2. Übungsblatt zur Vorlesung
Statistische Methoden der Datenanalyse
Abgabe: 01.11.2018 23:59

WS 2018/2019
Prof. W. Rhode

Zeit	Raum	Abgabe im Moodle; Mails mit Betreff: [SMD1819]
Di. 10-12	CP-03-150	tobias.hoinka@udo.edu, felix.geyer@udo.edu und jan.soedingrekso@udo.edu
Di. 16-18	CP-03-150	simone.mender@udo.edu und alicia.fattorini@udo.edu
Mi. 10-12	CP-03-150	mirco.huennefeld@udo.edu und kevin3.schmidt@udo.edu

Aufgabe 5: *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 Histogramm 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., 51),
         weights=data['hist'])
plt.show()
```

Aufgabe 6: *Zufallszahlengeneratoren* **5 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 $b = 3$ und $m = 1024$. Bestimmen Sie die Periodenlänge in Abhängigkeit des Parameters a, indem Sie für a Werte aus einem angemessenen Bereich verwenden. Stellen Sie den Zusammenhang von Periodenlänge und a in einem Plot dar. Wie groß ist die maximale Periodenlänge? Für welche Werte von a ist die Periodenlänge maximal? Lassen sich die erhaltenen Werte mit den Regeln für *gute* linear-kongruente Generatoren erklären? Hinweis: In dieser Aufgabe sollte der Startwert x_0 unverändert bleiben.

Verwenden Sie für die folgenden Aufgaben einen linear-kongruenten Zufallszahlengenerator mit den Parametern $a = 1601$, $b = 3456$ und $m = 10\,000$.

b) Erzeugen Sie so 10 000 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) Erstellen Sie Histogramme wie in c) und d) auch mit `numpy.random.uniform()`.

e) 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 7: *Zweidimensionale Gaußverteilung* **10 P.**

Eine zweidimensionale Gaußverteilung sei durch folgende Parameter gekennzeichnet:

$$\mu_x = 4, \quad \mu_y = 2, \quad \sigma_x = 3{,}5, \quad \sigma_y = 1{,}5 \quad \text{und} \quad \operatorname{Cov}(x, y) = 4{,}2$$

a) Wie groß ist der Korrelationskoeffizient?

b) Zeigen Sie, dass die Kurven konstanter Wahrscheinlichkeitsdichte Ellipsen sind.

c) Zeichnen Sie die Ellipse, bei der $f(x, y)$ auf das $1/\sqrt{e}$-fache des Maximums abgefallen ist. Zeichnen Sie die Werte μ_x, μ_y, $\mu_x \pm \sigma_x$ und $\mu_y \pm \sigma_y$ in Ihrer Zeichnung ein.

d) Geben Sie eine Rotationsmatrix $\boldsymbol{M}$ an, so dass die Variablen $(x', y')^\top = \boldsymbol{M}(x, y)^\top$ unkorreliert sind. Wie groß sind $\sigma_{x'}$ und $\sigma_{y'}$? Wie lang sind die Hauptachsen der Ellipse und welchen Winkel bilden sie mit den Koordinatenachsen? Zeigen sie hierfür unter anderem, dass

$$\alpha = \frac{1}{2} \arctan\left(-\frac{2 \cdot \operatorname{Cov}(x, y)}{\sigma_x^2 - \sigma_y^2}\right)$$

gilt. Zeichnen Sie $\sigma_{x'}$, $\sigma_{y'}$, und die Hauptachsen in die Zeichnung ein.

e) Wie lauten die bedingten Wahrscheinlichkeitsdichten $f(x \mid y)$ und $f(y \mid x)$? Zeichnen Sie diese Werte in die Zeichnung ein.

f) Wo liegen die bedingten Mittelwerte $E(x \mid y)$ und $E(y \mid x)$? Zeichnen Sie diese Werte in die Zeichnung ein.