EE1004 Lab 3

This is a self-learning lab with on-line tools and teacher support. In this lab, you will learn probability and statistics concepts on-line through doing experiments related to two famous problems: namely the Matching Problem and the Secretary Problem.

1) The Matching Problem*

The *matching experiment* is a random experiment that can the formulated in the following Matching Problem:

- Suppose that n male-female couples are at a party and that the males and females are randomly paired for a dance. A match occurs if a couple happens to be paired together.

These experiments are clearly equivalent from a mathematical point of view, and correspond to selecting a random permutation $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ of the population $\boldsymbol{D}_n = \{1, 2, \ldots, n\}$. Let us number the couples from 1 to n and $\boldsymbol{X}_i$ ($\boldsymbol{X}_i = 0, 1, 2, 3, \ldots$ or n) be the number of the woman paired with the i th man. We will say that a *match* occurs at position j if $\boldsymbol{X}_j = j$. Thus, the number of matches is the random variable N defined mathematically by

$$N_n = \sum_{j=1}^{n} I_j$$

where $I_j = \mathbf{1}(X_j = j)$ is the indicator variable for the event of match at position j. Our problem is to compute the mean, the standard deviation and the probability distribution of the number of matches (P_k) where $k = 0, 1, \ldots n$.

Example:

There are 5 couples ($n = 5$). Suppose we have the following matching results:

	Position 1	Position 2	Position 3	Position 4	Position 5
Male	Male 1	Male 2	Male 3	Male 4	Male 5
Female	Female 5	Female 2	Female 1	Female 4	Female 1

In this example, there are two matches (highlighted with red color), i.e. Couple 2 ($X_2 = j = 2$) and Couple 4 ($X_4 = j = 4$). Then, we have $I_2 = I_4 = 1$ and $I_1 = I_3 = I_5 = 0$. The number of matches is given by

$N_5 = I_1 + I_2 + I_3 + I_4 + I_5 = 2.$

With this experiment, the distribution of the number of matches is given by

$P_2 = 1$ and $P_0 = P_1 = P_3 = P_4 = P_5 = 0.$

Using the free on-line application (http://www.randomservices.org/random/apps/MatchExperiment.html), doing the matching experiment. You can go to the following to learn how to use the on-line applications (http://www.randomservices.org/random/apps/Instructions.html).

1. In the matching experiment, note how the **empirical/sample** mean, standard deviation and probability density function stabilize with the increase of the simulation times under different *n* values. For selected values of *n* (5, 10 and 20), run the simulation 1000 times and observe how the **empirical/sample** mean, standard deviation and probability density function stabilize (converge) to the **true/population** mean, standard deviation and probability density function. What theorem in probability theory can explain this phenomenon and what is about that theorem?

2. In the matching experiment, vary the parameter *n* (5, 10 and 20) and note the shape and location of the mean ± standard deviation bar as well as the probability density function bar. For selected values of the parameter, run the simulation 1000 times, observe and comment the empirical results compared to the true results as the simulation times increase.

3. In the matching experiment, vary the simulation times (10, 100 and 1000) and note the shape and location of the mean ± standard deviation bar and as well as the probability density function bar. For selected simulation times, run the simulation with $n = 5$ and write down in table and comment the sample mean and sample standard deviation when compared to the population mean and population standard deviation, respectively.

4. Eight married couples are randomly paired for a dance. Run the simulation 1000 times and write down each of the following:

   a. The sample and population probability density functions of the number of matches.
   b. The sample and the population means and standard deviations of the number of matches.
   c. The probability of at least 4 matches.

2) The Secretary Problem*

We have *n* *candidates* (applicants) for a secretary job. The assumptions are

a. The candidates are totally ordered from best to worst with no ties.
b. The candidates arrive sequentially in random order.
c. We can only determine the *relative ranks* of the candidates as they arrive. We cannot observe the *absolute ranks*.
d. Our goal is choosing the very best candidate.
e. Once a candidate is rejected, she is gone forever and cannot be recalled.
f. The number of candidates *n* is known.

The interesting questions you may ask are: What is an optimal strategy? What is the probability of success with this strategy? What happens to the strategy and the probability of success as *n* increases? In particular, when *n* is large, is there any reasonable hope of finding the best candidate?

Using the free on-line applications, doing the secretary experiment.

1. Play the secretary game (http://www.randomservices.org/random/apps/SecretaryGame.html) a number of times with selected values of *n* (2, 5 and 10). See if you can find a good strategy just by trial and error and comment what a reasonable type of strategy is.

2. In the secretary experiment (http://www.randomservices.org/random/apps/SecretaryExperiment.html), set the number of candidates to $n = 3$. Run the experiment 1000 times with each strategy $k \in \{1,2,3\}$ and write down the results. What is the optimal strategy (the optimal k value)?

3. In the secretary experiment (http://www.randomservices.org/random/apps/SecretaryExperiment.html), set the number of candidates to $n = 4$. Run the experiment 1000 times with each strategy $k \in \{1,2,3,4\}$ and write down the results. What is the optimal strategy (the optimal k value)?

4. As n increases, please comment the trends of the optimal k value and the optimal probability (with the optimal k value).

5. What are the approximate optimal strategy and the approximate probability of finding the best candidate?