

# Statistische Methoden der Datenanalyse

Markus Schumacher, Stan Lai, Florian Kiss

## Übung V

23.11.2012, 27.11.2012

### Anwesenheitsaufgaben

#### Aufgabe 34 Erzeugung gleichverteilter Zufallszahlen

Die Monte-Carlo-Methode ist hochgradig abhängig von der Erzeugung von Zufallszahlen. Im Prinzip ließen sich wahre Zufallszahlen mittels physikalischer Prozesse, wie zum Beispiel radioaktive Zerfälle erzeugen. In der Praxis werden allerdings deterministisch erzeugte *Pseudozufallszahlen* benutzt, da diese von einem Rechner erzeugt werden können.

Ein möglicher Algorithmus ist der *linear kongruente Generator*, der gleichverteilte Zahlen  $u_j$  zwischen 0 und 1 erzeugt, und nach folgender Rekursionsformel arbeitet:

$$i_j = (a \cdot i_{j-1} + c) \bmod m, \quad u_j = i_j/m,$$

wobei  $a, c$  und  $m$  positive ganze Zahlen sind. In dieser Übung wollen wir einen solchen Generator selber programmieren und ein paar grundlegende Eigenschaften testen. Gehen Sie dazu wie folgt vor:

- (i) Erstellen Sie ein ROOT-Skript, in dem Sie zuerst vier ganze Zahlen als Variablen bereitstellen:

```
int a = 171;
int c = 1;
int m = 65536;
int seed = 1;
```

- (ii) Erstellen Sie eine `for`-Schleife, in der Sie `nexp` mal wie folgt eine neue Zufallszahl ausrechnen:

```
seed = (a*seed + c) % m;
float zufall=float(seed)/float(m);
```

- (iii) Geben Sie sich die Zufallszahl am Bildschirm aus.

- (iv) Füllen Sie die Zufallszahlen in ein vor der `for`-Schleife erstelltes Histogramm. Zeichnen Sie das Histogramm auf dem Bildschirm. Stimmen die angezeigten Werte für Mittelwert und Varianz<sup>1</sup> mit Ihren Erwartungen überein? Sehen die Zahlen wirklich zufallsverteilt aus?

- (v) Nun soll mit einem einfachen Test die Güte dieses Zufallsgenerators überprüft werden. Dazu stellt man aufeinander folgende Zufallszahlen, z.B. gerade und ungerade in der Abfolge, durch einen Punkt in der  $xy$ -Ebene dar, dessen Koordinaten gegeben sind durch  $(x,y)_i = (u_{2i+1}, u_{2i})$ ,  $i \in \{0,1,2,3, \dots\}$ . Erstellen Sie dazu vor der `for`-Schleife ein Objekt vom Typ `TGraph`:

```
TGraph myscatter=TGraph(nexp);
```

Erzeugen Sie dann innerhalb der `for`-Schleife zwei Zufallszahlen und setzen Sie dann beispielsweise den  $i$ -ten Punkt des `TGraph` wie folgt:

```
myscatter.SetPoint(i,zufall1,zufall2);
```

<sup>1</sup>Das angezeigte RMS ist in ROOT tatsächlich die Wurzel aus der Varianz.

Dabei ist  $i$  die Zählvariable der `for`-Schleife und `zufall11` bzw. `zufall12` die Zufallszahlen. Stellen Sie den Scatterplot graphisch dar mit dem Befehl `myscatter.Draw("AP");`. Sie können auch den Korrelationsfaktor durch die Mitgliedsfunktion `myscatter.GetCorrelationFactor()` ausdrücken.

Ist der verwendete linear kongruente Generator mit den verwendeten Werten ein guter Zufallszahlengenerator? Benutzen Sie stattdessen folgende Werte für die Parameter:

```
int a = 205;
int c = 29573;
int m = 139968;
```

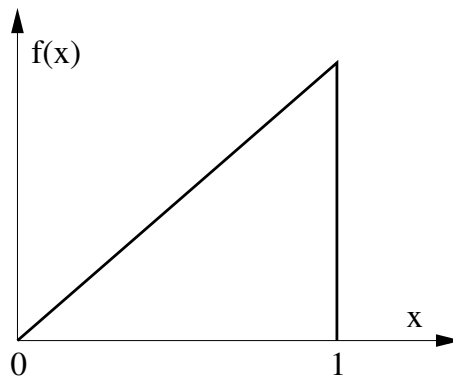
Verbessert sich das Verhalten des Zufallszahlengenerators damit?

- (vi) Vergleichen Sie jetzt diesen einfachen linear kongruenten Generator mit dem in ROOT implementierten Zufallsgenerator `TRandom3`. Ist dieser besser oder schlechter als der selbst programmierte Generator? Im Folgenden werden wir wieder den Zufallszahlengenerator `TRandom3` verwenden.

### Aufgabe 35 Gleichverteilung im Dreieck

Zufallszahlen können benutzt werden, um Funktionen numerisch zu integrieren. Zumeist wird dabei zunächst versucht, im Definitionsbereich einer Funktion gleichmäßig verteilte Zufallszahlen zu erzeugen.

Nehmen wir an, ein solcher Bereich sei durch ein Dreieck mit Steigung 1 gegeben.



Erzeugen Sie mit einem ROOT-Makro Zufallspunkte in diesem Wertebereich.

- (i) Stellen Sie einen Zufallszahlengenerator vom Typ `TRandom3` bereit. Erstellen Sie eine `for`-Schleife. Würfeln Sie darin zunächst die  $x$ -Koordinate gleichverteilt zwischen 0 und 1, würfeln eine andere Zufallszahl zwischen 0 und 1 und transformieren diese so, dass beim Bilden eines  $(x,y)$ -Wertepaares die  $y$ -Koordinate innerhalb des Dreiecks liegt. Prüfen Sie nach, inwiefern die so gebildeten Punkte wirklich gleichmäßig innerhalb des Dreiecks verteilt sind. Benutzen Sie dazu
- eine Punktwolke wie in Aufgabe 34 mittels `TGraph`,
  - ein zweidimensionales Histogramm mittels `TH2F`.

Erstellen Sie diese Objekte jeweils vor der `for`-Schleife, füllen die erhaltenen Zufallszahlenpaare innerhalb der `for`-Schleife hinein und stellen Sie sie nach der `for`-Schleife graphisch dar.

- (ii) Modifizieren Sie das Programm so, dass wenn ein Punkt ausserhalb des Dreiecks gewürfelt wurde, das erhaltene Wertepaar nicht benutzt wird. Verwenden Sie dazu eine `if`-Bedingung. Der Befehl `continue;` bewirkt einen sofortigen Abbruch der jeweiligen `for`-Schleifeniteration. Beachten Sie, dass man entsprechend doppelt so viele Zufallszahlen erstellen muss.
- (iii) Gibt es einen Trick, wie man ohne grossen Aufwand im Dreieck gleichverteilte Zufallspunkte erzeugen kann, ohne Zufallspunkte verwerfen zu müssen?

# Hausaufgaben

## Aufgabe 36 Erwartungstreue

8 Punkte

Gegeben Sei eine Stichprobe  $(x_1, \dots, x_n)$  vom Umfang  $n$ .

- (i) Zeigen Sie, dass der Schätzer für den Mittelwert der wahren Verteilung

$$\hat{\mu} = \frac{\sum_{i=1}^n x_i}{n} = \bar{x}$$

erwartungstreu ist.

- (ii) Nehmen Sie nun an, dass der wahre Wert des Mittelwertes der zugrundeliegenden Wahrscheinlichkeitsdichtefunktion bekannt ist. Zeigen Sie, dass der Schätzer für die Varianz der wahren Verteilung

$$\hat{V}(x) = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2$$

erwartungstreu ist.

- (iii) Nehmen Sie nun an, dass der wahre Wert des Mittelwertes nicht bekannt ist und stattdessen der Schätzer aus (i) verwendet wird. Zeigen Sie, dass die so erhaltene Varianz nicht erwartungstreu ist.
- (iv) Wie muss man den Schätzer für die Varianz modifizieren, um zu einem erwartungstreuen Schätzer zu kommen?
- (v) Zeigen Sie, dass ein erwartungstreuer Schätzer auf  $\langle (x - \mu)^3 \rangle$  gegeben ist durch

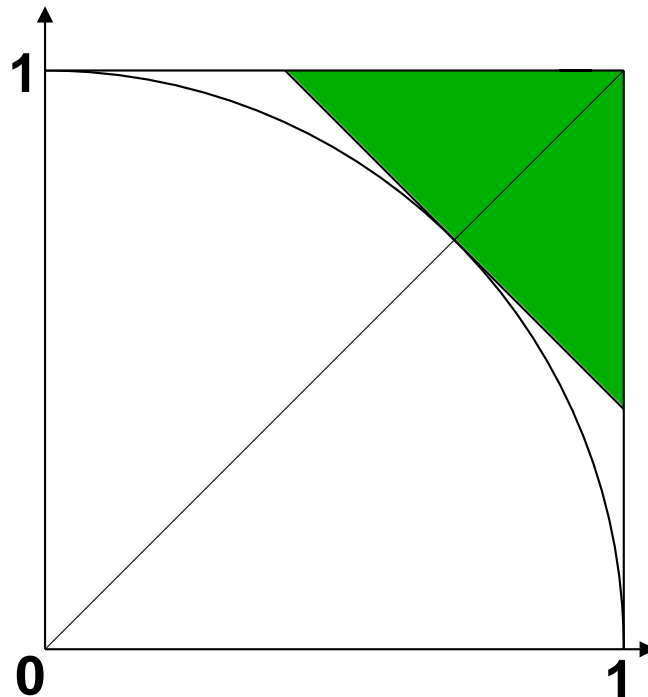
$$\frac{n}{(n-1)(n-2)} \sum_{i=1}^n (x_i - \bar{x})^3.$$

Tipp: Nehmen Sie ohne Beschränkung der Allgemeinheit  $\mu = 0$  an.

## Aufgabe 37 Effiziente Bestimmung von $\pi$

6 Punkte

Eine beliebige Möglichkeit zur Bestimmung der Kreiszahl  $\pi$  besteht im Ausnutzen der Monte-Carlo-Methode. Dazu kann man die von-Neumannsche Akzeptanz-Zurückweisungsmethode benutzen. Betrachtet sei dazu ein Quadrant des Einheitskreises, wie in der folgenden Abbildung skizziert.



Die Bestimmung von  $\pi$  geschieht nun dadurch, dass man im oberen rechten Quadranten  $N_0$  gleichverteilte Paare von Zufallszahlen würfelt. Von diesen liegen dann  $N$  innerhalb des Einheitskreises.

- (i) Geben Sie einen Schätzer  $\hat{\pi}$  für die Kreiszahl  $\pi$  basierend auf  $N$  und  $N_0$  an.

- (ii) Ermitteln Sie den relativen statistischen Fehler  $\Delta\hat{\pi}/\hat{\pi}$  auf diesen Schätzer in Abhängigkeit von  $N_0$  und  $\hat{\pi}$ . (Tipp: Es handelt sich um ein binomial verteiltes Zufallsexperiment! Nehmen Sie an, dass der Schätzer  $\hat{\pi}$  nahe am wahren Wert ist, und ersetzen Sie  $\hat{\pi}$  numerisch durch den bekannten Wert von  $\pi$ , um nur noch eine Abhängigkeit von  $N_0$  zu haben.)
- (iii) Nun besteht die Möglichkeit, nicht im gesamten Viereck zu würfeln, sondern wie skizziert, ein Dreieck oben rechts (symmetrisch bezüglich der Diagonalen) herauszuschneiden und nur innerhalb dieses Polygons gleichverteilte Punkte zu erzeugen. Geben Sie nun wieder einen Schätzer  $\hat{\pi}$  an. Was passiert mit dem relativen Fehler  $\Delta\hat{\pi}/\hat{\pi}$ ?
- (iv) Um die Bestimmung von  $\pi$  weiter zu optimieren, kann man weiterhin ein inneres Quadrat entfernen, welches durch den Berührungspunkt des äusseren Dreiecks an den Kreis sowie den Ursprung definiert werden kann. Geben Sie auch für diesen Fall einen Schätzer  $\hat{\pi}$  an. Was soll mit dem relativen Fehler  $\Delta\hat{\pi}/\hat{\pi}$  passieren? (keine Berechnung!)

**Aufgabe 38** *Integration zweier Gaussverteilungen*

**6 Punkte**

Betrachtet sei eine Wahrscheinlichkeitsdichtefunktion  $f(x)$ , die die Summe zweier Gaussverteilungen, jeweils mit identischem  $\sigma_1 = \sigma_2 = \sigma = 0.2$ , aber verschiedenen Mittelwerten  $\mu_1 = 1$  und  $\mu_2 = 3$ , sein soll:

$$f(x) = \frac{1}{4\sqrt{2\pi} \cdot 0.2^2} \left[ \exp\left(-\frac{(x-1)^2}{2 \cdot 0.2^2}\right) + 3 \exp\left(-\frac{(x-3)^2}{2 \cdot 0.2^2}\right) \right]$$

Benutzen Sie im Folgenden die Akzeptanz-Zurückweisungsmethode.

- (i) Berechnen Sie die Effizienz, wenn man  $f(x)$  im Bereich  $0 < x < 4$  integriert mittels der Akzeptanz-Zurückweisungsmethode mit gleichverteilten Zufallszahlen im Bereich  $0 < x < 4$  und  $0 < y < 1.5$ . Tipp: Die Kumulativfunktion  $\Phi(y)$  der Standardnormalverteilung ergibt sich für praktische Zwecke zu  $\Phi(5) = 1$ .
- (ii) Spalten Sie jetzt diesen Bereich in zwei Bereiche auf:  $0 < x < 2$ ,  $0 < y < 0.5$  und  $2 < x < 4$ ,  $0 < y < 1.5$ . Berechnen Sie die Effizienz für die Integration von  $f(x)$  über diese beiden disjunkten Bereiche.
- (iii) Beschreiben Sie in Worten, wie Sie ein Programm schreiben würden, um die Integration aus (ii) durchzuführen. (Sie brauchen keinen Programmcode aufzuschreiben!)