**Lab3 Report:**

**86/100**

1. **The Matching Problem**

Q1)

For n = 5:

True mean: 1.000

True standard deviation: 1.000

True probability density function:

| n (Number of couples) | Pk(k = n) |
| --- | --- |
| 0 | 0.367 |
| 1 | 0.375 |
| 2 | 0.167 |
| 3 | 0.083 |
| 4 | 0.000 |
| 5 | 0.008 |

Sample mean: 1.030

Standard deviation: 0.998

Probability density function:

| n (Number of couples) | Pk(k = n) |
| --- | --- |
| 0 | 0.348 |
| 1 | 0.385 |
| 2 | 0.172 |
| 3 | 0.087 |
| 4 | 0.000 |
| 5 | 0.008 |

For n = 10:

True mean: 1.000

True standard deviation: 1.000

True probability density function:

| n (Number of couples) | Pk(k = n) |
| --- | --- |
| 0 | 0.368 |
| 1 | 0.368 |
| 2 | 0.184 |
| 3 | 0.061 |
| 4 | 0.015 |
| 5 | 0.003 |
| 6 | 0.001 |
| 7 | 0.000 |
| 8 | 0.000 |
| 9 | 0.000 |
| 10 | 0.000 |

Sample mean: 0.946

Standard deviation: 0.943

Probability density function:

| n (Number of couples) | Pk(k = n) |
| --- | --- |
| 0 | 0.376 |
| 1 | 0.381 |
| 2 | 0.180 |
| 3 | 0.049 |
| 4 | 0.012 |
| 5 | 0.002 |
| 6 | 0.000 |
| 7 | 0.000 |
| 8 | 0.000 |
| 9 | 0.000 |
| 10 | 0.000 |

For n = 20:

True mean: 1.000

True standard deviation: 1.000

True probability density function:

| n (Number of couples) | Pk(k = n) |
| --- | --- |
| 0 | 0.368 |
| 1 | 0.368 |
| 2 | 0.184 |
| 3 | 0.061 |
| 4 | 0.015 |
| 5 | 0.003 |
| 6 | 0.001 |
| 7 | 0.000 |
| 8 | 0.000 |
| 9 | 0.000 |
| 10 | 0.000 |
| 11 | 0.000 |
| 12 | 0.000 |
| 13 | 0.000 |
| 14 | 0.000 |
| 15 | 0.000 |
| 16 | 0.000 |
| 17 | 0.000 |
| 18 | 0.000 |
| 19 | 0.000 |
| 20 | 0.000 |

Sample mean: 1.009

Standard deviation: 0.999

Probability density function:

| n (Number of couples) | Pk(k = n) |
| --- | --- |
| 0 | 0.359 |
| 1 | 0.378 |
| 2 | 0.184 |
| 3 | 0.057 |
| 4 | 0.018 |
| 5 | 0.004 |
| 6 | 0.000 |
| 7 | 0.000 |
| 8 | 0.000 |
| 9 | 0.001 |
| 10 | 0.000 |
| 11 | 0.000 |
| 12 | 0.000 |
| 13 | 0.000 |
| 14 | 0.000 |
| 15 | 0.000 |
| 16 | 0.000 |
| 17 | 0.000 |
| 18 | 0.000 |
| 19 | 0.000 |
| 20 | 0.000 |

When the selected values of n increasing under the same simulation times, the sample statistic and probability density function are more converge to the true population parameter and the probability density function. This phenomenon can be explained by the Law of Large Number.

By the Law of Large Number, when the sample size is larger or tends to infinity, then the sample statistic (such as mean and standard deviation) and the sample probability density function will converge and closer to the true population parameter (such as the population mean and standard deviation) and population probability density function.

Q2)

When the selected value of n is increasing, the shape and the location of the mean ± standard deviation bar and probability density function bar are narrower and more leftward. Furthermore, when the simulation times increase, then the empirical results will be closer to the true result, and the graph will more likely the same as the true result

| Selected value = 5 | Selected value = 10 | Selected value = 20 |
| --- | --- | --- |
|  |  |  |

Q3)

When the simulation time is increasing, the shape and location of the mean ± standard deviation bar and probability density function will be closer and more likely the same as the true mean ± standard deviation bar and probability density. For example, for the simulation time = 10, the empirical result such as sample mean and standard deviation are much different and far away from the population mean and standard deviation. However, when the simulation times increase to 100 or 1000, it will be closer and closer and much similar to the population parameters

|  | Population | Simulation time = 10 | Simulation time = 100 | Simulation time = 1000 |
| --- | --- | --- | --- | --- |
| Mean | 1.000 | 1.600 | 0.920 | 1.008 |
| Standard deviation | 1.000 | 1.075 | 1.012 | 1.009 |

| Simulation time = 10 | Simulation time = 100 | Simulation time = 1000 |
| --- | --- | --- |
|  |  |  |

Q4)

(a) Probability density functions

| n (Number of couples) | Pk(k = n) [Popolutaion] | Pk(k = n) [Sample] |
| --- | --- | --- |
| 0 | 0.368 | 0.334 |
| 1 | 0.368 | 0.389 |
| 2 | 0.184 | 0.188 |
| 3 | 0.061 | 0.068 |
| 4 | 0.016 | 0.018 |
| 5 | 0.003 | 0.003 |
| 6 | 0.001 | 0.000 |
| 7 | 0.000 | 0.000 |
| 8 | 0.000 | 0.000 |

(b) Sample and population mean and standard deviations

|  | Population | Sample |
| --- | --- | --- |
| Mean | 1.000 | 1.056 |
| Standard deviation | 1.000 | 1.001 |

(c) Probability of at least 3 matches:

P(N) =

= 0.089

1. **The Secretary Problem**

Q1)

For the good strategy of selected value of n = 2, we give up the first candidate and take as a sample. If the second candidates are better than all previous candidates, we choose that candidate to be the best.

For the selected value of n = 5, we give up the first two candidates and take them as a sample. If any candidate of the following three candidates is better than all previous candidates, we choose that candidate to be the best.

For the selected value of n = 10, we give up the first four candidates and take them as a sample. If any candidate of the following 6 candidates is better than all previous candidates, we choose that candidate to be the best.

To conclude, let say if we have n candidates, we normally choose a value k that is the median of the total candidates if n is an odd number or close to the median value if the n is an even number, to get a value k. After finding out the value of k, we give up all candidates before k and take them as a sample; if any n > k candidates are better than before, we will choose that candidates.

Q2)

| k (Optimal strategy) | 1 | 2 | 3 | 4 |
| --- | --- | --- | --- | --- |
| Optimal probability (Population) | 0.250 | 0.458 | 0.417 | 0.250 |
| Optimal probability (Sample) | 0.258 | 0.463 | 0.421 | 0.237 |

From the above table, the optimal probability is the highest when k = 2. Therefore, we can conclude that the optimal strategy for 4 candidates is k = 2

Q3)

| k (Optimal strategy) | 1 | 2 | 3 | 4 | 5 | 6 |
| --- | --- | --- | --- | --- | --- | --- |
| Optimal probability (Population) | 0.100 | 0.283 | 0.366 | 0.399 | 0.398 | 0.373 |
| Optimal probability (Sample) | 0.114 | 0.263 | 0.384 | 0.404 | 0.386 | 0.409 |

From the above table, the optimal probability is the highest when k = 4. Therefore, we can conclude that the optimal strategy for 10 candidates is k = 4

Q4)

When the value of n is increase, the optimal k value will increase, and the optimal probability will decrease. However, the optimal probability will not be decrease to zero.

Therefore, the value of n and the optimal k value is in a linear relationship that is directly proportional. Then, the relationship between the optimal k value and the optimal probability is a decreasing exponential relationship, as well as the relationship between the value of n and the optimal probability

Q5)

The approximate optimal strategy is to determine the optimal value of k for different n candidates and reject the first k candidates and take them as a sample. If the following candidates are better than all previous candidates, we will take that candidate as the best candidate.

Furthermore, for the approximate probability of finding the best candidate by calculating , then we will get about 37%. Therefore, the approximate probability is about 37%. So if we need to find out the optimal k value, we can calculate to find out the approximate optimal strategy