5 Simulation

5.1 Algorithm

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

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 $\mu = 1000$.

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

3. Sample a set of samples with size n from subpopulation 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 estimators: Arithmetic Mean (AM), Weighted Mean (WM) and Maximum Likelihood Estimator (MLE).

18

5. Repeat 3. 4. for the set *Repeated Times* for each *Ratio*.

6. Calculate the Mean, Standard Deviation and Root MSE for each candidate estimator. Also draw plots of sampling distributions for each candidate estimator.

The process on generating data of Perimeter of mitochondria was 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, randomly chose N elements of Circularity of mitochondria from $Beta(\alpha, \beta)$ and then substitute these N elements Circularity and the N elements of Area simulated before in the formula, Perimeter $= \sqrt{4\pi}\sqrt{\frac{\text{Area}}{\text{Circularity}}}$, to obtain N elements of Perimeter as subpopulation. Then, the mean of the N elements of Perimeter were our interested parameter, μ_P. After getting the subpopulation of Perimeter, n elements of Perimeter were chosen with sampling probability proportional to its corresponding Area. Again, sampling with and sampling without replacement both were considered in our simulation.

**The Algorithm for simulating the sampling distribution of Perimeter:**

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

2. Generate N samples from $Exp(\mu)$ 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 α and β are set to be 15 and 5 by observing the data we have.

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

19

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 estimators: Arithmetic Mean (AM), Weighted Mean (WM), Delta Method Estimator (DME), 2nd Order Taylor's Approximation Estimator (2TAE).

6. Repeat 3. 4. for the set *Repeated Times* for each *Ratio*.

7. Calculate the Mean, Standard Deviation and Root MSE for each candidate estimator. Also draw plots of sampling distributions for each candidate estimator.

5.2 Results for the simulation study

After repeating sampling with and without replacement, sampling distribution of each candidate estimator was constructed and their Biasedness, Standard Deviation and Root MSE were calculated. Figure 5.2.1, Table 5.2.1, Figure 5.2.2, and Table 5.2.2 shows that when sampling with replacement, MLE and 2TAE are always the best estimators for μ_A and μ_P and when sample size becomes larger, their performance on standard deviation (Std.) would also become better. As for the performance of AM, like our expectation, severely overestimates the true mean no matter how sample size changes. Conversly, when it is sampling without replacement, the performance of MLE and 2TAE estimators are not worse only when sample size compared to population size is small since when sample size is small, sampling without replacement is similar to sampling with replacement. For AM, even though its expected value approximates to the true mean as sample size becomes larger, it is still a biased estimator if the sample size is not equal to the population size.

To sum up, in this simulation, when it is sampling with replacement, MLE and the 2TAE perform best and AM has the worst performance. Nevertheless, MLE and 2TAE can only be obtained under strong assumptions on distributions of Area and Circularity. Hence, we also include WM in our following distribution for its sufficient performance and its nonparametric assumption.
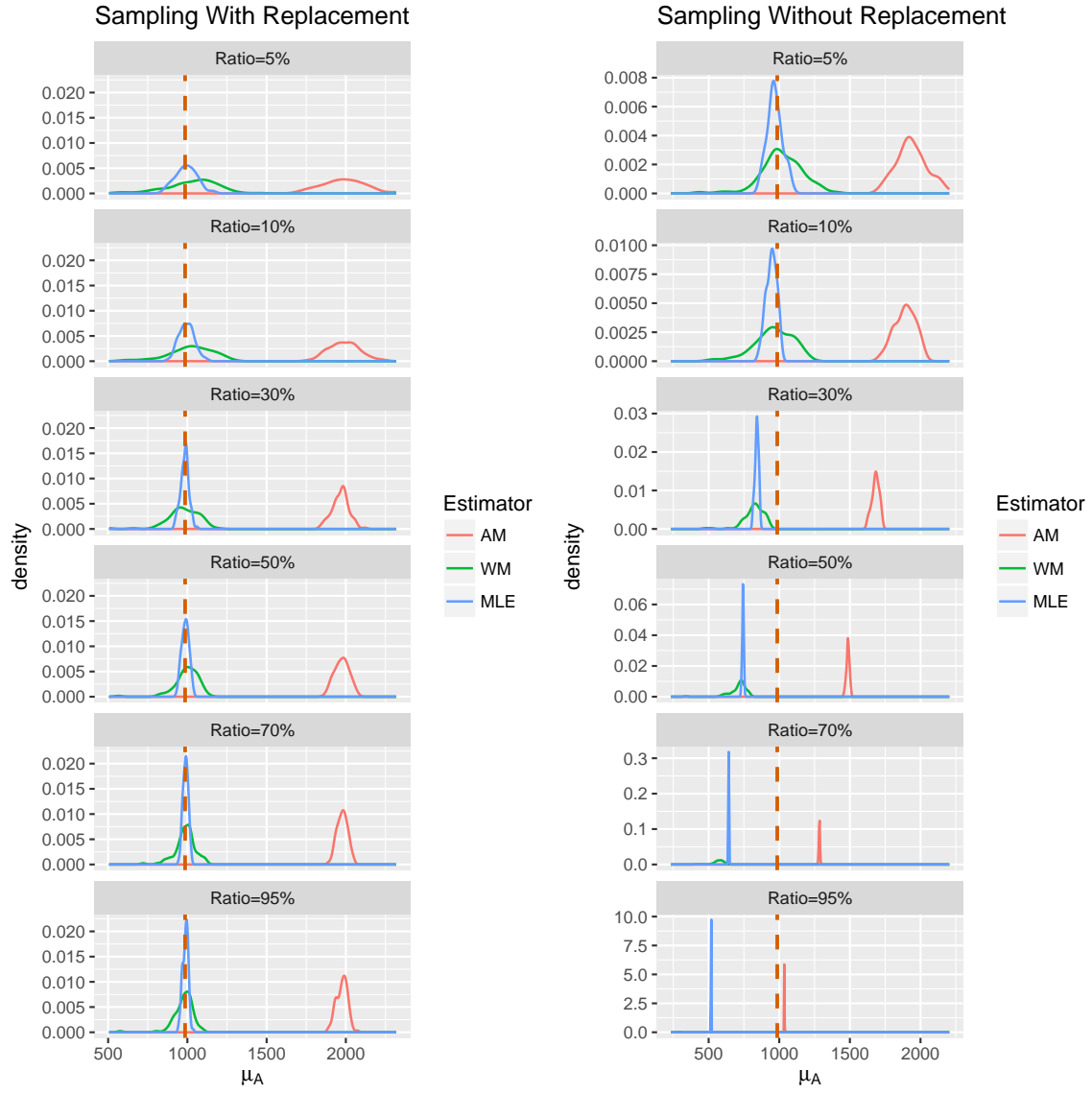
Figure 5.2.1: Sampling Distributions for Area with Different Ratios, N=2000, μ_A=985.75

Sampling **WITH** Replacement					Sampling **WITHOUT** Replacement				
Ratio	Estimator	Bias	Std. Dev.	Root MSE	Ratio	Estimator	Bias	Std. Dev.	Root MSE
5%	AM	1000	133.9	1000	5%	AM	951.12	106.95	951.18
	WM	45.2	162	47		WM	33.71	150.39	35.87
	MLE	7.1	67	10.8		MLE	-17.31	53.47	18.79
10%	AM	1003.9	96.3	1003.9	10%	AM	903.12	75.07	903.17
	WM	35	140.4	37		WM	-25.54	137.37	28.10
	MLE	9.1	48.1	11.4		MLE	-41.31	37.53	41.76
30%	AM	980.2	52.6	980.2	30%	AM	696.02	26.21	696.04
	WM	-8.7	100.6	13.2		WM	-163.47	74.31	163.70
	MLE	-2.8	26.3	5.8		MLE	-144.87	13.10	144.91
50%	AM	988.4	46	988.5	50%	AM	500.30	10.95	500.31
	WM	11.2	78	14.3		WM	-274.76	60.98	274.87
	MLE	1.3	23	5		MLE	-242.73	5.47	242.74
70%	AM	991.7	33.2	991.8	70%	AM	298.43	3.52	298.43
	WM	0.9	65.1	8.1		WM	-416.20	41.13	416.25
	MLE	3	16.6	5.1		MLE	-343.66	1.76	343.66
95%	AM	989.2	34.6	989.2	95%	AM	49.93	0.17	49.93
	WM	-8.9	65	12		WM	-697.56	18.23	697.58
	MLE	1.7	17.3	4.5		MLE	-467.91	0.09	467.91

Table 5.2.1: Performance Table for Area
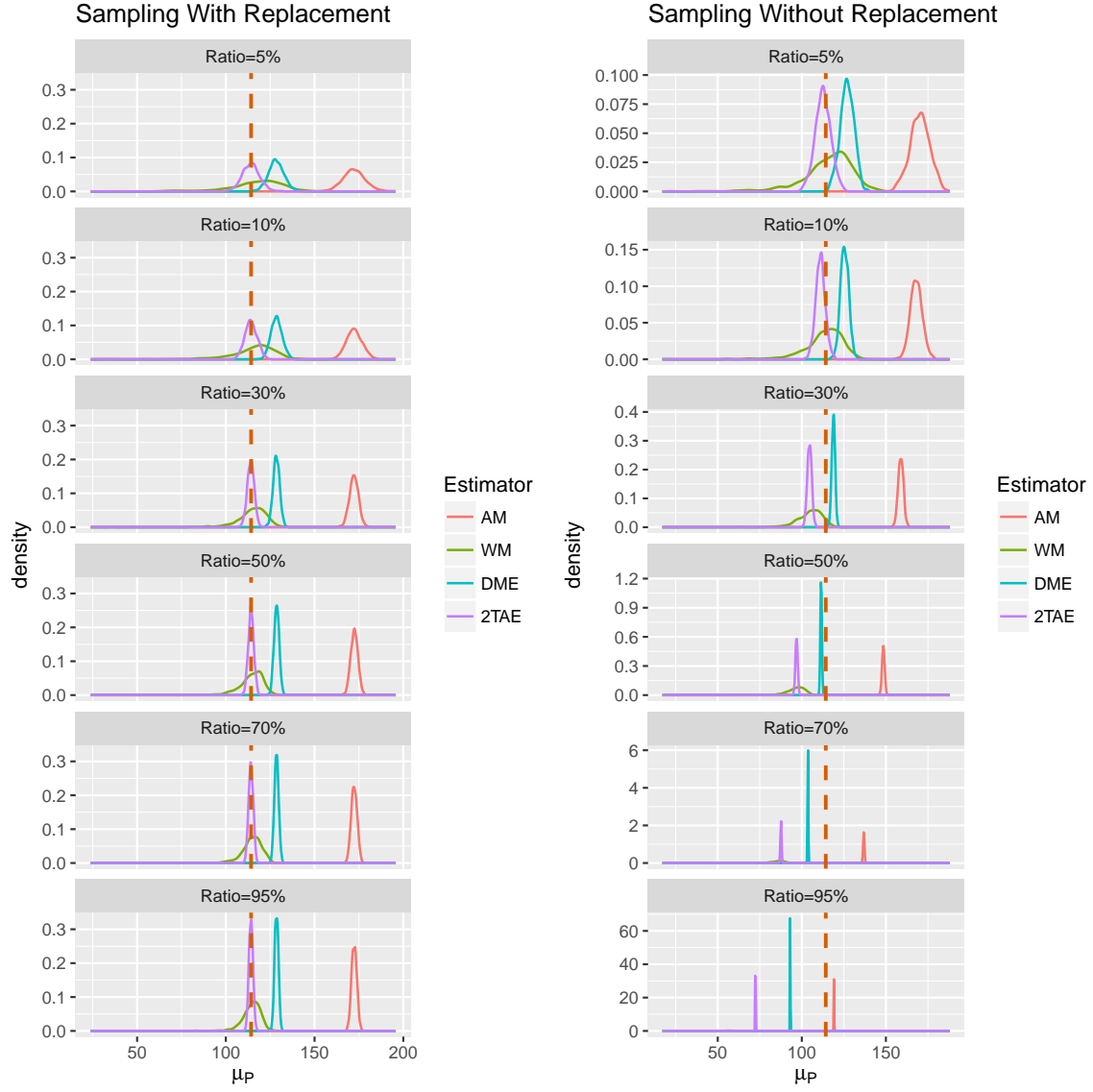
Figure 5.2.2: Sampling Distributions for Perimeter with Different Ratios, N=2000, μ_P=114.24

	Sampling WITH Replacement				Sampling WITHOUT Replacement			
Ratio	Estimator	Bias	Std. Dev.	Root MSE	Ratio	Estimator	Bias	Std. Dev.	Root MSE
5%	AM	58.05	6.30	58.11	5%	AM	55.95	5.80	56.01
	WM	2.93	15.81	4.94		WM	1.64	14.76	4.18
	DME	14.23	4.43	14.38		DME	12.70	4.07	12.86
	2TAE	-0.10	4.91	2.22		2TAE	-1.63	4.57	2.69
10%	AM	58.12	4.42	58.16	10%	AM	53.79	3.58	53.82
	WM	1.87	12.25	3.97		WM	-0.87	11.71	3.53
	DME	14.34	3.15	14.45		DME	11.17	2.47	11.28
	2TAE	-0.07	3.44	1.85		2TAE	-3.15	2.85	3.58
30%	AM	57.94	2.59	57.96	30%	AM	44.61	1.56	44.62
	WM	0.72	8.82	3.06		WM	-9.35	7.53	9.75
	DME	14.24	1.86	14.31		DME	4.55	0.98	4.65
	2TAE	-0.22	2.00	1.43		2TAE	-9.66	1.31	9.73
50%	AM	58.17	2.02	58.19	50%	AM	34.24	0.76	34.25
	WM	0.42	6.32	2.55		WM	-17.78	5.69	17.94
	DME	14.39	1.45	14.44		DME	-2.76	0.43	2.84
	2TAE	-0.04	1.56	1.25		2TAE	-17.38	0.66	17.39
70%	AM	58.09	1.65	58.10	70%	AM	22.60	0.35	22.61
	WM	0.42	5.90	2.46		WM	-28.71	4.66	28.79
	DME	14.33	1.18	14.37		DME	-10.61	0.17	10.62
	2TAE	-0.10	1.29	1.14		2TAE	-26.65	0.31	26.65
95%	AM	58.09	1.47	58.11	95%	AM	4.85	0.05	4.85
	WM	0.38	5.18	2.31		WM	-58.41	2.85	58.44
	DME	14.33	1.05	14.36		DME	-21.19	0.02	21.19
	2TAE	-0.09	1.13	1.07		2TAE	-41.83	0.03	41.83

Table 5.2.2: Performance Table for Perimeter