Karlsruher Institut für Technologie

**Institut für Stochastik**
Statistik WS 17/18
Dr. B. Ebner · A. Glauner, M. Sc.

# Praktikumsblatt 2

**Aufgabe 5**
Zeichnen Sie für $-1 \leq x \leq 1$ die Exponentialfunktion $f(x) = \exp(x)$ und die Parabel $g(x) = x^2/2 + x + 1$ mit dem `plot`-Befehl gemeinsam in einem Bild. Wiederholen Sie dies, wobei Sie den `curve`-Befehl verwenden.

**Aufgabe 6**
Untersuchen Sie die Verteilung der Daten in den ersten zwei Spalten des Data Frames **hills**, indem Sie

(a) Histogramme

(b) Dichteplots

(c) Boxplots

(d) QQ-Plots

verwenden. Wiederholen Sie (a) bis (d) mit den logarithmierten Daten. Was sind die Vor- und Nachteile der einzelnen Darstellungen?

**Aufgabe 7**
Die Wahrscheinlichkeiten der $k\sigma$-Bereiche einer Normalverteilung mit Mittelwert $\mu$ und Varianz $\sigma^2$ sind durch

$$P(\mu - k \cdot \sigma \leq X \leq \mu + k \cdot \sigma),\ k \in \mathbb{N},\ X \sim \mathcal{N}(\mu, \sigma^2)$$

gegeben.

(a) Berechnen Sie diese Wahrscheinlichkeiten für $\mu = 3$, $\sigma^2 = 4$ und $k = 1, \ldots, 5$ ohne Verwendung einer for-Schleife.

(b) Zeichnen Sie die Dichte der Normalverteilung aus Teilaufgabe a) und visualisieren Sie die Wahrscheinlichkeit des $k\sigma$-Bereichs für $k = 1, 2, 3$ durch Einfärben der entsprechenden Fläche unter der Dichte.

**Aufgabe 8**
Zeichnen Sie die Dichten von Lognormalverteilungen $LN(0.5, 1)$, $LN(2, 1)$ und $LN(0.5, 2)$ mit dem Befehl `curve` gemeinsam in einem Bild.

**Aufgabe Z1**
Bei einem Boxplot wird die Länge der Box durch das obere und das untere (empirische) Quartil begrenzt, die Länge der Box entspricht also dem Interquartilsabstand. Datenpunkte, die mehr als das 1,5-Fache des Interquartilsabstands von der Box entfernt liegen, werden einzeln markiert, da sie als Ausreißer (in Bezug auf eine Normalverteilung) angesehen werden.
Nehmen Sie an, es liegen ihnen ein großer Datensatz aus einer normalverteilten Grundgesamtheit vor. Bestimmen Sie entweder theoretisch oder durch Simulation, welcher Anteil der Daten bei einem Boxplot unter dieser Verteilungsannahme ungefähr als Ausreißer deklariert wird.

**Aufgabe Z2**
Schreiben Sie eine Funktion, die die Verteilungsfunktion

$$G(z) \quad = \quad 1 - 2\sum_{j=1}^{\infty}(-1)^{j+1}e^{-2j^2z^2} \quad (z > 0)$$

der Kolmogorov-Smirnov-Verteilung berechnet, und plotten Sie $G$.

**Aufgabe Z3**

(a) Beim Geburtstagsproblem ist nach der Wahrscheinlichkeit $p_k$ gefragt, dass unter $k$ rein zufällig ausgewählten Personen mindestens zwei am selben Tag Geburtstag haben. Diese ist für $k \geq 2$ gegeben durch

$$p_k \quad = \quad 1 - \prod_{j=0}^{k-1}\left(\frac{365-j}{365}\right) \;=\; 1 - \prod_{j=1}^{k-1}\left(1 - \frac{j}{365}\right),$$

falls die Annahme einer Gleichverteilung über alle 365 Tage gemacht wird. Bestimmen Sie $p_k$ für $k = 1, \ldots, 50$ unter dieser Annahme und plotten Sie das Ergebnis.

(b) Approximieren Sie die Wahrscheinlichkeit $p_{23}$, indem Sie das Experiment 10000 mal simulieren.
**Hinweis:** Ziehen mit Zurücklegen: Funktion `sample()`.
Bestimmung gleicher Elemente in einem Vektor: Funktion `duplicated()`.

(c) Wiederholen Sie die Simulation, ohne die Annahme einer Gleichverteilung über alle 365 Tage zu machen. Verwenden Sie dazu die Option `prob` in der Funktion `sample()`.

**Aufgabe Z4**
Visualisieren Sie das empirische Gesetz von der Stabilisierung relativer Häufigkeiten. Simulieren Sie dazu eine große Anzahl von unabhängigen Wiederholungen eines Treffer/Niete-Experiments, und stellen Sie die relative Häufigkeit für wachsenden Stichprobenumfang graphisch dar.