BST 222 Final Project

Simulation Study of HPSH's generosity in preparing dollar meals for its students

Group Members: Bowen Gu, Jojo Kennedy, Pluto Zhang

I. The real-world problem

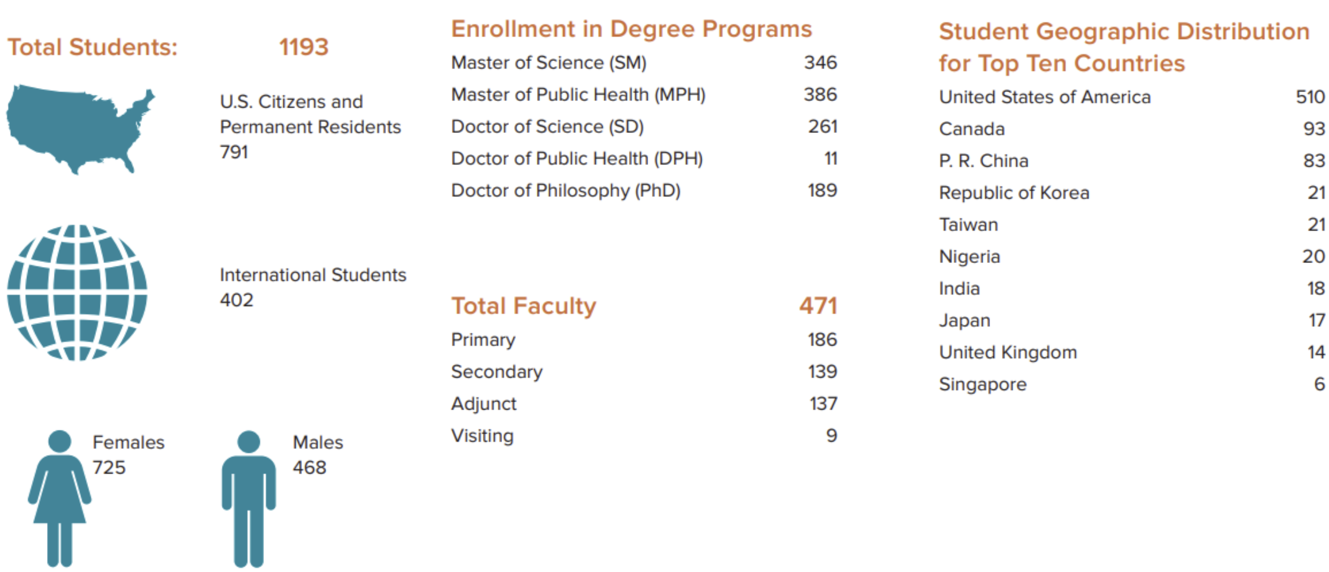
In Harvard T.H.Chan School of Public Health (HSPH), the cafeteria at Kresge provided dollar meals for both its students and faculty members. Although this is a boon for the students and faculty members in HSPH, how much will HSPH pay by providing the dollar meals? In this simulation study, we aim to estimate the profit that HSPH will lose due to providing dollar meals.

II. The methods

To estimate the profit that HSPH will lose due to providing dollar meals, we established the following model:

$$Y_i = (\beta - 1)X_i + a\varepsilon_i$$

where Y_i is the total profit that HSPH will lose due to providing dollar meals on day i, β is the actual cost of the dollar meal (we use $\beta - 1$ since HSPH will gain 1 dollar by selling one dollar meal), X_i is the number of people who purchased the dollar meal on day i, a is a constant that is used to adjust for the magnitude of food waste, which does not change through time, and ε_i is profit loss caused by food waste (e.g. some raw materials for the dollar meals are unconsumed and wasted) on day i. Here, we assumed that X_i follows a normal distribution $N(\mu,\sigma^2)$ with μ = 500 and σ = 50. The above number estimation is based on the following facts about HSPH, as the number of students and faculty members is around 1500 in total:



We assume that about $\frac{1}{3}$ of the students and faculty members (out of 1500) will purchase the dollar meal on average for each day, and the standard deviation is about $\frac{1}{10}$ of the mean.
For εi, we assume that it follows a standard half-normal distribution since we want the profit loss due to food waste to be non-negative and have a decreasing probability as the profit loss goes. The probability of losing a little food is high while the probability of losing much food is low. The constant *a* is set to be 100 to mimic a more realistic profit loss due to large amounts of food consumption and food waste.
We used the following three methods to estimate the parameter β:

- Estimator 1: $\sum_{i=1}^{n} (Xi\,Yi) \,/\, \sum_{i=1}^{n} (Xi^2) \;-\; \left(\frac{a\mu\sqrt{\frac{2}{\pi}}}{\sigma^2+\mu^2}\right)$
- Estimator 2: $\sum_{i=1}^{n} (Yi \;-\; a\sqrt{\frac{2}{\pi}}) \,/\, \sum_{i=1}^{n} Xi$
- Estimator 3: $\frac{1}{n}\sum_{i=1}^{n} ((Yi \;-\; a\sqrt{\frac{2}{\pi}}) \,/\, Xi)$

See the appendix section to see how we came up with the estimators above, why these estimators are unbiased, and what the variance of these estimators are.

III. The metrics

The metrics used in this study are the estimators' biases, variances, and MSEs. Plots are generated for each metric to compare the accuracy of each estimator.

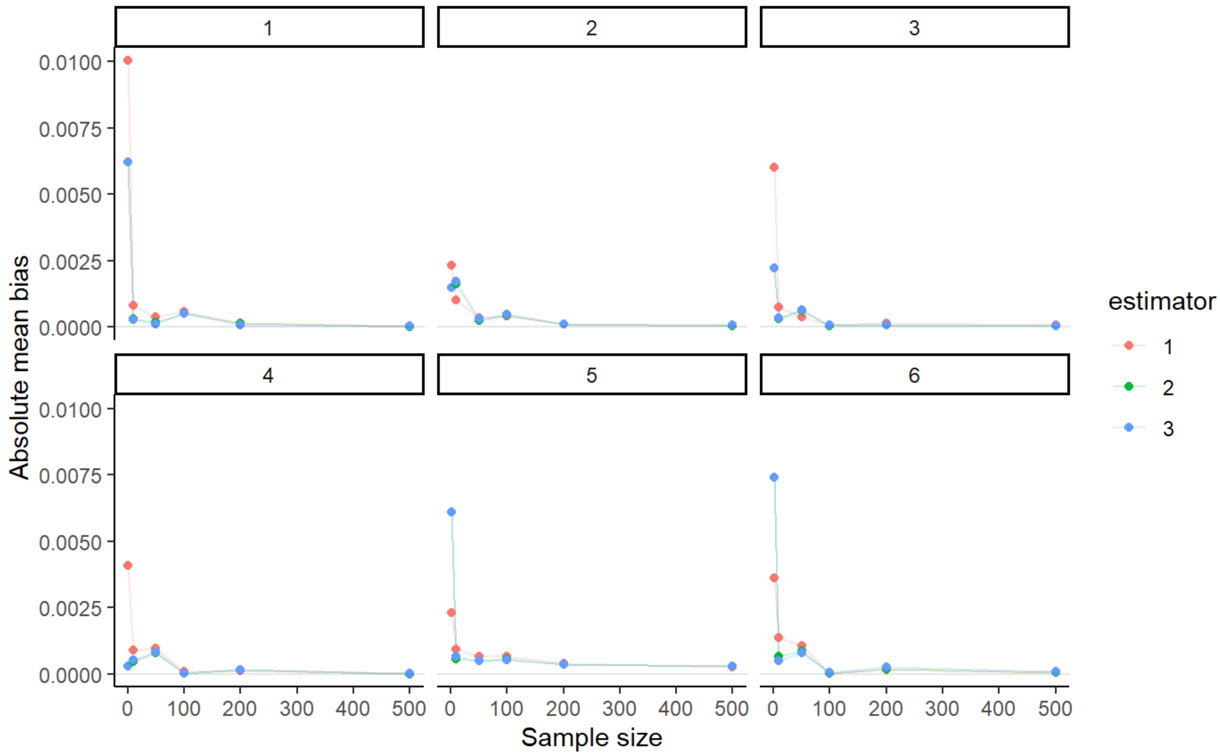
IV. The simulation design

1. The random data generating methods:
   The *x* variable values are generated from a normal distribution with sample sizes n = (1, 10, 50, 100, 200, 500), with the mean of 500 and the standard deviation of 50.
   The ε variable values are generated from a half-normal distribution with sample sizes n = (1, 10, 50, 100, 200, 500), with the mean of 0 and the standard deviation of 1.
   For each β in β = (1, 2, 3, 4, 5, 6), the *y* variable values are generated via the model ***Yi = (β - 1)Xi + aεi***
   Where *a* is a constant that equals 100.
   This procedure is simulated 1000 times, and the average of the biases, variances, and MSEs of the three estimators after the simulations are compared.

2. The simulation process:
   First, we used the distribution mentioned previously to generate data. After that, we saved the generated data in a data frame, then we calculated the three estimators based on the generated data together with other parameters mentioned above.
   For each simulation, we calculated their bias, variance, and MSE for the three estimators.
   We calculated the average of the biases, variances, and MSEs of the three estimators after the simulations.
   Finally, we plotted the results with respect to the sample size *n* and evaluated how accurate these estimators are, by comparing the estimator's mean biases, variances and MSEs at each n and β values.

V. The results and discussions

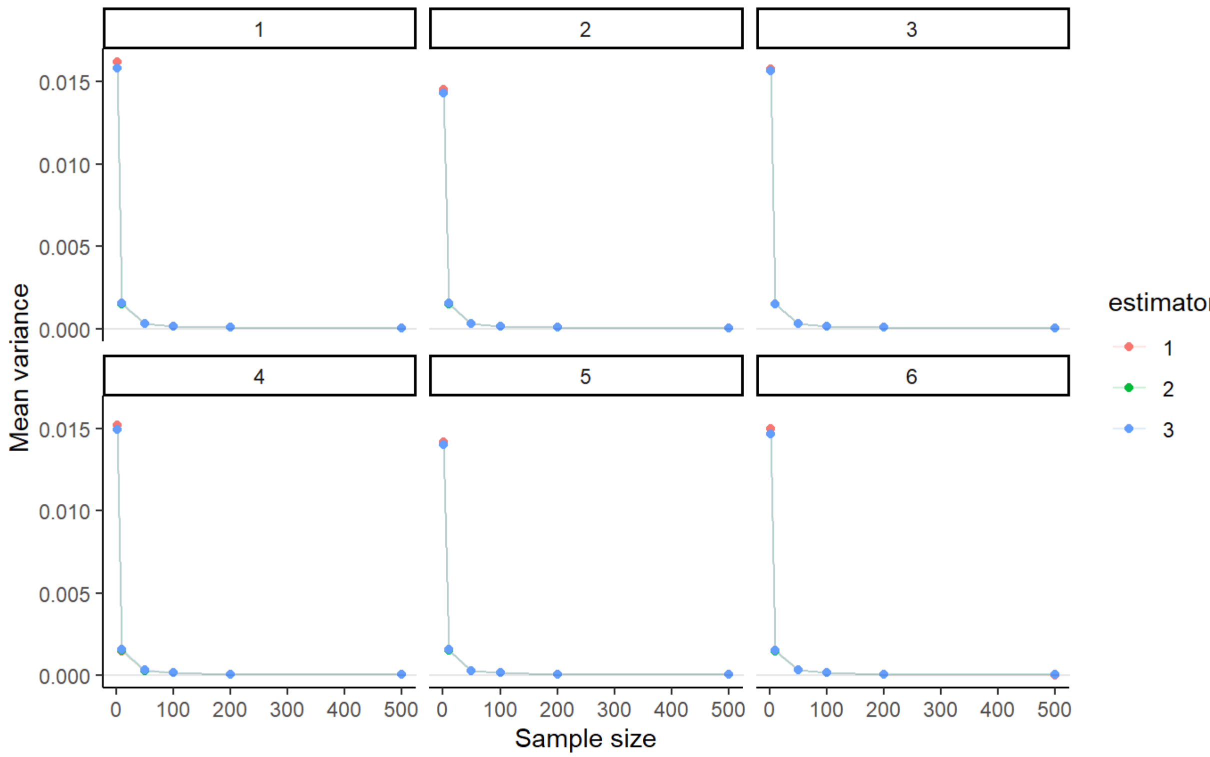
1. For the biases, we got the following plot (Plot 1), where the x-axis is the sample size, the y-axis is the absolute value of the mean bias, and the titles are the values of β. We choose to plot the absolute value of the mean bias instead of the mean bias to make a better comparison since bias can be both greater or smaller than 0.



Plot 1: the absolute value of the mean bias of the three estimation methods with regard to different values of β

From the plot we can see, as the sample size (n) increases, the absolute value of the mean bias of each of the three estimators gradually reduced to 0 for each value of β, meaning that all three estimators are unbiased estimators. Also, there is not much split when the sample size (n) is relatively large (> 100), meaning that there is no obvious superiority of any of the three estimators by comparing the absolute values of the mean bias.

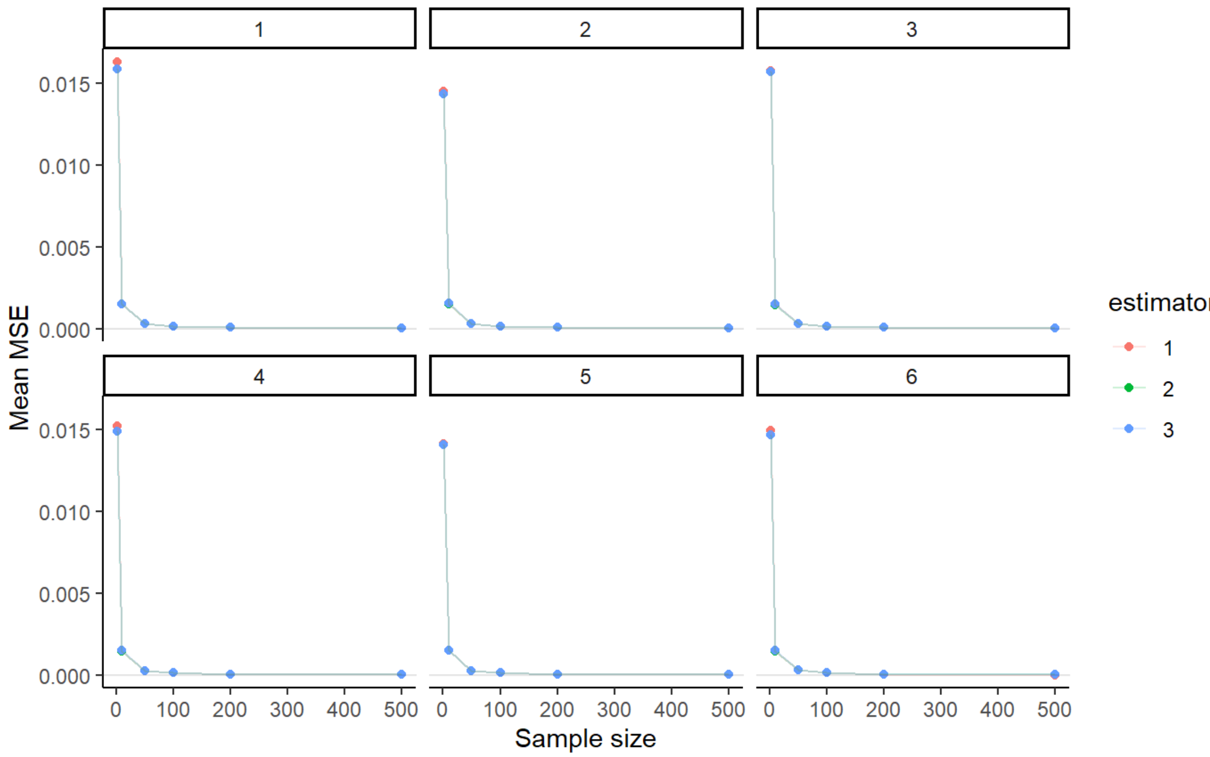
2. For the variances, we got the following plot (Plot 2), where the x-axis is the sample size, the y-axis is the mean variance, and the titles are the values of β.



Plot 2: the mean variance of the three estimation methods with regard to different values of β

From the plot we can see, as the sample size(n) increases, the mean variance drops and approaches 0 as the sample size increases, for every value of β, which is anticipated. The estimators don't yield much difference for every value of β and at every sample size n. Thus, by comparing the mean variance, there is no obvious superiority of any of the three estimators.

3. For the mean MSE, we got the following plot (Plot 3), where the x-axis is the sample size, the y-axis is the mean MSE, and the titles are the values of β.



Plot 3: the plots of mean MSE of three estimators

From the plot we can see that it is very similar to the mean variance case, as the sample size(n) increases, the mean MSE drops and approaches 0 as the sample size increases, for every value of β, which is anticipated. The estimators don't yield much difference for every value of β and at every sample size n. Thus, by comparing the mean MSE, there is no obvious superiority of any of the three estimators.

VI. Conclusions

Based on the discussions above, we can see that the three estimators have very similar performance. According to the results from the absolute values of mean bias, the mean variance, and the mean MSE, we cannot see significant superiority of any estimators over the other two.

VII. References

1. Larry Han, BST 222 Lab: Simulation study.rmd
2. About HSPH BY THE NUMBERS · 2013 - Harvard University, https://hpac.harvard.edu/files/hpac/files/hsph_fastfacts.pdf

BST 222 Final Project Appendix

Bowen Gu, Jojo Kennedy, Pluto Zhang

November 2022

1 The issue with our MLE estimator

$$f_x(\lambda) = \lambda e^{-\lambda x}, \tag{1}$$

Thus,

$$\frac{1}{a}\left(Y_i - \beta X_i\right), \tag{2}$$

follows:

$$\exp(\lambda). \tag{3}$$

The likelihood function goes:

$$L(\beta) = \prod_{i=1}^{n} \lambda e^{-\lambda\left[\frac{1}{a}(Y_i - \beta X_i)\right]}, \tag{4}$$

$$\ell(\beta) = \sum_{i=1}^{n} \log(\lambda) - \lambda\left[\frac{1}{a}\left(Y_i - \beta X_i\right)\right]. \tag{5}$$

$$\frac{\partial l(\beta)}{\partial \beta} = \sum_{i=1}^{n} \frac{\lambda}{a} X_i = 0. \tag{6}$$

Thus, beta is gone and there is no MLE.

2 Why we don't use MOM estimator

Since:

$$\frac{1}{a}\left(Y_i - \beta X_i\right), \tag{7}$$

follows:

$$\exp(\lambda). \tag{8}$$

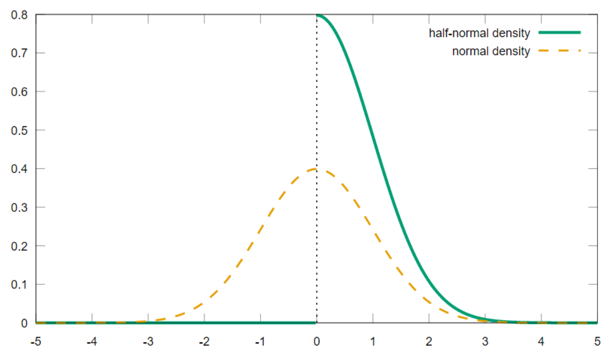
Thus,

$$\overline{\frac{1}{a}\left(Y_i - \beta X_i\right)} = \frac{1}{\lambda}. \tag{9}$$

Just like $\bar{x} = \mu$ in normal distribution We can now have $\hat{\lambda}_{MOM} = \frac{1}{\overline{\frac{1}{a}(Y_i - \beta X_i)}}$ but there is no way we can get $\hat{\beta}_{MOM}$

3 After we change the distribution of residuals to half-normal distribution

The graph representing the half-normal distribution:



$$f_x(\sigma) = \frac{\sqrt{2}}{\sigma\sqrt{\pi}} e^{-\frac{x^2}{2\sigma^2}} (x > 0). \tag{10}$$

Mean is:

$$= \frac{\sigma\sqrt{2}}{\sqrt{\pi}}. \tag{11}$$

Median:

$$= \sigma^2 \left(1 - \frac{2}{\pi}\right). \tag{12}$$

and for our model,

$$Y_i = \beta X_i + a\epsilon_i. \tag{13}$$

first, let's calculate the MLE of beta:

$$L(\beta) = \prod_{i=1}^{n} \frac{\sqrt{2}}{\sigma\sqrt{\pi}} e^{-\frac{\frac{1}{a^2}(Y_i - \beta X_i)^2}{2\sigma^2}}. \tag{14}$$

$$\begin{aligned} l(\beta) &= \sum_{i=1}^{n} \left(\log\left(\frac{\sqrt{2}}{\sigma^2\sqrt{\pi}}\right) - \frac{(Y_i - \beta X_i)^2}{2a^2\sigma^2} \right), \\ &= n\log\left(\frac{\sqrt{2}}{\sigma\sqrt{\pi}}\right) - \sum_{i=1}^{n} \frac{Y_i^2 - 2\beta y_i X_i + \beta^2 X_i^2}{2a^2\sigma^2}. \end{aligned} \tag{15}$$

$$\frac{\partial l(\beta)}{\partial \beta} = \sum_{i=1}^{n} \frac{Y_i X_i}{a^2\sigma^2} - \sum_{i=1}^{n} \frac{\beta X_i^2}{a^2\sigma^2}. \tag{16}$$

$$\Rightarrow \hat{\beta}_{MLE} = \frac{\sum_{i=1}^{n} (X_i Y_i)}{\sum_{i=1}^{n} X_i^2}. \tag{17}$$

Then, we calculate the expected value of the MLE

$$\begin{aligned}
E\left(\hat{\beta}_{MLE}\right) &= \frac{\sum_{i=1}^{n} E\left[X_i\left(\beta X_i + a\varepsilon_i\right)\right]}{\sum_{i=1}^{n} E\left(X_i^2\right)}, \\
&= \frac{\sum_{i=1}^{n} \left(\beta E\left(X_i^2\right) + aE\left(X_i\right)E\left(\varepsilon_i\right)\right]}{\sum_{i=1}^{n} E\left(X_i^2\right)}, \\
&= \beta + \frac{\sum_{i=1}^{n} aE\left(X_i\right)E\left(\varepsilon_i\right)}{\sum_{i=1}^{n} E\left(X_i^2\right)}, \\
&= \frac{na\mu\frac{\sqrt{2}}{\sigma\sqrt{\pi}}}{\sum_{i=1}^{n}\left[Var\left(X_i\right) + E\left(X_i\right)^2\right]}, \\
E\left(\hat{\beta}_{MLE}\right) &= \beta + \frac{a\mu\frac{\sqrt{2}}{\sigma\sqrt{\pi}}}{\sigma_1^2 + \mu^2}.
\end{aligned} \tag{18}$$

Thus, bias of the MLE is:

$$\frac{a\mu\frac{\sqrt{2}}{\sigma\sqrt{\pi}}}{\sigma_1^2 + \mu^2}. \tag{19}$$

And in our case,

$$\sigma = 1. \tag{20}$$

stands for the variance of the half normal distribution

$$\sigma_1 = 50. \tag{21}$$

which stands for the variance of the number of students dining each day

$$\mu = 500. \tag{22}$$

stands for the mean number of students dining each day

Third, let's calculate the variance of MLE of beta:

$$\begin{aligned}
Var\left(\hat{\beta}_{MLE}\right) &= Var\left(\frac{\sum_{i=1}^{n} X_i Y_i}{\sum_{i=1}^{n} X_i^2}\right), \\
&= Var\left(\sum_{i=1}^{n} a_i Y_i\right).
\end{aligned} \tag{23}$$

where $a_i = \frac{X_i}{\sum_{i=1}^n X_i^2}$ is a constant. Thus,

$$\begin{aligned} Var\left(\hat{\beta}_{MLE}\right) &= \sum_{i=1}^n Var\left(a_i Y_i\right), \\ &= \sum_{i=1}^n a_i^2 Var\left(Y_i\right), \\ &= \sum_{i=1}^n a_i^2 Var\left(\beta X_i + a\varepsilon_i\right), \\ &= \sum_{i=1}^n a_i^2 Var\left(a\varepsilon_i\right). \end{aligned} \tag{24}$$

where βX_i is also a constant.

Thus,

$$\begin{aligned} Var\left(\hat{\beta}_{MLE}\right) &= \sum_{i=1}^n a_i^2 a^2 Var\left(\varepsilon_i\right), \\ &= \sum_{i=1}^n a_i^2 a^2 \sigma^2 \left(1 - \frac{2}{\pi}\right), \\ &= \sum_{i=1}^n \left(\frac{X_i}{\sum_{i=1}^n X_i^2}\right)^2 a^2 \sigma^2 \left(1 - \frac{2}{\pi}\right), \\ &= a^2\sigma^2 \left(1 - \frac{2}{\pi}\right) \sum_{i=1}^n \frac{X_i^2}{\left(\sum_{i=1}^n X_i^2\right)^2}, \\ &= a^2\sigma^2 \left(1 - \frac{2}{\pi}\right) \frac{\sum_{i=1}^n X_i^2}{\left(\sum_{i=1}^n X_i^2\right)^2}, \\ &= \frac{a^2\sigma^2\left(1 - \frac{2}{\pi}\right)}{\sum_{i=1}^n X_i^2}, \\ Var\left(\hat{\beta}_{MLE}\right) &= \frac{Var\left(Y_i\right)}{\sum_{i=1}^n X_i^2}. \end{aligned} \tag{25}$$

4 First Estimator: $\hat{\beta}_1$

Thus, after getting an idea of the variance and bias of the $\hat{\beta}_{MLE}$, we adjust it to be:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n X_i Y_i}{\sum_{i=1}^n X_i^2} - \frac{a\mu\sqrt{\frac{2}{\pi}}}{\sigma_1^2 + \mu^2} \tag{26}$$

We want to check that $\hat{\beta}_1$ is unbiased. To find the bias of $\hat{\beta}_1$, we calculate its expected value:

$$
\begin{aligned}
E\left(\hat{\beta}_1\right) &= E\left(\frac{\sum_{i=1}^n X_i y_i}{\sum_{i=1}^n X_1^2}\right) - \frac{\alpha\mu\sqrt{\frac{2}{\pi}}}{\sigma_1^2+\mu^2}, \\
&= \beta + \frac{\frac{\alpha\mu}{\sigma}\sqrt{\frac{2}{\pi}}}{\sigma_1^2+\mu^2} - \frac{\alpha\mu\sqrt{\frac{2}{\pi}}}{\sigma_1^2+\mu^2}.
\end{aligned}
\tag{27}
$$

$\sigma = 1$

Thus,

$$E\left(\hat{\beta}_1\right) = \beta \tag{28}$$

This estimator is unbiased now.

Next, we want to know the estimator's variance:

$$
\begin{aligned}
Var\left(\hat{\beta}_1\right) &= Var\left(\frac{\sum_{i=1}^n X_i Y_i}{\sum_{i=1}^n X_i^2}\right) - 0, \\
&= Var\left(\frac{\sum_{i=1}^n X_i Y_i}{\sum_{i=1}^n X_i^2}\right), \\
Var\left(\hat{\beta}_1\right) &= \frac{Var\left(Y_i\right)}{\sum_{i=1}^n X_i^2}.
\end{aligned}
\tag{29}
$$

Thus, in our code, we use $Var\left(\hat{\beta}_1\right) = \frac{Var(Y_i)}{\sum_{i=1}^n X_i^2}$.

Note that $\hat{\beta}_1$ has the form of MLE, but is no longer MLE, since we have adjusted for it already. We call this estimator $\hat{\beta}_1$.

5 Second Estimator: $\hat{\beta}_2$

$\hat{\beta}_2$ originates from $\hat{\beta}_2'$:

$$\hat{\beta}_2' = \frac{\sum_{i=1}^n Y_i}{\sum_{i=1}^n X_i} \tag{30}$$

We need an estimator that is unbiased, so we want to check the bias of $\hat{\beta}_2'$.

To find the bias of $\hat{\beta}_2'$, we calculate its expected value.

$$\begin{aligned} E\left(\hat{\beta}_2'\right) &= E\left(\frac{\sum_{i=1}^n Y_i}{\sum_{i=1}^n X_i}\right) \\ &= \frac{\sum_{i=1}^n E\left(Y_i\right)}{\sum_{i=1}^n E\left(X_i\right)} \\ &= \frac{\sum_{i=1}^n E\left(\beta X_i + a\varepsilon_i\right)}{\sum_{i=1}^n E\left(X_i\right)} \\ &= \beta + \frac{a\sum_{i=1}^n E\left(\varepsilon_i\right)}{\sum_{i=1}^n E\left(X_i\right)} \\ &= \beta + \frac{a\sum_{i=1}^n \frac{\sigma\sqrt{2}}{\sqrt{\pi}}}{\sum_{i=1}^n E\left(X_i\right)} \\ &= \beta + \frac{an\frac{\sigma\sqrt{2}}{\sqrt{\pi}}}{n\mu} \\ E\left(\hat{\beta}_2'\right) &= \beta + \frac{a\sigma\sqrt{\frac{2}{\pi}}}{\mu} \end{aligned} \tag{31}$$

Therefore, the bias of $\hat{\beta}_2'$ is $\frac{a\sigma\sqrt{\frac{2}{\pi}}}{\mu}$.

Next, we want to know the variance of $\hat{\beta}_2'$.

$$\operatorname{Var}\left(\hat{\beta}_2'\right) = \operatorname{Var}\left(\frac{\sum_{i=1}^n Y_i}{\sum_{i=1}^n X_i}\right) \tag{32}$$

$\sum_{i=1}^n X_i$ is a constant.

Thus,

$$\operatorname{Var}\left(\hat{\beta}_2'\right) = \left(\frac{1}{\sum_{i=1}^n X_i}\right)^2 \sum_{i=1}^n \operatorname{Var}\left(Y_i\right) \tag{33}$$

$Y_i = \beta X_i + a\varepsilon_i$

Thus,

$$\operatorname{Var}\left(\hat{\beta}_2'\right) = \left(\frac{1}{\sum_{i=1}^n X_i}\right)^2 \sum_{i=1}^n \operatorname{Var}\left(\beta X_i + a\varepsilon_i\right) \tag{34}$$

βX_i is a constant. Constants have no variance.

Thus,

$$\begin{aligned}
\operatorname{Var}\left(\hat{\beta}_2'\right) &= \left(\frac{1}{\sum_{i=1}^n X_i}\right)^2 \sum_{i=1}^n \operatorname{Var}\left(0 + a\varepsilon_i\right) \\
&= \left(\frac{1}{\sum_{i=1}^n X_i}\right)^2 a^2 \sum_{i=1}^n \operatorname{Var}\left(\varepsilon_i\right) \\
&= \left(\frac{1}{\sum_{i=1}^n X_i}\right)^2 a^2 n\sigma^2 \left(1 - \frac{2}{\pi}\right) \\
\operatorname{Var}\left(\hat{\beta}_2'\right) &= \frac{n \operatorname{Var}\left(Y_i\right)}{\left(\sum_{i=1}^n X_i\right)^2}
\end{aligned} \tag{35}$$

Thus, after getting an idea of the variance and bias of the $\hat{\beta}_2'$, we adjust $\hat{\beta}_2$ to be:

$$\hat{\beta}_2 = \frac{\sum_{i=1}^n \left(Y_i - a\sqrt{\frac{2}{\pi}}\right)}{\sum_{i=1}^n X_i} \tag{36}$$

We want to know the bias of $\hat{\beta}_2$. To find the bias of $\hat{\beta}_2$, we calculate its expected value.

$$\begin{aligned}
E\left(\hat{\beta}_2\right) &= E\left(\frac{\sum_{i=1}^n \left(Y_i - a\sqrt{\frac{2}{\pi}}\right)}{\sum_{i=1}^n X_i}\right) \\
&= \frac{\sum_{i=1}^n E\left(Y_i - a\sqrt{\frac{2}{\pi}}\right)}{\sum_{i=1}^n E\left(X_i\right)} \\
&= \frac{\sum_{i=1}^n \left[E\left(Y_i\right) - a\sqrt{\frac{2}{\pi}}\right]}{\sum_{i=1}^n E\left(X_i\right)}
\end{aligned} \tag{37}$$

$Y_i = \beta X_i + a\varepsilon_i$

Thus,

$$\begin{aligned}
E\left(\hat{\beta}_2\right) &= \frac{\sum_{i=1}^n \left[E\left(\beta X_i + a\varepsilon_i\right) - a\sqrt{\frac{2}{\pi}}\right]}{\sum_{i=1}^n E\left(X_i\right)} \\
&= \beta + \frac{\sum_{i=2}^n \left[aE\left(\varepsilon_i\right) - a\sqrt{\frac{2}{\pi}}\right]}{\sum_{i=1}^n E\left(X_i\right)} \\
&= \beta + \frac{na\frac{\sigma\sqrt{2}}{\sqrt{\pi}} - a\sqrt{\frac{2}{\pi}}}{n\mu}
\end{aligned} \tag{38}$$

$\sigma = 1$

Thus,

$$\begin{aligned}
E\left(\hat{\beta}_2\right) &= \beta + \frac{1}{\mu} \\
E\left(\hat{\beta}_3\right) &= \beta
\end{aligned} \tag{39}$$

This estimator is unbiased now.

Next, we want to know the variance of $\hat{\beta}_2$.

$$\begin{aligned}
\operatorname{Var}\left(\hat{\beta}_2\right) &= \operatorname{Var}\left(\frac{\sum_{i=1}^{n}\left(Y_i - a\sqrt{\frac{2}{\pi}}\right)}{\sum_{i=1}^{n} X_i}\right) \\
&= \left(\frac{1}{\sum_{i=1}^{n} X_i}\right)^2 \sum_{i=1}^{n} \operatorname{Var}\left(Y_i - a\sqrt{\frac{2}{\pi}}\right) \\
&= \left(\frac{1}{\sum_{i=1}^{n} X_i}\right)^2 \sum_{i=1}^{n} \operatorname{Var}\left(Y_i\right) \\
\operatorname{Var}\left(\hat{\beta}_2\right) &= \frac{n \operatorname{Var}\left(Y_i\right)}{\left(\sum_{i=1}^{n} X_i\right)^2}
\end{aligned} \tag{40}$$

Thus, in our code, we use $\operatorname{Var}\left(\hat{\beta}_2\right) = \frac{n \operatorname{Var}(Y_i)}{\left(\sum_{i=1}^{n} X_i\right)^2}$.

6 Third Estimator: $\hat{\beta}_3$

$\hat{\beta}_3$ originates from $\hat{\beta}_3'$:

$$\hat{\beta}_3' = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{Y_i}{X_i}\right) \tag{41}$$

We need an estimator that is unbiased, so we want to check the bias of $\hat{\beta}_3'$. To find the bias of $\hat{\beta}_3'$, we calculate its expected value.

$$\begin{aligned}
E\left(\hat{\beta}_3'\right) &= E\left(\frac{1}{n}\sum_{i=1}^{n}\left(\frac{Y_i}{X_i}\right)\right) \\
&= \frac{1}{n}\sum_{i=1}^{n}\frac{E\left(Y_i\right)}{E\left(X_i\right)}
\end{aligned} \tag{42}$$

$Y_i = \beta X_i + a\varepsilon_i$

Thus,

$$\begin{aligned}
E\left(\hat{\beta}_3'\right) &= \frac{1}{n}\sum_{i=1}^{n}\frac{E\left(\beta X_i + a\varepsilon_i\right)}{E\left(X_i\right)} \\
&= \frac{1}{n}\sum_{i=1}^{n}\left[\beta + \frac{aE\left(\varepsilon_i\right)}{E\left(X_i\right)}\right] \\
&= \frac{1}{n}\sum_{i=1}^{n}\left[\beta + \frac{a\sigma\sqrt{\frac{2}{\pi}}}{\mu}\right] \\
E\left(\hat{\beta}_3'\right) &= \beta + \frac{a\sigma\sqrt{\frac{2}{\pi}}}{\mu}
\end{aligned} \tag{43}$$

Therefore, the bias of the $\hat{\beta}_3'$ is $\frac{a\sigma\sqrt{\frac{2}{\pi}}}{\mu}$.

Next, we want to know the variance of $\hat{\beta}_3'$.

$$\begin{aligned}
\operatorname{Var}\left(\hat{\beta}_3'\right) &= \operatorname{Var}\left(\frac{1}{n}\sum_{i=1}^{n}\left(\frac{Y_i}{X_i}\right)\right) \\
&= \frac{1}{n^2}\operatorname{Var}\left(\sum_{i=1}^{n}\left(\frac{Y_i}{X_i}\right)\right) \\
&= \frac{1}{n^2}\operatorname{Var}\left(\frac{Y_1+Y_2+\cdots+Y_n}{X_1+X_2+\ldots+X_n}\right) \\
&= \frac{1}{n^2}\operatorname{Var}\left(\frac{Y_1}{X_1+X_2+\ldots+X_n}+\cdots+\frac{Y_n}{X_1+X_2+\ldots+X_r}\right) \\
&= \frac{1}{n^2}\frac{1}{(X_1+X_{21}++X_n)^2}\operatorname{Var}(Y_i) \\
\operatorname{Var}\left(\hat{\beta}_3'\right) &= \frac{\operatorname{Var}(Y_i)}{n^2}\frac{1}{\sum_{i=1}^{n}X_i^2}
\end{aligned} \tag{44}$$

Thus, after getting an idea of the variance and bias of the $\hat{\beta}_3'$, we adjust $\hat{\beta}_3$ to be:

$$\hat{\beta}_3 = \frac{1}{n}\sum_{i=1}^{n}\frac{Y_i - a\sqrt{\frac{2}{\pi}}}{X_i} \tag{45}$$

We want to check that $\hat{\beta}_3$ is unbiased. To find the bias of $\hat{\beta}_3$, we calculate its expected value.

$$\begin{aligned}
E\left(\hat{\beta}_3\right) &= E\left(\frac{1}{n}\sum_{i=1}^{n}\frac{Y_i - a\sqrt{\frac{2}{\pi}}}{X_i}\right) \\
&= \frac{1}{n}\sum_{i=1}^{n}E\left(\frac{Y_i - a\sqrt{\frac{2}{\pi}})}{X_i}\right) \\
&= \frac{1}{n}\sum_{i=1}^{n}\frac{E(Y_i) - a\sqrt{\frac{2}{\pi}}}{E(X_i)}
\end{aligned} \tag{46}$$

$Y_i = \beta X_i + a\varepsilon_i$

Thus,

$$
\begin{aligned}
E\left(\hat{\beta}_3\right) &= \frac{1}{n}\sum_{i=1}^{n}\frac{E\left(\beta X_i + a\varepsilon_i\right) - a\sqrt{\frac{2}{\pi}}}{E\left(X_i\right)} \\
&= \frac{1}{n}\sum_{i=1}^{n}\frac{\beta E\left(X_i\right) + aE\left(\varepsilon_i\right) - a\sqrt{\frac{2}{\pi}}}{E\left(X_i\right)} \\
&= \frac{1}{n}\sum_{i=1}^{n}\beta + \frac{a\sigma\sqrt{\frac{2}{\pi}} - a\sqrt{\frac{2}{\pi}}}{\mu}
\end{aligned} \tag{47}
$$

$\sigma = 1$

Thus,

$$
\begin{aligned}
E\left(\hat{\beta}_3\right) &= \frac{1}{n}n\beta \\
E\left(\hat{\beta}_3\right) &= \beta
\end{aligned} \tag{48}
$$

This estimator is unbiased now.

Next, we want to know the variance of $\hat{\beta}_3'$.

$$
\begin{aligned}
\operatorname{Var}\left(\hat{\beta}_3\right) &= \operatorname{Var}\left(\frac{1}{n}\sum_{i=1}^{n}\frac{Y_i - a\sqrt{\frac{2}{\pi}}}{X_i}\right) \\
&= \frac{1}{n^2}\frac{1}{\sum_{i=1}^{n}{X_i}^2}\operatorname{Var}\left(Y_i - a\sqrt{\frac{2}{\pi}}\right) \\
&= \frac{1}{n^2}\frac{1}{\sum_{i=1}^{n}{X_i}^2}\operatorname{Var}\left(Y_i\right) \\
\operatorname{Var}\left(\hat{\beta}_3\right) &= \frac{\operatorname{Var}\left(Y_i\right)}{n^2}\frac{1}{\sum_{i=1}^{n}{X_i}^2}
\end{aligned} \tag{49}
$$

Thus, in our code, we use $\operatorname{Var}\left(\hat{\beta}_3\right) = \frac{\operatorname{Var}(Y_i)}{n^2}\frac{1}{\sum_{i=1}^{n}{X_i}^2}$.