Practical 1 (9 Feb 2015, Monday)

Question 1

i. Let X be a Bernoulli variable with probability of success p, denoted by $X{\sim}Bernoulli(p)$. Show that $E(X) = p$ and $E(X^2) = p$.

ii. How can we obtain the variance of X, $var(X)$, from $E(X)$ and $E(X^2)$?

Question 2

i. What is the physical meaning of a binomial random variable? How is it related to Bernoulli random variables? Google it if you are not sure.

ii. Let $X{\sim}Binomial(n,p)$. What are the possible outcomes?

iii. Let $X{\sim}Binomial(n,p)$. The probability mass function of X is $f_{X(k)} = \binom{n}{k} p^k (1-p)^{n-k}$. Show that $\sum_{k=0}^{n} f_X(k) = 1$

iv. Now we would like to visualise the binomial pmf on the computer. Try running the following code to plot a binomial pmf in R.

```
x<-0:10

y<-dbinom(x, size=10, prob=0.4)

plot(x, y, pch=16, ylab='probability', xlab='outcome')
```

v. Does the pmf sum up to one? You may use the `sum()` function.

vi. We can generate the outcomes of a binomial random variables using `rbinom()`. If you are not familiar with the function you can always look up the help file. For instance, `?rbinom`. Let us sample 1000 outcomes from the binomial distribution with n=10 and p=0.4 by the following code:

```
x<-rbinom(1000, size=10, prob=0.4)
x
```

vii. You will now see 1000 random numbers on your screen. Actually you will get a different set of random numbers if you rerun the code every time. Let us visualise these numbers with a histogram, `hist()`.

viii. Does the histogram look like the graph obtained in (iv)? We can also calculate the sample mean and variance using `mean()` and `var()` function.

Question 3

i. Let $X \sim Poisson(\lambda)$. Calculate (by hand) the expected value of X, $E(X)$. (Hint: Taylor's expansion may help)

ii. Let $X \sim Poisson(\lambda = 2)$. Remember that Poisson distribution models the number of events happened within a time interval with λ the average rate of occurrence. Plot the pmf in R using the `dpois()` function for x between 0 to 10.

iii. Calculate $Pr(X = 4)$ and $Pr(X \leq 3)$.

Question 4

i. Suppose X follows an exponential distribution with λ. X is a continuous random variable that models the time between events. The probability density function is $f_X(x) = \lambda\exp(-\lambda x)$ with support $[0, \infty)$. Calculate the expected value of X.
$E(X) = \int_0^\infty x f_X(x)dx =$

ii. Show that the cumulative density function of X is $F_X(x) = 1 - \exp(-\lambda x)$.

iii. Let $X \sim Exp(\lambda = 2)$. Plot the pdf of X in R for $0 \leq x \leq 5$ with interval 0.01. Here are some functions that may be useful: `seq()`, `dexp()`

iv. Calculate the probability from pdf. It seems that we have never discussed how we can obtain a probability from a pdf. If X is a continuous random variable, then it is pointless to calculate the probability of X takes a certain number as it is always zero. This is because there are infinitely many outcomes for a continuous random variable and you can never exactly be a particular number. What we can calculate is the probability of X lies between an interval, say, $\Pr(0 \leq X \leq 1)$. In general, the probability of X lies between two numbers, a and b, is the area under the pdf curve: $\Pr(a \leq X \leq b) = \int_a^b f_X(x)dx$. Let us calculate the probability of $\Pr(0 \leq X \leq 1)$ for $X \sim Exp(\lambda = 2)$. Use `integrate()` in R to do this.

```
integrate(dexp, lower=0, upper=1, rate=2)
```

Question 5

i. Let $X \sim N(\mu, \sigma^2)$. We call X the "standard normal" if $\mu = 0$ and $\sigma^2 = 1$. Let us plot the pdf of the standard normal distribution, from $x = -3$ to $x = 3$.

ii. What is $\Pr(2 \leq X \leq 3)$? And what is $\Pr(-1.96 \leq X \leq 1.96)$? Use `integrate()` to help you with this.

iii. Verify you answer with the following commands. Are they the same? So now you don't need statistical tables anymore as they are all built-in in R.

```
pnorm(3)-prnorm(2)

pnorm(1.96)-pnorm(-1.96)
```

Question 6 (Adapted from Mick Crawley GLM course)

The central limit theorem states that for any distribution with a finite mean and variance, then the mean of a random sample from that distribution tends to be a normally distributed.

Let us consider the negative binomial distribution. We draw 1000 negative binomial random numbers with r=1 and p=0.2 and plot the histogram.

```
y<-rnbinom(1000, 1, 0.2)

hist(y)
```

It is far from normal as it skews to a side. The central limit theorem says the mean of samples will follow a normal distribution even for a badly behaved distribution like this. What does it actually mean? Let us consider the following code:

```
# GENERATE 30000 NETAGIVE BINOMIAL RANDOM NUMBERS

y<-rnbinom(30*1000, 1, 0.2)

# PUT THEM IN A 1000-BY-300 MATRIX

y.matrix<-matrix(y, nr=1000)

# SO THERE ARE 30 ELEMENTS IN A ROW, AND I CALCULATE THE MEAN OF
EACH ROW

# SO WE HAVE 1000 OF ROW AVERAGES,

# AND WE PLOT THE HISTOGRAM OF THSES AVERAGES

y.mean<-apply(y.matrix, 1, mean)

hist(y.mean)
```

Does the histogram look more 'normal' now? This is why the normal distribution is so famous and widely used.