1 Probability distributions and R (AJ Chapter 5)

The (*) questions from sheets 1 to 4 form homework 1.

1. Install R and RStudio on your laptop.

2. Introduction to R. Work through the following R instructions and find out what they do.

   (a)
   ```
   print("Hello World!")
   ?rnorm
   #I am a comment
   cat("Hello", "World!",123)
   ```
   (b)
   ```
   1+1
   3/3==1
   3.3/3==1.1
   c(1,2,3)+4
   c(1,2,3)==c(3,2,1)
   ```
   (c)
   ```
   x <- 1:10
   x+1
   x**2
   sin(x)
   sapply(x,sin)
   ```
   (d)
   ```
   x=1:20
   x[4]
   x[-4]sudo apt-get install libglu1-mesa-dev
   x[seq(2,13,2)]
   x[-seq(2,13,2)]
   ```
   (e)
   ```
   #Plotting in R
   x <- seq(0,10,0.1)
   y <- x**2
   plot(x,y)
   plot(x,y,type="l",col=2)
   ```
   (f)
   ```
   #Random numbers, and summary statistics
   x <- rnorm(10**6)
   mean(x)
   summary(x)
   hist(x,40)
   ```
   (g) sudo apt-get install libglu1-mesa-dev
   ```
   #Plot random points in [0,1]^2. Why are there clusters?
   x <- runif(1000)
   y <- runif(1000)
   plot(x,y,pch=15,cex=0.3)
   ```

(h)
```
#Functions, if-statements and loops in R.
parity <- function(x) {
  if (x %% 2 == 0) {
    cat("Number", x, "is even.\n")
  } else {
    cat("Number", x, "is odd.\n")
  }
}

for (x in 1:10) {
  parity(x)
}

x<-1
while(x<=10) {
  cat(x)
  x <- x+1
}

replicate(10,"Hello")
```

(i)
```
#Matrices in R
rbind(1:3,4:6,7:9)
cbind(1:3,4:6,7:9)

n <- 3
a <- matrix(rnorm(n**2),n,n)  # A random matrix
a * a                         # Entry-wise multiplication
a %*% a                       # Matrix multiplication
a[2,]                         # Second row of a
a[,2]                         # Second column of a
a[1:2,1:2]                    # The top left corner of a

n <- 1000
a <- matrix(rnorm(n**2),n,n)  # A bigger random matrix
s <- (a+t(a))/n               # Form a symmetric random matrix:
                              # t(.) means transpose
hist(eigen(s)$values)         # What shape do you get?
```

Continued....

3. Generate 10^6 samples from:

   (a) The normal distribution with mean 2 and standard deviation 3. What percentage are positive? [Check your answers are reasonable probabilistically.]

   (b) The Poisson distribution with mean 3. What percentage are zero?

   (c) The uniform distribution on [-1,2]. What percentage are negative?

   (d) The uniform distribution on the set 1,2,3,4,5,6. What percentage are 6s?

   (e) The lottery distribution, i.e. 6 numbers without replacement from 1,2,...,49. If your £2 ticket matches at least three numbers you win £25. What percentage contain at least three of the numbers 1,2,3,4,5,6 (i.e. what is your probability of winning the prize at least £25 if you bought a ticket with the numbers 1,2,3,4,5,6, or any numbers for that matter by symmetry).

4. (*) Jumping in a triangle: Let T denote the triangle with corners $c_1 = (0,0) \in \mathbb{R}^2$, $c_2 = (1,0) \in \mathbb{R}^2$, and $c_3 = (1/2, \sqrt{3}/2) \in \mathbb{R}^2$. Start at $x_1 = c_1$. For $k = 2, 3, ...,$some big number:

   - Pick one of the three corners uniformly at random: call it C_k.
   - Jump half way from x_{k-1} to the choosen corner, i.e.

   $$x_k = \frac{1}{2}(x_{k-1} + C_k).$$

   (a) What shape do you think you will get?

   (b) Make a plot of the x_k and give the code used to generate it.

   (c) What is the resulting shape called?

   (d) What type of process is this?

5. (*) Estimating the median. Let $X_1, \ldots, X_n \sim f(x;\theta)$ where

   $$f(x;\theta) = g(x - \theta)$$

   and g is a continuous p.d.f. with median zero. Suppose n is odd so that $n = 2r + 1$ for some integer r.

   (a) Use the formula for the $(r+1)$-th out of $2r+1$ order statistic

   $$f_{(r+1)}(x) = \frac{(2r+1)!}{r!r!} F(x)^r (1 - F(x))^r f(x)$$

   to find a (simple) approximation for the distribution of $X_{(r+1)}$. How does your answer depend on g and r?

(b) Use your answer from part (a) construct an approximate confidence interval for the median θ of the $f(\cdot, \theta)$ distribution. i.e. Find statistics $L = L(X_1, \ldots, X_n)$ and $R = R(X_1, \ldots, X_n)$ such that

$$\mathbb{P}(L \leq \theta \leq U) = 0.95.$$

(c) Test your confidence interval using R for two different distributions and a few different values of r. Give the code you used, your test results and any relevant observations.

6. Probability recap: Independence (Ch 1 of AJ's notes)

   (a) Define what it means for two events to be independent.
   (b) Define what it means for n events to be independent.
   (c) Which events are independent of themselves?
   (d) Define what it means for two random variables to be independent.
   (e) Find random independent random variables X and Y such that they have covariance 0 and are *not* independent.
   (f) Find events A, B and C such that (i) A and B are independent (ii) B and C are independent (iii) A and C are independent, and (iv) A,B,C are not independent.
   (g) Find events A, B and C such that $\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C)$ but A, B and C are not independent. If you can make up a story where A, B and C are natural/interesting events, you get bonus points.

7. Read Chapter 5 of AJ's notes.