Übungen zur Vorlesung „Stochastik für die Informatik“

Abgabe der Lösungen zu den S-Aufgaben: Donnerstag, 15. Dezember 2011, zu Beginn der Vorlesung

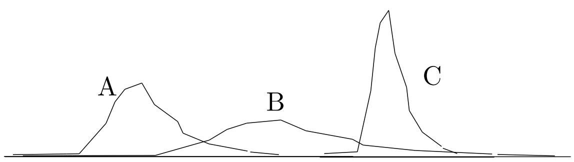
25. S 100 Punkte werden rein zufällig (d.h. unabhängig und uniform) auf den Einheitskreis $A := \{(x,y) : x^2 + y^2 \le 1\}$ verteilt. Ihre Abstände vom Zentrum $(0,0)$ sind $R_1, \dots, R_{100}$.
a) Wo ist die Dichte der Zufallsvariablen R_1 am größten?
b) Finden Sie (näherungsweise) Zahlen r und δ so, dass $\frac{1}{100}(R_1 + \dots R_{100})$ mit Wahrscheinlichkeit 0.95 in das Intervall $[r-\delta, r+\delta]$ fällt.

26. (i) Zeigen Sie: Für $x := (x_1, \dots, x_n) \in \mathbb{R}^n$ und $\bar{x} := \frac{1}{n}\sum_{i=1}^n x_i$ ist $\bar{x}\mathbf{1} := (\bar{x}, \dots, \bar{x})$ die *Orthogonalprojektion* von x auf $D := \{(a, \dots, a) : a \in \mathbb{R}\}$ 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: $(*)\quad \sum_{i=1}^n (x_i - c)^2 = \sum_{i=1}^n (x_i - \bar{x})^2 + n(\bar{x} - c)^2.$
(iii) Seien $X_1, \dots, X_n$ paarweise unkorreliert und identisch verteilt mit Erwartungswert μ und Varianz $\sigma^2 < \infty$, und sei $\bar{X} := \frac{1}{n}(X_1 + \dots + X_n)$ ihr Stichprobenmittel. Zeigen Sie:

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

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

27. (Aus einer Klausur "Statistik für Biologen") Es geht um die Flügelbreite in drei Vogelpopulationen A, B, C. Bei A ist die Stichprobengröße $n_A = 200$, bei B ist sie $n_B = 400$, bei C ist sie $n_C = 20$.



Die drei Verteilungen sind durch die drei skizzierten Dichtepolygone beschrieben. a) Die drei Standardabweichungen σ_A, σ_B und σ_C stehen in einem ungefähr ganzzahligen Verhältnis ?:?:1. In welchem?
b) In welcher der drei Situationen ist der Standardfehler (des Mittelwertes) am kleinsten? (Bitte die Antwort kurz begründen!)
c) Die Standardabweichung in der Stichprobe B ist $s_B = 2$ cm, der Stichprobenmittelwert m_B ist 50 cm. Geben Sie ein Intervall an, das den Populationsmittelwert μ_B mit ca. 95% Wahrscheinlichkeit enthält.
d) Wie variabel ist die Differenz der Stichprobenmittelwerte von A und B? Schätzen Sie die Standardabweichung der Differenz $m_A - m_B$. (Hinweis: Den Wert für s_A gewinnt man aus den Angaben in a) und c).)

28. S Das zufällige Paar (X,Y) hat die in der Tabelle angegebenen Verteilungsgewichte.

a) Bestimmen Sie μ_X, μ_Y, σ_X, σ_Y und $\kappa_{X,Y}$.

b) Es geht darum, Y "möglichst gut auf (affin) lineare Weise auf der Basis von X vorherzusagen", d.h. Koeffizienten c und d so zu finden, dass der erwartete quadratische Fehler $\mathbf{E}[(Y - cX - d)^2]$ minimal wird. Lösen Sie diese Aufgabe.

y \ x	0	1	2	3	4
0	0.1	0.05	0.05	0.05	0
1	0.05	0.1	0.1	0.2	0.05
2	0	0.05	0.05	0.05	0.1