Vorbesprechung: 17/18. April 2013

Aufgabe 1

Für grossangelegte Simulationen müssen im allgemeinen **pseudozufällige** Zahlen generiert werden; diese Zahlen heissen pseudozufällig, da sie mit Hilfe eines Algorithmus erzeugt werden und daher nicht "wirklich" zufällig sind. Angenommen wir möchten die Performance von Warteschlangennetzwerken beurteilen mit Hilfe einer Simulation, dann müssen wir zufällige Zeitintervalle zwischen den Ankünften von Kunden generieren. Wir nehmen an, diese Zeitintervalle folgen einer Exponentialverteilung.
Verfügen wir über keinen Zufallsgenerator für exponentialverteilte Zufallszahlen, dann können wir exponentialverteilte Zufallszahlen mit Hilfe von gleichmässig im Intervall $[0, 1]$ verteilten Zufallszahlen erzeugen, und zwar mit folgender Überlegung: Sei U eine uniform auf dem Intervall $[0, 1]$ verteilte Zufallsvariable, und sei $X = F_X^{-1}(U)$, wobei $F_X(x) = 1 - e^{-\lambda\, x}$ die kumulative Verteilungsfunktion der Exponentialverteilung ist. Dann gilt

$$P(X \leq x) = P(F_X^{-1}(U) \leq x) = P(U \leq F_X(x)) = F_U(F_X(x)) = F_X(x)\ ,$$

wobei wir im letzten Schritt benutzt haben, dass die kumulative Verteilungsfunktion der uniformen Verteilung gegeben ist durch $F_U(u) = u$, falls $u \in [0, 1]$. Die Zufallsvariable, die durch $X = F_X^{-1}(U)$ definiert wurde, folgt also einer Exponentialverteilung.

(**a**) Lösen Sie $F_X(x) = 1 - e^{-\lambda\, x} = u$ nach x auf, d.h., bestimmen Sie die Funktion $F_X^{-1}(\,\cdot\,)$. Generieren Sie nun Zufallszahlen $X \sim \mathrm{Exp}(\lambda = 2)$ mit $F_X^{-1}(U)$, wobei U uniform auf $[0, 1]$ verteilt ist. *Hinweis:* Konsultieren Sie Serie 2 Aufgabe 2 für den **R**−Befehl `runif`.

(**b**) Überprüfen Sie nun mit einem qq-Plot, ob die in Teilaufgabe (a) generierten Zahlen exponentialverteilt sind.
R-Hinweis:

```
## Bestimmen der theoretischen Quantilen der Exponentialverteilung mit
## lambda=2: wenn z.B. n=100 Datenpunkte vorliegen, dann werden die
## theoretischen Quantile generiert mit
n <- 100
qexp((seq(1,n,by=1)-0.5)/n,rate=2)
## die empirischen Quantilen fuer den qq-plot erhalten Sie, indem Sie die
## Datenpunkte (dargestellt als Kompoenenten eines Vektors x) der Groesse
## nach ordnen, und zwar mit der Funktion
sort(x)
```

Aufgabe 2

Ein Statistiker beobachtet, dass ein Angler innerhalb von 2 Stunden 15 Fische fängt. Er nimmt an, dass es sich um einen Poissonprozess handelt und überlegt sich:

(a) Mit welcher Wahrscheinlichkeit dauert es länger als 12 Minuten, bis der nächste Fisch anbeisst?

(b) Mit welcher Wahrscheinlichkeit beissen innerhalb der nächsten 12 Minuten genau 2 Fische an?

Aufgabe 3

(a) Gegeben sind zwei unabhängige Zufallsvariable X und Y mit den Kennwerten $\mu_X = 40$, $\sigma_X = 15$, $\mu_Y = 85$ und $\sigma_Y = 18$. Berechnen Sie $\mathrm{E}(X + 2Y)$, $\mathrm{Var}(X + 2Y)$ und $\mathrm{E}(X^2)$.

(b) Ein Werk produziert rechteckige Glasscheiben, deren Länge X und Breite Y (in mm gemessen) voneinander unabhängig produktionsbedingten Schwankungen unterliegen. Es gilt $\mu_X = 1000$, $\sigma_X = 0.02$, $\mu_Y = 500$, $\sigma_Y = 0.01$. Wie gross sind Erwartungswert und Standardabweichung des Umfangs U ?

(c) (*Zusatzaufgabe*) Bestimmen Sie die Wahrscheinlichkeitsdichte $f_X(x)$ der Zufallsvariablen $X = Z^2$, wobei $Z \sim \mathcal{N}(0, 1)$.

Aufgabe 4

In dieser Aufgabe untersuchen Sie die Wirkung des Zentralen Grenzwertsatzes mittels Simulation. Gehen Sie von einer Zufallsvariablen X aus, die folgendermassen verteilt ist: die Werte 0, 10 und 11 werden je mit einer Wahrscheinlichkeit $\frac{1}{3}$ angenommen. Simulieren Sie nun die Verteilung von X sowie die Verteilung des Mittelwerts $\overline{X}$ von mehreren X.

(a) Simulieren Sie X. Stellen Sie die Verteilung von X mittels eines Histogramms von 1000 Realisierungen von X dar, und vergleichen Sie sie mittels des Normalplots mit der Normalverteilung.

```
> par(mfrow=c(4,2))     # Mehrere Grafiken neben- und untereinander
> werte <- c(0,10,11)  # moegliche Werte von X
> sim <- sample(werte,1000, replace = TRUE)  # X simulieren
> hist(sim, main=paste("Original"))            # Histogramm erstellen
> qqnorm(sim)           # Normalplot erstellen
```

(b) Simulieren Sie nun $\overline{X} = \frac{X_1+X_2+X_3+X_4+X_5}{5}$, wobei die X_i die gleiche Verteilung haben wie X und unabhängig sind. Stellen Sie die Verteilung von $\overline{X}$ anhand von 1000 Realisierung von $\overline{X}$ dar, und vergleichen Sie mit der Normalverteilung.

```
> n<-5
> sim<-matrix(sample(werte,n*1000,replace=TRUE),ncol=n)
>      # X_1,...,X_n simulieren und in einer n-spaltigen Matrix
>      # (mit 1000 Zeilen) anordnen
> sim.mean<- apply(sim,1,"mean")   #In jeder Matrixzeile Mittelwert berechnen
> hist(sim.mean)
> title(paste("Mittelwerte von",n,"Beobachtungen"))
> qqnorm(sim.mean)
```

(**c**) Simulieren Sie nun die Verteilung von $\overline{X}$ auch für die Fälle, wo $\overline{X}$ das Mittel von 10 resp. 200 X_i ist.