4 Simulation

4.2 Algorithm and Candidate Estimates

To simulate the reality of how mitochondria were sampled, two-staged sampling was assumed in this project because of the fact that finite mitochondria are in a cell. For the first stage, we randomly chose N elements as subpopulation from $Exp(\theta)$ which is the distribution of Area of mitochondria. These N elements are like the total mitochondria in the cell with value as their Area. Therefore, the mean Area of the subpopulation, θ_A, are our interested parameter rather than θ. In the second stage, n elements were sampled from the subpopulation with sampling probability proportional to the Area of elements. Though the observed data we have are from sampling without replacement, sampling with replacement was also done in this project in order to obtain more insights of the estimates performance.

The process of generating data of the Perimeter of mitochondria is similar but based not only on the distribution of Area but also the distribution of Circularity of mitochondria and their independent relationship. In the first stage, N elements are also randomly chosen from $Beta(15, 5)$ with value as Circularity of mitochondria and then are plugged in the formula with the N elements of Area, $Circularity = \frac{4\pi Area}{Perimeter^2}$, to obtain N elements of Perimeter. The mean of the N elements of Perimeter are our interested parameter, θ_P. After getting the subpopulation of Perimeter, n elements of Perimeter are chosen with sampling probability proportional to its corresponding Area of elements. Also, sampling with and sampling without replacement both are considered in our simulation.

The algorithm for simulating the sampling distribution of Area:

1. Set $N = 2000$; *Ratio* between N and n are $(5\%, 10\%, 30\%, 50\%, 70\%, 95\%)$; *Repeated Times* $= 1000$ and $\theta = 1000$.

2. Generate N samples from $Exp(\theta)$ as subpopulation of Area and calculate subpopulation mean, θ_A, as the known parameter.

3. Sample a set of samples with size n from sobpopulation with sampling probability proportional to the value of Area with and without replacement. n is the product of N and a certain *Ratio*.

4. For each set of samples, calculate the candidate estiamtes.

5. Repeat 3. 4. for the set *RepeatedTimes* for each *Ratio*.

6. Calculate the Mean, Variance and MSE of the candidate estimates. Also draw plots of sampling distributions of each candidate estimate.

The algorithm for simulating the sampling distribution of Perimeter:

1. Set N to be 2000; *Ratio* between N and n are $(5\%, 10\%, 30\%, 50\%, 70\%, 95\%)$; *RepeatedTimes* are 1000 and θ is 1000.

2. Generate N samples from $Exp(\theta)$ distribution as subpopulation of Area and N samples from $Beta(\alpha, \beta)$ as subpopulation of Circularity. Assume the observed Circularity data we have are representative enough for the population of Circularity, and by observing the data α and β are set to be 15 and 5.

3. Plug the generated N elements of Area and N elements of Circularity into the formula, $Circularity = \frac{4\pi Area}{Perimeter^2}$, and obtain N elements of Perimeter. Calculate the mean of N elements of Perimeter, θ_P, and treat it as the true mean of Perimeter.

4. Sample a set of samples with size n from subpopulation of Perimeter with sampling probability proportional to Area with and without replacement. n is the product of N and a certain Ratio.

5. For each set of samples, calculate the candidate estimates.

6. Repeat 3. 4. for the set *RepeatedTimes* for each *Ratio*.

7. Calculate the Mean, Variance and MSE of the candidate estimates. Also draw plots of sampling distributions of each candidate estimate.