1: The hypergeometric distribution

(a) Solve:

A hung jury is one that is unable to reach a unanimous decision. Suppose that a pool of twenty-five potential jurors is assigned to a murder case where the evidence is so overwhelming against the defendant that twenty-three of the twenty-five would return a guilty verdict. The other two potential jurors would vote to acquit regardless of the facts. What is the probability that a twelve-member panel chosen at random from the pool of twenty-five will be unable to reach a unanimous decision?

(b) R commands:

The `dhyper` function returns the hypergeometric probability distribution.

Using proper graphical representations, show the convergence of the hypergeometric distribution to the binomial distribution when N is large, computing and comparing hypergeometric and binomial probabilities for $N \in (50, 100, 1000)$ using a rate of success equal to 0.2.

Suggestion: the `par` function can be used to split a graphical device in two parts:

```
par(mfrow = c(2,1))          #to split the device in two rows
command for the first plot   #plot on the first row of the device
command for the second plot  #plot on the second row of the device
par(mfrow = c(1,1))          #to reset the device in order to show only a plot
```

2: The Poisson distribution

Solve:

(a) Assume that the number of hits, X, that a baseball team makes in a nine-inning game has a Poisson distribution. If the probability that a team makes zero hits is $\frac{1}{3}$, what are their chances of getting two or more hits?

(b) If a typist averages one misspelling in every 3250 words, what are the chances that a 6000-word report is free of all such errors? Answer the question two ways: first, by using an exact binomial analysis, and second, by using a Poisson approximation. Does the similarity (or dissimilarity) of the two answers surprise you? Explain.

R commands:

The `dpois` function returns the Poisson probability distribution. Compute the binomial probabilities using the binomial law and its Poisson approximation for:

- $n = 5$ and $\pi = \frac{1}{5}$,
- $n = 100$ and $\pi = \frac{1}{100}$,

and compare the results in terms of goodness of the approximation.

3: The geometric distribution

Reason:

Let $X \sim Geo(\pi)$. Why $P(X > n) = (1 - \pi)^n$?

Solve:

A couple of dice is thrown until the sum of the results exceeds 6. Let X denotes the number of throws needed for this. Compute $P(X = 7)$.

4: The negative binomial distribution (and surroundings)

A collection of 50 tracks is stored in my mp–3 device. In particular, there are 30 songs by U2, 15 songs by Tanita Tikaram and 5 sonates by Mozart.

The device offers two options for selecting tracks randomly: shuffle with replacement (you can listen to the same track twice) and shuffle without replacement (you cannot listen to the same track twice). Compute the probabilities that I need to wait for 30 tracks in order to listen to Mozart 5 times

(a) selecting the shuffle with replacement option,

(b) selecting the shuffle without replacement option.

5: The exponential distribution

Solve:

Suppose that commercial airplane crashes in a certain country occur at the rate of 2.5 per year.

(a) Is it reasonable to assume that such crashes are Poisson events? Explain.

(b) What is the probability that four or more crashes will occur next year?

(c) What is the probability that the next two crashes will occur within three months of one another?

(b) R commands:

The `dexp` function returns the exponential density distribution.

Plot the exponential density function for $\lambda = (1, 5, 10, 30, 50, 100)$ and shortly compare the obtained densities.

6: The normal distribution (and surroundings)

(a) R commands:

The `dnorm` function returns the normal density distribution.

Plot the normal density function for the following values of the parameters on the same plot: $(\mu = 10, \sigma^2 = 2)$, $(\mu = 12, \sigma = 2)$ and $(\mu = 10, \sigma = 1.5)$ and shortly compare the obtained densities.

(b) R commands:

The `pnorm` function returns the cumulative distribution function of the normal r.v.

Let $X \sim N(\mu = 10, \sigma^2 = 2)$ and $Z \sim N(\mu = 0, \sigma^2 = 1)$.

Exploiting the `pnorm` function, computes:

- $P(X < 10)$	- $P(Z < 0)$
- $P(X > 10)$	- $P(Z > 0)$
- $P(8 < X < 12)$	- $P(-1 < Z < 1)$
- $P(6 < X < 14)$	- $P(-2 < Z < 2)$
- $P(4 < X < 16)$	- $P(-3 < Z < 3)$

Why the results for the X and the Z are equivalent on each row? Shortly explain.

(c) R commands:

The `pnorm` function returns the cumulative distribution function of the normal r.v.

Plot the normal cumulative distribution function for the following values of the parameters on the same plot: $(\mu = 10, \sigma^2 = 2)$, $(\mu = 12, \sigma = 2)$ and $(\mu = 10, \sigma = 1.5)$ and shortly compare the obtained functions.

7: The central limit theorem

Using the R tutorial that graphically shows the central limit theorem for a uniform distribution (available in the Diary of class section), plot the mean of:

- a t distribution with 3 degrees of freedom
  (NOTE: search on Wikipedia for the mean and standard deviation)
- a negative exponential distribution with $\lambda = 1.2$
  (NOTE: see Section 6.5 for the mean and standard deviation)

Do it for $n = 10$, $n = 20$, $n = 30$ and $n = 100$. What are the differences between the two distributions?