2. Übungsblatt zur Vorlesung
Statistische Methoden der Datenanalyse
Abgabe: 15.11.2015, 23:59 Uhr

WS 2016/2017
Prof. W. Rhode

Zeit	Raum	Abgabe im Moodle; Mails mit Betreff: [SMD1617]
Mo. 10-12	CP-O3-150	thorben.menne@udo.edu
Di. 10-12	CP-O3-150	maximilian.noethe@udo.edu und tobias.hoinka@udo.edu
Di. 16-18	CP-O3-150	mathis.boerner@udo.edu
Di. 16-18	P1-02-111	philipp.schlunder@udo.edu und maximilian.meier@udo.edu

Aufgabe 1: *Zufallszahlen verschiedener Verteilungen* **4 P.**

Die Zufallsvariable x möge der Wahrscheinlichkeitsdichte

$$f(x) = \begin{cases} 1 & 0 \leq x \leq 1 \\ 0 & \text{sonst} \end{cases}$$

(Gleichverteilung zwischen 0 und 1) genügen.

a) Mit welcher Wahrscheinlichkeit nimmt x einen Wert zwischen $\frac{1}{2}$ und $\frac{1}{3}$ an?

b) Mit welcher Wahrscheinlichkeit nimmt x den exakten Wert $\frac{1}{2}$ an?

c) Mit welcher Wahrscheinlichkeit liefert ein Zufallsgenerator auf einem Computer den exakten Wert $\frac{1}{2}$? Der Generator soll sein Ergebnis in Form einer binären Gleitkommazahl mit einer Mantisse von 23 Binärstellen darstellen.

d) Mit welcher Wahrscheinlichkeit liefert derselbe Zufallsgenerator den exakten Wert $\frac{2}{3}$?

Aufgabe 2: *Zufallszahlengeneratoren* **6 P.**

Linear-kongruente Zufallszahlengeneratoren erzeugen eine neue ganzzahlige Zufallszahl aus der vorhergehenden durch die Vorschrift

$$x_n = (a \cdot x_{n-1} + b) \mod m.$$

Division durch m ergibt dann eine zwischen 0 und 1 gleichverteilte reelle Zufallszahl.

a) Programmieren Sie einen solchen Zufallszahlengenerator mit $a = 1601$, $b = 3456$ und $m = 10\,000$.

b) Erzeugen Sie so 10000 Zufallszahlen und stellen Sie diese als Histogramm dar. Entspricht das Ergebnis den Anforderungen an einen guten Zufallszahlengenerator? Hängt es vom Startwert x_0 ab, und wenn ja, wie?

c) Stellen Sie Paare bzw. Tripletts aufeinanderfolgender Zufallszahlen als zweidimensionales bzw. dreidimensionales Streudiagramm (engl. scatter plot) dar. Entspricht das Ergebnis den Anforderungen an einen guten Zufallszahlengenerator?

d) Vergleichen Sie diesen Generator mit dem in `root` implementierten `TRandom.Rndm()`. Sehen Sie sich dazu den Quelltext an. Um was für einen Generator handelt es sich und welche Periodizität hat dieser Generator?

e) Erstellen Sie Histogramme wie in (b) und (c) auch mit `TRandom.Rndm()`.

f) Wie oft liefert der Zufallsgenerator aus Aufgabenteil a) den exakten Wert $\frac{1}{2}$? Hängt diese Anzahl vom Startwert ab? Geben Sie einen möglichen Startwert an, sodass der Generator $\frac{1}{2}$ erzeugen kann.

Beispiel für ein dreidimensionales Streudiagramm in `matplotlib`:

```python
import matplotlib.pyplot as plt
import numpy as np
from mpl_toolkits.mplot3d import Axes3D

x, y, z = np.random.normal(size=(3, 1000))

fig = plt.figure()
ax = fig.add_subplot(111, projection='3d')

ax.init_view(45, 30)  # Elevation, Rotation
ax.scatter(
  x, y, z,
  lw=0,   # no lines around points
  s=5,    # smaller points
)

plt.show()
```

Aufgabe 3: *Gleichverteilung* **5 P.**

Gegeben sei ein Zufallszahlengenerator, der gleichverteilte Zahlen z von 0 bis 1 liefert. Geben Sie **effiziente Algorithmen** an, und implementieren Sie diese, mit denen Sie Zufallszahlen erzeugen können, die den folgenden Verteilungen gehorchen:

a) Eine Gleichverteilung in den Grenzen $x_{\min}$ bis $x_{\max}$

b) Exponentialgesetz: $f(t) = Ne^{-t/\tau}$ in den Grenzen 0 bis ∞ (N = Normierungskonstante)

c) Potenzgesetz: $f(x) = Nx^{-n}$ in den Grenzen x_{min} bis x_{max} ($n \geq 2$, N = Normierungskonstante)

d) Cauchy-Verteilung:

$$f(x) = \frac{1}{\pi}\frac{1}{1+x^2}$$

in den Grenzen $-\infty$ bis ∞

e) Die durch das (im Moodle unter *empirisches_histogramm.npy* zu findene) Histogramm gegebene empirische Verteilung. Die Datei enthält Binzentren (*bin_mid*) und die Höhen (*hist*). Das Histogram besteht aus 50 Bins zwischen 0,0 und 1,0.

Zum Einlesen und Darstellen dieses Histogramms können Sie z.B. so vorgehen:

```
from matplotlib import pyplot as plt
import numpy as np
data = np.load("empirisches_histogramm.npy")
plt.hist(data['bin_mid'], bins=np.linspace(0., 1., 50),
         weights=data['hist'])
plt.show()
```

Aufgabe 4: *Fehlerfortpflanzung* **5 P.**

Die Parameter einer Ausgleichsgeraden $y = a_0 + a_1 x$ wurden zu $a_0 = 1{,}0 \pm 0{,}2$ und $a_1 = 1{,}0 \pm 0{,}2$ bestimmt. Der Korrelationskoefffizient ist $\rho = -0{,}8$. Bestimmen Sie die Unsicherheit eines Wertes y als Funktion von x.

a) Bestimmen Sie das Resultat analytisch sowohl unter Berücksichtigung der Korrelation als auch unter Vernachlässigung der Korrelation.

b) Bestimmen Sie das Resultat numerisch mit einer Monte Carlo Simulation. Visualisieren Sie die Parameter a_0 und a_1 in einem Scatter-Plot.

c) Bestimmen Sie die Vorhersagen y (Mittelwert und Standardabweichung) für feste $x = -3, 0, +3$ numerisch sowie analytisch und vergleichen Sie diese.