Statistik für Informatiker **SS 2017**

Prof. Dr. Matthias Birkner
M.Sc. Frederik Klement

Blatt 0

Dieses Blatt wird nicht abgegeben und wird nicht bepunktet. Es dient als Einstieg und wird in der zweiten Woche des Semesters in den Übungen besprochen. Versuchen Sie die Aufgaben mithilfe Ihrer bisherigen Kenntnisse zu lösen, im Laufe der Vorlesung werden wir erst noch den passenden formalen mathematischen Apparat entwickeln. Dennoch wollen wir Ihnen besonders die Aufgaben 1 und 5 ans Herz legen, um sich mit der Programmiersprache **R** vertraut zumachen.

Aufgabe 1
Installieren Sie **R** und gegebenfalls R-Studio oder einen anderen Editor Ihrer Wahl.

Aufgabe 2
Ein Programm erzeugt einen zufälligen String der Länge 12. Jeder Character wird hierbei zufällig und unabhängig von den restlichen Charactern aus der Menge $\{\texttt{H}, \texttt{e}, \texttt{l}, \texttt{o}, \texttt{W}, \texttt{r}, \texttt{d}, !, \_\}$ gezogen, wobei jedes Element diesselbe Wahrscheinlichkeit besitzt und mehrmals gezogen werden kann. Mit welcher Wahrscheinlichkeit werden folgende Strings gebildet:

a) `Hello_World!` b) `orW_eoH!_!Hl` c) `rdrdrdrdrdrd`

Aufgabe 3
Sie betrachten zwei primitive Suchalgorithmen **S1** und **S2**, denen ein Feld V der Länge N übergeben wird und ein Element e. Daraufhin wird überprüft, ob das Element in dem Feld V enthalten ist. Beide Algorithmen basieren auf Zufall:

S1: In jedem Schritt wird zufällig ein Index $i \in \{1, 2, ..., N\}$ gewählt, wobei jeder Index diesselbe Wahrscheinlichkeit besitzt, und überprüft, ob es sich bei $V[i]$ um das Element e handelt. Der vorige Schritt wird solange wiederholt, bis ein Eintrag mit dem Element e gefunden wurde.

S2: Ähnlich wie der Algorithmus S1, aber diesmal wählt der Algorithmus zufällig in jeden Schritt einen Index aus, welcher noch nicht in einem vorigen Schritt überprüft worden ist. Der Algorithmus endet, falls er das Element e gefunden hat oder alle Indizes überprüft worden sind.

a) Nehmen Sie an, ihr Feld besitzt N Einträge und k davon enthalten das Element e. Bestimmen Sie für beide Algorithmen die Wahrscheinlichkeit genau nach n Schritten zu enden.

b) Um die Algorithmen zu verbessern, soll vorzeitig abgebrochen werden. Bestimmen Sie in Abhängigkeit von N die kleinste Zahl m, so dass unter der Annahme, $k = 1$, die Wahrscheinlichkeit, dass das Element e nach m Schritten nicht gefunden worde, kleiner oder gleich 0.01 ist.

Aufgabe 4
Betrachten Sie die vier Würfel $W1$, $W2$, $W3$ und $W4$, welche folgende Zahlen auf ihren Seitenflächen $(s_1, s_2, s_3, s_4, s_5, s_6)$ eingraviert besitzen:

$W1$**:** (3,3,3,3,3,3) $W2$**:** (4,4,4,4,0,0) $W3$**:** (5,5,5,1,1,1) $W4$**:** (6,6,2,2,2,2)

Die Wahrscheinlichkeit, dass einer der Würfel auf eine vorgegebene Seite fällt, beträgt jeweils $\frac{1}{6}$.

Ihr Freund und Sie spielen gegeneinander. Zuerst wählen Sie einen Würfel, dann Ihr Freund einen von den verbleibenden. Sie beide machen jeweils einen Wurf und der Spieler mit der höheren Zahl gewinnt. Ihr Wurf und der Ihres Freundes sind unabhängig voneinander. Bestimmen Sie die Wahrscheinlichkeit, dass Sie gewinnen, falls

a) Sie $W1$ und Ihr Freund $W2$ gewählt haben.

b) Sie $W2$ und Ihr Freund $W3$ gewählt haben.

c) Sie $W3$ und Ihr Freund $W4$ gewählt haben.

d) Sie $W4$ und Ihr Freund $W1$ gewählt haben.

Ist es ein Vorteil anzufangen? Wenn Sie Lust haben, können Sie ein kleines R-Programm schreiben, um Ihre Ergebnisse durch Simulationen zu überprüfen. Hierbei könnte der Befehl `sample` nützlich sein.

Aufgabe 5 Sie sollen das folgende Integral berechnen:

$$\int_0^1 \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}\, dt. \tag{1}$$

Nachdem es Ihnen nach einigen Versuchen nicht gelungen ist, einen geschlossenen Ausdruck für eine Stammfunktion von $\frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$ zu finden, greifen Sie nach anderen Mitteln.

a): Schreiben Sie in **R** eine Funktion `norm(t)`, welche den Wert $f(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$ berechnet.

b): Zeichnen Sie mithilfe der Funktion `plot` und `norm(t)` den Graphen der Funktion $t \mapsto f(t)$ auf dem Interval $[0, 1]$. Zerlegen Sie dazu das Interval $[0, 1]$ mithilfe von `seq` in Stützstellen, für welche Sie den Wert von f berechnen. Damit `plot` eine Linie statt Punkten zeichnet, übergeben Sie den Parameter `type="l"`. Übergeben Sie für die Achsen `xlim=c(0,1)` und `ylim=c(0,1)`.

c): Simulieren Sie mit dem Befehl `runif` 1000 Paare $(X_1, Y_1), ..., (X_{1000}, Y_{1000})$ von unabhängigen und auf $[0, 1]$ uniform verteilten Zufallsvariablen. Zeichnen Sie diese Paare als Punkte in den Plot von **b)** mit `points`. Färben Sie dabei die Punkte, welche unter dem Graphen von f liegen rot, die anderen grün. Nutzen Sie hierzu die Parameter `col="red"` und `col="green"`. **Hinweis:** Seien `x` und `y` die Vektoren mit den von Ihnen simulierten Werten. Wenn Sie den Befehl `y<=norm(x)` verwenden, erhalten Sie einen Vektor mit den Indizies der rot zu färbenden Punkte. Das erspart Ihnen eine `for`-Schleife.

d): Die Wahrscheinlichkeit, dass ein Punkt rot wird, ist genau der Wert des Integrals. Schätzen Sie also das Integral mithilfe von $\hat{p} = N/1000$, wobei es sich bei N um die Zahl der roten Punkte handelt.

e): Tatsächlich ist f die Dichte der Standardnormalverteilung, welche in der Vorlesung noch eine bedeutende Rolle spielen wird. Die Funktion `pnorm(x)` berechnet den Wert

$$\int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}\, dt.$$

Berechnen Sie mit `pnorm(x)` das Integral (1) und berechnen Sie den Unterschied zwischen diesem Wert und Ihrer Schätzung $\hat{p}$ aus Teil **d)**.