3. Übungsblatt zur Vorlesung
Statistische Methoden der Datenanalyse A
Abgabe: 21.05.2019 23:59

SoSe 2019
Prof. W. Rhode

Zeit	Raum	Abgabe im Moodle; Mails mit Betreff: [SMD19]
Do.14–16	CP-03–150	kevin3.schmidt@udo.edu und maximilian.sackel@udo.edu
Fr. 10–12	CP-03–150	tobias.hoinka@udo.edu und noah.biederbeck@udo.edu
Fr. 16–18	CP-03–150	felix.geyer@udo.edu und rune.dominik@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.csv* zu findende) Histogramm gegebene empirische Verteilung. Die Datei enthält Binzentren (*binmid*) und die Höhen (*counts*). Das Histogramm besteht aus 50 Bins zwischen 0,0 und 1,0.

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: *Zufallszahlen verschiedener Verteilungen* **5 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 8: *Zwei Populationen* **5 P.**

Gegeben seien zwei Populationen von jeweils 10 000 Punkten in einer Ebene. Die Population P_0 sei eine zweidimensionale, korrelierte Gaußverteilung mit:

$$\mu_x = 0, \quad \mu_y = 3, \quad \sigma_x = 3{,}5, \quad \sigma_y = 2{,}6 \quad \text{und Korrelation} \quad \rho = 0{,}9$$

Die zweite Verteilung P_1 ist gegeben durch eine Gaußverteilung in x mit

$$\mu_x = 6 \quad \text{und} \quad \sigma_x = 3{,}5,$$

und einer Gaußverteilung in y, deren Mittelwert linear von x abhängt:

$$\mathbf{E}[y|x] = \mu_{y|x} = a + bx \quad \text{mit} \quad a = -0.5, b = 0.6 \quad \text{und} \quad \mathbf{Var}[y|x] = \sigma^2_{y|x} = 1$$

(a) Stellen Sie die beiden Populationen zusammen in einem zweidimensionalen Scatter-Plot dar.

(b) Berechnen Sie die Stichproben-Mittelwerte und -Varianzen von x und y sowie die Stichproben-Kovarianz und den -Korrelationskoeffizienten für die Einzelpopulationen und die Gesamtheit beider Populationen.

(c) Erzeugen Sie eine weitere Population mit den Eigenschaften der Population P_0, diesmal jedoch nur mit 1000 Punkten. Erstellen Sie anschließend ein HDF5-File mit drei Keys und speichern Sie die drei erzeugten Populationen unter eindeutigen Bezeichnern ab. Nutzen sie dafür das Python Paket `pandas`, siehe Python Hands-On.