Statistik für Informatiker **SS 2017**
Prof. Dr. Matthias Birkner
M.Sc. Frederik Klement

Blatt 7

Aufgabe 1 (2 + 2 + 2 Punkte)
Betrachten Sie folgendes R-Programm:

```
FINDMAX
INPUT: Unsortierter Vektor x=(x1,x2,...,xn) mit n>=2
OUTPUT: max(x1,x2,...,xn)
```

```
1 FindMax<-function(x){
2   max<-x[1]
3   for(i in 2:length(x)){
4     if(max<x[i]) max<-x[i]
5   }
6   return(max)}
```

Für $n \in \mathbb{N}$ sei S_n die Menge aller Permutationen der Zahlen $1, 2, ..., n$, d.h.

$$S_n := \{\sigma : \{1, 2, ..., n\} \to \{1, 2, ..., n\} : \sigma(i) \neq \sigma(j) \text{ für } i \neq j\}.$$

Sei σ eine zufällige Permutation, welche uniform auf S_n verteilt ist. Sei M die Zahl der Vertauschungen in der Zeile 4. Berechnen Sie den Erwartungswert $\mathbb{E}[M]$, wenn der Inputvektor gegeben ist durch $X = (\sigma(1), \sigma(2), ..., \sigma(n))$.

a): Argumentieren Sie zuerst, dass in der i-Iteration der Schleife eine Vertauschung stattfindet, falls $\sigma(i+1) > \max\{\sigma(1), \sigma(2), ..., \sigma(i)\}$.

b): Bestimmen Sie $P(\sigma(i+1) > \max\{\sigma(1), \sigma(2), ..., \sigma(i)\})$ für $i \in \{1, 2, ..., n-1\}$.

c): Argumentieren Sie, dass $M = \sum_{i=1}^{n-1} I_{\{\sigma(i+1) > \max\{\sigma(1),\sigma(2),...,\sigma(i)\}\}}$ und berechnen Sie mithilfe dieser Darstellung und den Wert $\mathbb{E}[M]$.
Hinweis: Sei A ein Ereignis, dann gilt $\mathbb{E}[I_A] = P(A)$.

Aufgabe 2 (3 + 2 Punkte)
Seien $(X_i)_{i \in \mathbb{N}}$ eine Folge von unabhängigen, geometrisch verteilten Zufallsvariablen mit $X_i \sim \mathrm{Geom}_p$ mit $p \in (0, 1)$ für $i \in \mathbb{N}$. Sei

$$S_n = \sum_{i=1}^{n} X_i.$$

a): Zeigen Sie, dass S_n negativ binomialverteilt ist mit $S_n \sim \mathrm{NegBin}_{n,p}$, d.h. für alle $k \in \mathbb{N}_0$ gilt:

$$P(S_n = k) = \binom{n+k-1}{k} p^n (1-p)^k.$$

Hinweis: Sie können vollständige Induktion und die Faltungsformel verwenden oder Sie argumentieren mit einer geeigneten Münzwurffolge wie in Aufgabe 2 von Blatt 3.

b): Seien nun S_n und S'_m zwei unabhängige Zufallsvariablen mit $S_n \sim \mathrm{NegBin}_{n,p}$ und $S'_m \sim \mathrm{NegBin}_{m,p}$ für $n, m \in \mathbb{N}$. Argumentieren Sie, dass für $\tilde{S} := S_n + S'_m$ gilt $\tilde{S} \sim \mathrm{NegBin}_{n+m,p}$.

Aufgabe 3 (1 + 2 + 2 Punkte)
Für die Inversionsmethode benötigt man die Verteilungsfunktion, falls man jedoch nur die Dichte kennt, kann man die Verwerfungsmethode verwenden. Für die Verwerfungsmethode brauchen wir eine zweite, einfachere Dichte g und eine Konstante $C \geq 1$, so dass für alle $x \in \mathbb{R}$ gilt

$$f(x) \leq Cg(x). \tag{1}$$

Sei z.B. $f_{n,m}$ die Dichte der $\beta_{n,m}$-Verteilung für $n, m \in \mathbb{N}$ und g die Dichte der Uniformverteilung auf $(0, 1)$, d.h.

$$f_{n,m}(x) = \frac{1}{B(n,m)} x^{n-1}(1-x)^{m-1}\mathbf{1}_{(0,1)}(x), \qquad g(x) = \mathbf{1}_{(0,1)}(x),$$

mit $B(n, m) = \int_0^1 x^n(1 - x)^m dx$, wobei $B(n, m)$ die Euler'sche Betafunktion ist. Dann gilt (1) für $C := \max\limits_{x \in (0,1)} f_{n,m}(x)$. Um nun eine Zufallsvariable mit Dichte f zu erhalten, gehen wir wie folgt vor:

1: Wir erzeugen eine Folge von unabhängigen Zufallsvariable $(U_1, X_1), (U_2, X_2), ...$, wobei $U_1, U_2, ...$ uniform auf $(0, 1)$ verteilt sind und $X_1, X_2, ...$ die Dichte g besitzen.

2: Dann bestimmen wir $K = \min\{k : U_k \leq \frac{f(X_k)}{Cg(X_k)}\}$ und geben X_K zurück.

a): Schreiben Sie die Funktion `beta(n,m,x)`, welche den Wert $f_{n,m}(x)$ berechnet. Verwenden Sie hierbei $B(n, m) = \frac{\Gamma(n)\Gamma(m)}{\Gamma(n+m)} = \frac{(n-1)!(m-1)!}{(m+n-1)!}$. Verwenden Sie `factorial`.

b): Schreiben Sie die Funktion `rejection(n,m,C)`, die Zufallsvariable mit der Dichte $\beta_{n,m}$ erzeugt mithilfe der Verwerfungsmethode. Die Funktion soll neben X_K auch K zurückgeben. Hierbei steht `C` für die verwendete Konstante und g soll stets die Dichte der Uniformverteilung auf $(0, 1)$ sein.

c): Erzeugen Sie zweimal 10^4 Stichproben von $\beta_{4,7}$, einmal für die Konstante $C_1 = 3$ und $C_2 = 5$. Plotten Sie für beide Konstanten jeweils ein Histogramm mit `freq=F` der erzeugten Stichproben und geben Sie in der Überschrift den Mittelwert $\hat{k}$ von K an. Zeichnen Sie auch die Dichte von $\beta_{4,7}$. Plotten Sie beide Histogramme in ein Bild mit `par(mfrow=c(2,1))`.

Achtung: Zusätzlich zur elektronischen Abgabe per Mail, drucken Sie bitte Ihren Code aus und reichen Sie diesen zusammen mit Ihrer restlichen Abgabe ein.

Abgabe: 10 Uhr 16.06.2017