

# Statistische Methoden der Datenanalyse

Markus Schumacher, Stan Lai, Florian Kiss

## Übung II

5.11.2013, 6.11.2013

### Anwesenheitsaufgaben

#### Aufgabe 9 *Wechseln oder nicht — mal simuliert!*

In einer Spielshow gibt es drei Türen. Hinter einer Tür wartet ein Gewinn, hinter den beiden anderen Türen befinden sich Niete. Nachdem der Kandidat eine Tür ausgewählt hat, öffnet der Moderator eine Tür mit einer Niete (aber nicht die Tür, die der Kandidat gewählt hat). Der Kandidat darf erneut eine der beiden übrigen Türen wählen.

In dieser Programmierübung wollen wir diese Ihnen bereits bekannte Aufgabenstellung mithilfe einer Simulation bearbeiten. Dazu wollen wir die Spielshowsituation  $N$  mal nacheinander zufällig nachstellen. Dazu erst folgende Grundbetrachtungen:

- Wir nehmen an, dass die drei Türen als Tür 0, Tür 1 und Tür 2 bezeichnet seien, so dass wir eine Beziehung zwischen den ganzen Zahlen und den Türen erhalten.
- Ohne Beschränkung der Allgemeinheit wähle der Kandidat jedesmal Tür 0.
- Die wahre Platzierung des Preises wird in jedem Zufallsexperiment zufällig ermittelt.
- Sollte der Kandidat bereits richtig liegen, öffnet der Showmaster zufällig Tür 1 oder Tür 2. Sollte der Preis nicht hinter Tür 0 liegen, so bleibt dem Showmaster selbstverständlich nur eine Alternative.

Gehen Sie nun wie folgt vor:

- (i) Zunächst wollen wir die Spielshow nur einmal nachstellen. Programmieren Sie ein ROOT-Skript mit den folgenden Komponenten:
  - a) Einen Zufallszahlengenerator vom Typ `TRandom3`.
  - b) Ermitteln Sie mittels dieses Zufallszahlengenerators die wahre Platzierung des Preises als Zufallszahl zwischen 0 und 2 (`Integer(3)`). Merken Sie sich diese Zahl in einer Variablen `int ngewinn`.
  - c) Erstellen Sie eine weitere ganze Zahl, der Sie zunächst den Wert 0 zuweisen (`int nnotshow=0`). Sollte der Kandidat bereits richtig liegen, ermitteln Sie eine weitere Zufallszahl zwischen 1 und 2 (`Integer(2)+1`), die angibt, welches Tor nicht geöffnet wird und merken Sie sich diese Zahl mittels `nnotshow`. Ob der Kandidat bereits richtig liegt, oder nicht, ermitteln Sie mittels einer `if`-Abfrage. Sollte der Kandidat falsch liegen, so merken Sie sich die Tür, die automatisch geöffnet wird in der Variablen `nnotshow`.
  - d) Gehen Sie nun davon aus, dass der Kandidat die Tür wechselt. Ermitteln Sie, ob der Kandidat gewonnen hat, oder nicht und geben Sie sich das Resultat auf dem Bildschirm aus. (Tipp: Vergleichen Sie `nnotshow` mit `ngewinn`.)
- (ii) Nun möchten wir dieses Experiment  $N$  ausführen, für den Anfang z.B.  $N = 10$ . Gehen Sie dazu wie folgt vor:
  - a) Erstellen Sie sich am Anfang des Skriptes eine Variable `int npoints=10`.

- b) Erstellen Sie eine `for`-Schleife, die `npoints` mal das Experiment durchführt. Stellen Sie sicher, dass der Zufallszahlengenerator ausserhalb der `for`-Schleife erstellt wird, da sonst bei jedem Zufallsexperiment identische Zufallszahlen erstellt werden. Lassen Sie wie bereits zuvor für jedes Zufallsexperiment am Bildschirm ausgeben, ob der Kandidat gewonnen hat, oder nicht.
- (iii) Ändern Sie nun den Startwert für den Zufallszahlengenerator mittels `SetSeed`. Beachten Sie, dass sich die Abfolge der Ergebnisse ändert.
- (iv) Nun wollen wir ermitteln, wie oft der Kandidat bei dieser Strategie (Wechseln) im Mittel gewinnen würde. Dazu wollen wir zunächst nur eine Zahl ermitteln, den Mittelwert der Stichprobe  $x_1, \dots, x_N$ , wobei die  $x_i$  entweder 0 (verloren) oder 1 (gewonnen) sein können. Erstellen Sie sich dazu ausserhalb der `for`-Schleife eine Fließkommazahl `float mittelwert=0;`. Zählen Sie für jedes Zufallsexperiment zu dieser Zahl 1 hinzu, wenn der Kandidat gewinnt und 0, falls nicht. Dividieren Sie am Ende durch die Anzahl der Zufallsexperimente und lassen Sie sich den erhaltenen Mittelwert am Bildschirm ausgeben. Erhöhen Sie `npoints`, was beobachten Sie?
- (v) Letztendlich wollen wir noch den Mittelwert in Abhängigkeit von der Stichprobengröße graphisch darstellen. Dazu machen wir uns zunutze, dass man den Mittelwert bei Vergrößerung einer Stichprobe aus dem alten Mittelwert und der Größe der Stichprobe ermitteln kann:

$$\bar{x}_N = \bar{x}_{N-1} + \frac{x_n - \bar{x}_{N-1}}{N}$$

Gehen Sie nun wie folgt vor:

- a) Benutzen Sie diese Rekursionsformel um an jedem Punkt der `for`-Schleife den aktuellen Mittelwert zu berechnen.
- b) Erstellen Sie ausserhalb der `for`-Schleife ein Histogramm mit `npoints` bins zwischen 0 und `npoints`.
- c) Benutzen Sie innerhalb der `for`-Schleife die Methode `SetBinContent(i,mittelwert)`, um das `i`-te Bin auf den jeweils aktuellen Mittelwert zu setzen.
- d) Zeichnen Sie schliesslich das Histogramm am Ende des Skriptes. Probieren Sie verschiedene Werte für `npoints`. Ändern Sie den Startwert des Zufallszahlengenerators. Was beobachten Sie?

### Aufgabe 10 *Binomialverteilung*

In dieser Programmierübung soll die sogenannte Binomialverteilung zufällig erzeugt werden, indem flach verteilte Zufallszahlen zwischen 0 und 1 erzeugt werden. Die Binomialverteilung beschreibt die Wahrscheinlichkeit, bei  $N$  Versuchen, wobei bei jedem Versuch ein bestimmtes Ereignis mit einer Wahrscheinlichkeit  $\mu$  auftritt genau  $n$  mal einen Erfolg zu haben. Als Beispiel kann man sich die Wahrscheinlichkeit vorstellen, bei  $N = 20$  Würfelwürfen genau  $n = 5$  mal eine sechs zu würfeln ( $\mu = 1/6$  bei einem normalen Würfel).

Um die Binomialverteilung zu erzeugen gehen Sie nun wie folgt vor:

- (i) Zunächst wollen wir das Zufallsexperiment nur einmal durchführen. Erstellen Sie ein ROOT-Skript, das die folgenden Aufgaben durchführt:
  - a) Erstellen eines Zufallszahlengenerators vom Typ `TRandom3`.
  - b) Definition einer ganzen Zahl `int nwuerfe=20;`
  - c) Definition einer Fließkommazahl `float mu=1./6.;`
  - d) Definition einer ganzen Zahl `int nsechs=0;`, die benutzt werden soll, um sich die Anzahl der Erfolge zu merken.
  - e) Erstellen Sie dann eine `for`-Schleife, die genau `nwuerfe` mal mittels der `Uniform()`-Funktion des Zufallszahlengenerators eine Zufallszahl zwischen 0 und 1 erstellt. Sollte diese Zufallszahl kleiner sein als `mu`, so wird `nsechs` um eins erhöht. Geben Sie nach der `for`-Schleife das Resultat für `nsechs` am Bildschirm aus.
- (ii) Nun wollen wir dieses Zufallsexperiment sehr oft durchführen. Gehen Sie dazu wie folgt vor:
  - a) Erstellen Sie eine weitere `for`-Schleife, die alle anderen Schritte, die zuvor ausgeführt wurden beispielsweise 100 mal ausführt. Beachten Sie allerdings, dass Sie weiterhin den Zufallszahlengenerator vor dieser `for`-Schleife erstellen, da sich sonst die Sequenz der Zufallszahlen in jedem Experiment wiederholen würde. Geben Sie weiterhin für jedes Zufallsexperiment den Wert von `nsechs` am Bildschirm aus, um zu testen, dass sich dieser ändert.

- b) Erstellen Sie nun außerhalb beider `for`-Schleifen ein Histogramm mit `nwuerfe+1` Bins von `-0.5` bis `nwuerfe+0.5`. Füllen Sie für jedes Zufallsexperiment den Wert von `nsechs` in das Histogramm. Normieren Sie das Histogramm auf eins und stellen Sie es am Bildschirm dar.
- c) Verändern Sie nun einige der Parameter. Was passiert, wenn man `nwuerfe` erhöht?

# Hausaufgaben

## Aufgabe 11 *Rekursive Mittelwert- und Varianzberechnung*

5 Punkte

Gegeben sei eine Stichprobe  $(x_1, \dots, x_n)$ . Zeigen Sie:

- (i) Wenn  $\bar{x}_n$  der Mittelwert der ersten  $n$  Elemente ist, dann gilt:

$$\bar{x}_n = \bar{x}_{n-1} + \frac{x_n - \bar{x}_{n-1}}{n}$$

- (ii) Wenn  $\sigma_n^2$  die Varianz der ersten  $n$  Elemente ist, dann gilt:

$$\sigma_n^2 = \frac{(n-1)\sigma_{n-1}^2 + (x_n - \bar{x}_n)(x_n - \bar{x}_{n-1})}{n}$$

Diese rekursive Berechnung vermeidet numerische Instabilitäten bei sehr großen Stichproben.

## Aufgabe 12 *SC Freiburg*

5 Punkte

In der Saison 2012-13, hat der SC Freiburg den 5. Platz in der Tabelle erreicht, und hat sich sogar für die Europa-Liga qualifiziert. In der Bundesliga spielt man 34 mal pro Jahr, und die Torverteilung des SC Freiburg ist in der folgenden Tabelle zusammengefasst:

Tore $k$ im Spiel	Anzahl der Spiele $n_k$ mit $k$ Tore erzielt
0	9
1	13
2	6
3	5
4	0
5	1
$\geq 6$	0

- (i) Berechnen Sie die relativen Häufigkeiten der Anzahl der Tore pro Spiel für den SC Freiburg. Was ist die durchschnittliche Anzahl der Tore pro Spiel?
- (ii) Nehmen Sie diese mittlere Anzahl der Tore  $k$  als Mittelwert einer Poissonverteilung an und berechnen Sie die Wahrscheinlichkeiten  $p_k$  für  $k$  Tore pro Spiel im Jahr 2012-13. Wieviele Spiele mit  $k$  Toren würde man gemäß dieser Wahrscheinlichkeiten erhalten?
- (iii) Können Sie sich vorstellen, warum die Anzahl der Tore pro Spiel nicht unbedingt von einer Poissonverteilung beschrieben wird?

## Aufgabe 13 *Marginal- und bedingte Wahrscheinlichkeitsdichtefunktionen*

5 Punkte

Betrachten Sie die multivariate Wahrscheinlichkeitsdichtefunktion (WDF):

$$f(x,y) = \begin{cases} x+y & 0 < x < 1 \text{ und } 0 < y < 1 \\ 0 & \text{sonst.} \end{cases}$$

- (i) Sind  $x$  und  $y$  unabhängig? Erläutern Sie Ihre Antwort.
- (ii) Wie lauten die marginale WDFs  $f_x(x)$  und  $f_y(y)$ ? Stellen Sie Ihre Formen in einer einfachen Skizze dar.
- (iii) Wie lauten die bedingte WDFs  $g(x|y)$  und  $h(y|x)$  und wie sind die mit Bayes Theorem verbunden? Zeigen Sie dass diese Beziehung konsistent mit den marginalen WDFs ist, die Sie im vorherigen Teil gefunden haben.

**Aufgabe 14** *Mellinfaltung***5 Punkte**

Gegeben seien zwei unabhängige Zufallsvariablen  $x$  und  $y$ , die beide gleichverteilt sind zwischen 0 und 1, d.h. ihre Wahrscheinlichkeitsdichtefunktion ist gegeben durch

$$g(x) = \begin{cases} 1 & 0 < x < 1 \\ 0 & \text{sonst} \end{cases} \quad (1)$$

und analog durch  $h(y)$ . Benutzen Sie die Mellin-Faltung, um zu zeigen, dass die Wahrscheinlichkeitsdichtefunktion  $f(z)$  von  $z = xy$  gegeben ist durch:

$$f(z) = \begin{cases} -\ln z & 0 < z < 1 \\ 0 & \text{sonst.} \end{cases} \quad (2)$$