Y-\delta, Y+\delta]$ zu überdecken, (für jedes p) nicht kleiner ist als 0.95.

c) Welches approximative 95%–Konfidenzintervall erhält man für die Erfolgswahrscheinlichkeit p in einem n-fachen p-Münzwurf aus der relativen Anzahl $\bar{X}_n$ der Erfolge, wenn mit der Normalapproximation gerechnet wird?

d) (Als Kür) Bestimmen Sie per Simulationsexperiment für $n = 25$ und $p = 3/4$ die tatsächliche Überdeckungswahrscheinlichkeit des in c) gefundenen approximativen 95%–Konfidenzintervalls.

26. S. In 30 Laboratorien werden Messungen durchgeführt, man stellt sich vor, dass sie als unabhängige, identisch verteilte Zufallsvariable (mit endlichem Erwartungswert) zustande kommen, also von der Form $\mu + Z_i$ sind, mit unabhängigen, identisch verteilten, *zentrierten* Zufallsvariablen Z_i. Die Standardabweichung einer einzelnen Messung wird mit $\sigma = 0.3$ angegeben. Es geht darum, den Modellparameter μ “mit Konfidenz” zu schätzen

a) Wie groß ist der *Standardfehler*, d.h. die Standardabweichung des Stichprobenmittelwertes $\bar{X}$ aus den 30 Messungen?
b) Geben Sie ein zufälliges Intervall $[\bar{X}-c, \bar{X}+c]$ an, das den Parameter μ mit Wahrscheinlichkeit ≈ 0.95 enthält.
c) Aus den 30 Messungen ergab sich der Stichprobenmittelwert 27,9. Angenommen $\mu = 28$. Wie wahrscheinlich ist dann eine mindestens so große Abweichung des Stichprobenmittelwertes von μ wie die beobachtete? (Hinweis: Der R-Befehl `pnorm` ist hilfreich.)

27. (i) Zeigen Sie: Für $x := (x_1, \ldots, x_n) \in \mathbb{R}^n$ und $\bar{x} := \frac{1}{n}\sum_{i=1}^n x_i$ ist $\bar{x}\mathbf{1} := (\bar{x}, \ldots, \bar{x})$ die *Orthogonalprojektion* von x auf $D := \{(a, \ldots, a) : a \in \mathbb{R}\}$ in dem Sinn, dass das euklidische Skalarprodukt $\langle x - \bar{x}\mathbf{1}, y\rangle$ für alle $y \in D$ verschwindet.

(ii) Folgern Sie aus (i), dass für jedes $c \in \mathbb{R}$ gilt: $\quad (*) \quad \sum\limits_{i=1}^{n}(x_i - c)^2 = \sum\limits_{i=1}^{n}(x_i - \bar{x})^2 + n(\bar{x} - c)^2.$

(iii) Seien $X_1, \ldots, X_n$ paarweise unkorreliert und identisch verteilt mit Erwartungswert μ und Varianz $\sigma^2 < \infty$, und sei $\bar{X} := \frac{1}{n}(X_1 + \cdots + X_n)$ ihr Stichprobenmittel. Zeigen Sie:

$$\frac{1}{n-1}\big((X_1 - \bar{X})^2 + \cdots + (X_n - \bar{X})^2\big) \text{ hat Erwartungswert } \sigma^2.$$

Hinweis: Benutzen Sie die Zerlegung $(*)$ mit $c := \mu$.

28. S. a) Z_1 und Z_2 sind unkorreliert mit Erwartungswert 0 und Varianz 1. Die Zufallsvariablen X und Y sind definiert durch $X = 2Z_1 + Z_2$, $Y = Z_1 - Z_2 + 1$.
Finden Sie unter allen Zufallsvariablen $h(X)$ der Form $h(X) = \beta_0 + \beta_1 X$ diejenige, für die $\mathbf{E}((Y - h(X))^2)$ minimal wird.

b) Besteht die eingangs von a) fomulierte Eigenschaft von Z_1 und Z_2, wenn Z_1 und Z_2 unabhängig und $\mathcal{N}(0,1)$-verteilt sind? Wie sind in diesem Fall X und Y verteilt?