Prof. Antony Unwin, Alexander Pilhöfer
Lehrstuhl für Rechnerorientierte Statistik und Datenanalyse
Institut für Mathematik
Universität Augsburg
http://stats.math.uni-augsburg.de/

Wahrscheinlichkeitstheorie
Übungsblatt 10

Abgabe: *Donnerstag 19. Januar 2012, bis spätestens 12.00 Uhr*; Briefkasten: WTheorie oder per email

Die Aufgaben können auch in 2er-Gruppen bearbeitet und abgegeben werden!

1. (a) Sei $X \sim Exp(\lambda)$ und $\lambda \sim Exp(\mu)$. Geben Sie eine Formel für die Dichte $g_X(x)$ von X an! **(1P)**
   (b) Sei X eine Stichprobe des Umfangs 9 aus $B(2, 0.5)$. Was ist dann $P(Median(X) \neq 2)$? **(1P)**
   (c) Sei $\Omega = \{1, 2, 5\}$ mit $p_1 = 1/3$, $p_2 = 1/2$ und $p_5 = 1/6$. Bestimmen sie die WEF! **(1P)**
   (d) A. und C. würfeln jeweils 10 Mal. Mit welcher Wskt. gewinnt A. alle Würfe? **(1P)**

2. **Verschollen (5P)**

   Ein Flugzeug fängt während eines Fluges über das Eismeer Feuer. Der Pilot muss mit dem Fallschirm abspringen, da er auf dem Eis nicht notlanden kann. Durch die Erderwärmung ist der Eispanzer in viele Eisschollen zerbrochen. Der Pilot nimmt an, die Größe der Eisschollen G sei exponentialverteilt mit $\lambda = 1/2$, so dass die meisten Eisschollen zu klein sind.

   (a) Welche Größe ist für die Eisscholle, auf der der Pilot landet, zu erwarten?
   (b) Die tatsächliche Tragfähigkeit T (in kg) einer Eisscholle sei als normalverteilt $N(40G, \sigma^2)$ angenommen. Wie muss der 80kg schwere Pilot die Verteilung der Eisschollenfläche revidieren, wenn er lediglich weiß, dass er nicht untergegangen ist?
   (c) Ist es wahrscheinlich, dass der Pilot untergehen wird (ohne explizite Berechnung)?

3. **Simulation (5P)** Als Simulationsmethode wird das folgende Vorgehen in R vorgeschlagen:

```
x <- runif(2)
for(i in 2:9000){
       x[i<-i+1]<- (31*x[i]+13*x[i-1]+3)%% 127
}
```

   (a) Erläutern Sie das Vorgehen! Handelt es sich bei den erzeugten Zahlen tatsächlich um Zufallszahlen?
   (b) Installieren Sie das Paket `rgl` und laden Sie es:

```
install.packages("rgl")
library(rgl)
```

   Erstellen Sie mit Hilfe der Funktion `plot3d(x,y,z)` ein dreidimensionales Streudiagramm für geeignete Teilstichproben von x. Sehen die Daten zufällig aus?
   (Tipp: die Funktion `seq(from, to, by)` ist hilfreich!)

4. **Klausur (5P)**

   Studenten schreiben eine Klausur. Nehmen wir an, dass die Leistung von jedem an einem bestimmten Tag normalverteilt $N(\mu, \sigma^2)$ ist. Um die Prüfung zu bestehen muss jemand mindestens m Punkte erreichen.

   (a) Wie groß ist die Wahrscheinlichkeit, dass jemand die Prüfung besteht, obgleich seine erwartete Leistung μ kleiner als m ist (= m - d)?

   (b) Wie groß ist die Wahrscheinlichkeit, dass jemand durchfällt, obgleich seine erwartete Leistung μ größer als m ist (= m + d)?

   (c) Plotten Sie diese Wahrscheinlichkeiten gegen μ für $m = 40$, und $\sigma^2 = 10$.

5. **Verteilungen (5P)**

   Zwei bedeutsame stetige Verteilungen sind die Betaverteilung und die log-Normalverteilung, die nun betrachtet werden sollen.

   (a) Die Betaverteilung auf $[0, 1]$ mit Parametern p und q ist definiert durch die Dichte

$$
\begin{aligned}
f(x) &= \frac{1}{B(p,q)} x^{p-1}(1-x)^{q-1} \\
B(p,q) &= \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)} = \int_0^1 z^{p-1}(1-z)^{q-1} dz
\end{aligned}
$$

   Wozu dient der Vorfaktor $B(p, q)$? Zeichnen Sie in R die Dichte für die Parametrisierungen $(1, 1)$, $(0.5, 0.5)$, $(4, 2)$ und $(4, 1)$.

   (b) Zeichen Sie in R eine Betaverteilung auf einem allgemeinen Intervall $[a, b]$! Verwenden Sie dazu eine geeignete Transformation von x!

   (c) Geben Sie für die verschiedenen Parametrisierungen eine Abschätzung ($\pi \cdot Daumen$) für die Lage von Erwartungswert und Median an! In welchem Fall ist die Varianz am größten?

   (d) Eine Zufallsvariable X genügt einer log-Normalverteilung, wenn $Y = log(X) \sim N(\mu, \sigma^2)$. Wie sieht die Form der Dichte aus? Zeichnen Sie auch diese in R!

   (e) Betrachten Sie folgende Situationen und Sachverhalte:

   - In einer Studie wurde das Einkommen von Angestellten untersucht.
   - In einer Spielshow müssen die Kandidaten verschiedene Aufgaben in jeweils maximal 2 Minuten lösen. Je schneller sie sind, desto mehr Punkte erhalten sie. Am Ende gewinnt derjenige mit den meisten Punkten.

   Welche der obigen Verteilungen und Parametrisierungen könnte in den beiden Fällen jeweils relevant sein?