Albert-Ludwigs-Universität Freiburg Wintersemester 2011/2012

Statistische Methoden der Datenanalyse

Markus Schumacher

Übung V

Markus Warsinsky

28.11.2011

Anwesenheitsaufgaben

Aufgabe 33 *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:

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

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[1] 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 $(x,y)_i = (u_{2i+1},u_{2i})$, $i \in \{0,1,2,3,\ldots\}$. Erstellen Sie dazu vor der `for`-Schleife ein Objekt vom Typ `TGraph`:

```
TGraph myscatter=TGraph(nexp/2);
```

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);
```

Dabei ist `i` die Zählvariable der `for`-Schleife und `zufall1` bzw. `zufall2` die Zufallszahlen. Stellen Sie den Scatterplot graphisch dar mit dem Befehl `myscatter.Draw("AP");`.

[1] Das angezeigte RMS ist in ROOT tatsächlich die Wurzel aus der Varianz.

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;
```

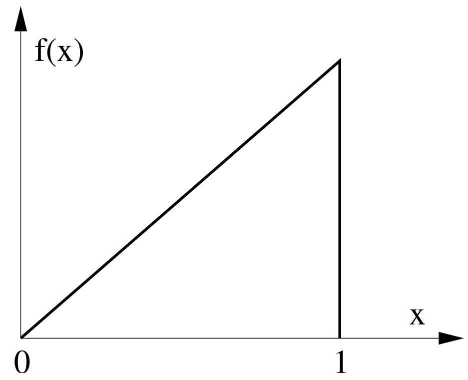
Verbessert sich das Verhalten des Zufallszahlengenerators damit?

(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 34 *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

a) eine Punktwolke wie in Aufgabe 33 mittels `TGraph`,

b) 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 35 *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!)

Aufgabe 36 *Integration der Planck-Verteilung* **7 Punkte**

Wir betrachten die Erzeugung von Photonen, die dem Gesetz von der Strahlung des Schwarzen Körpers folgen. Die Verteilung der skalierten frequenz $x = h\nu/kT$ ist dann gegeben durch die Plancksche Strahlungsformel:

$$f(x) = c\frac{x^3}{e^x - 1}$$

wobei c eine Normierungskonstante ist. Im folgenden soll besprochen werden, wie man diese Funktion möglichst effizient zwischen $x = 0$ und einem bestimmten x_{max} integrieren könnte.

(i) Warum lässt sich die Transformationsmethode nicht anwenden?

(ii) Skizzieren Sie den Verlauf der Funktion.

(iii) Ermitteln Sie einen Ausdruck für die Lage des Maximums und dessen Höhe. Die erhaltenen Gleichungen lassen sich nur numerisch lösen. Als Lage des Maximums ergibt sich $x \approx 2{,}82$ mit $f_{\text{max}} \approx 1{,}42c$.

(iv) Beschreiben Sie, wie man nach der Anzeptanz-Zurückweisungsmethode vorgehen würde, um die Funktion zu integrieren bzw. Zufallszahlen zu erzeugen. Wie ist die Effizienz gegeben? Warum wird dieses Verfahren für hohe x sehr ineffizient?

(v) Nun soll ein Ansatz mit einer stückweise definierten Majorante gemacht werden. Im unteren Bereich bis zu einem Punkt x_1 machen wir den Ansatz $g_I(x) = f_{\text{max}}$, oberhalb von x_1 hingegen den folgenden Ansatz:

   a) Betrachten Sie zunächst das asymptotische Verhalten von $f(x)$ für große x. Welchen Einfluss hat der x^3-Term?

   b) Warum könnte der Ansatz

   $$g_{II}(x) = K \cdot c \cdot x^{-\epsilon} \exp(-x^{1-\epsilon})$$

   ein geeigneter Ansatz für eine Majorante sein?

   c) Ermitteln Sie die Transformationsvorschrift, die zu dieser Majorante gehört.

   d) Beschreiben Sie nun noch das allgemeine Verfahren für die Erzeugung von Zufallsvariablen bzw. zur Monte-Carlo-Integration. Wir wird entschieden, ob eine Zufallszahl $x < x_1$ oder $x > x_1$ genommen wird?

   e) Wie groß ist der Effizienzgewinn durch diese Majorante im Gebiet $x > x_1$ für $c = 1$, $K = 200$, $\epsilon = 0{,}1$, $x_{max} = 20$, $x_1 = 6$?

Aufgabe 37 *Fehler auf die Varianz* **7 Punkte**

In Aufgabe 30 wurde gezeigt, dass die Stichprobenvarianz

$$s^2 = \widehat{\sigma^2} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

ein erwartungstreuer Schätzer der Varianz σ^2 ist. Zeigen Sie, dass die Varianz von s^2 gegeben ist durch

$$V[s^2] = E[s^4] - (E[s^2])^2 = \frac{1}{n}\left(\mu_4 - \frac{n-3}{n-1}\mu_2^2\right),$$

wobei $\mu_k = E[(x-\mu)^k]$ das k-te zentrale Moment von x ist. Gehen Sie dzu wie folgt vor:

(i) Zeigen Sie, dass s^2 geschrieben werden kann als:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n} x_i^2 - \frac{1}{n(n-1)}\sum_{i,j=1}^{n} x_i x_j$$

(ii) Zeigen Sie dann, dass der Erwartungswert von s^4 geschrieben werden kann als:

$$E[s^4] = \frac{1}{(n-1)^2}\sum_{i,j=1}^{n} E[x_i^2 x_j^2] - \frac{2}{n(n-1)^2}\sum_{i,j,k=1}^{n} E[x_i x_j x_k^2] + \frac{1}{n^2(n-1)^2}\sum_{i,j,k,l=1}^{n} E[x_i x_j x_k x_l]$$

(iii) Zählen Sie, wieviele Terme in jeder Summe die algebraischen Momente μ_4' oder $\mu_2'^2$ mindestens in der ersten Potenz enthalten. Beachten Sie, dass alle anderen Terme mindestens eine Potenz von μ enthalten.

(iv) Drücken Sie das Resultat in Form der zentralen Momente μ_2 und μ_4 aus, indem Sie $\mu = 0$ setzen.

(v) Ziehen Sie von dem Resultat $(E[s^2])^2$ ab. $E[s^2]$ sollte Ihnen bekannt sein, da s^2 erwartungstreu ist.

(vi) Ermitteln Sie die Varianz von s^2 im Fall einer Gaussverteilung. Benutzen Sie dazu, dass das 4-te zentrale Moment der Gaussverteilung gegeben ist durch $\mu_4 = 3\sigma^4$.