Biased and Asymptotically Unbiased Estimators

Omari Payne & Chandler Larkin

Biased and Unbiased Estimators

$$Bias = E(\widehat{\theta}_n) - \theta$$

Unbiased if the expected value of the Observed Estimator is equal to the Expected Estimator

In general, you must take many samples to determine if the estimator is biased

Asymptotically Unbiased if obs estimator is equal to the exp estimator as n -> ∞

$$\lim_{n \to \infty} E(\widehat{\theta}_n) = \theta$$

Asymptotically Normal

$$Z_n = \frac{\widehat{\theta}_n - \theta}{\frac{\sigma}{\sqrt{n}}}$$

An estimator is asymptotically normal if the limit of the CDF as n approaches ∞ is The Standard normal distribution for all x

$$\lim_{n \to \infty} F_{Z_n}(x) = \Phi(x)$$

5.11. (Do not submit; for mini-lectures) Consider estimating the population variance from a random sample of size n. Assume that the model is $N(\mu, \sigma^2)$, and take the estimator $\hat{\sigma}_n^2 = 1/n \sum_{i=1}^{n}(X_i - \bar{X}_n)^2$.

(a) Create a computer program to generate random values of $\hat{\sigma}_n^2$.

(b) Use simulations to "empirically" show that $\hat{\sigma}_n^2$ is biased.

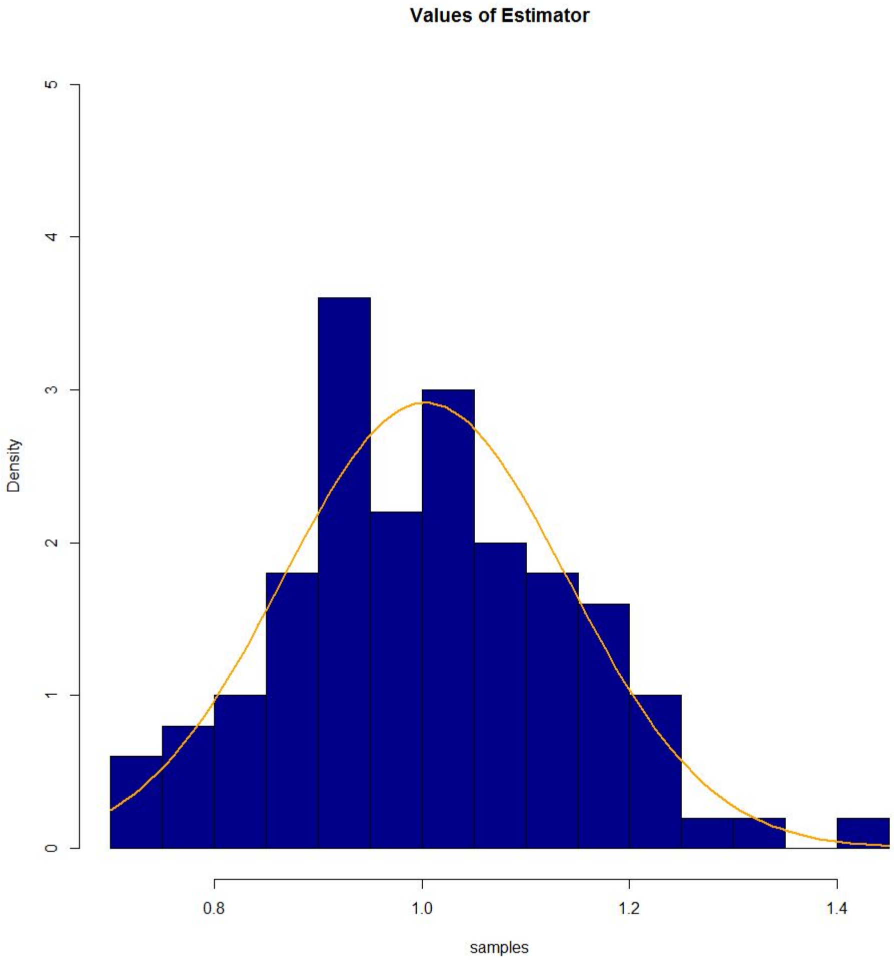
(c) Use simulations to empirically demonstrate that $\hat{\sigma}_n^2$ is consistent, thus asymptotically unbiased, and asymptotically normally distributed.

Repeat (a-c), but assume that the model is Exponential(θ).

Generated Estimators

$$\hat{\sigma}_n^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x}_n)^2}{n}$$

100 estimators of a normal distribution with 100 random values



Biased Estimators

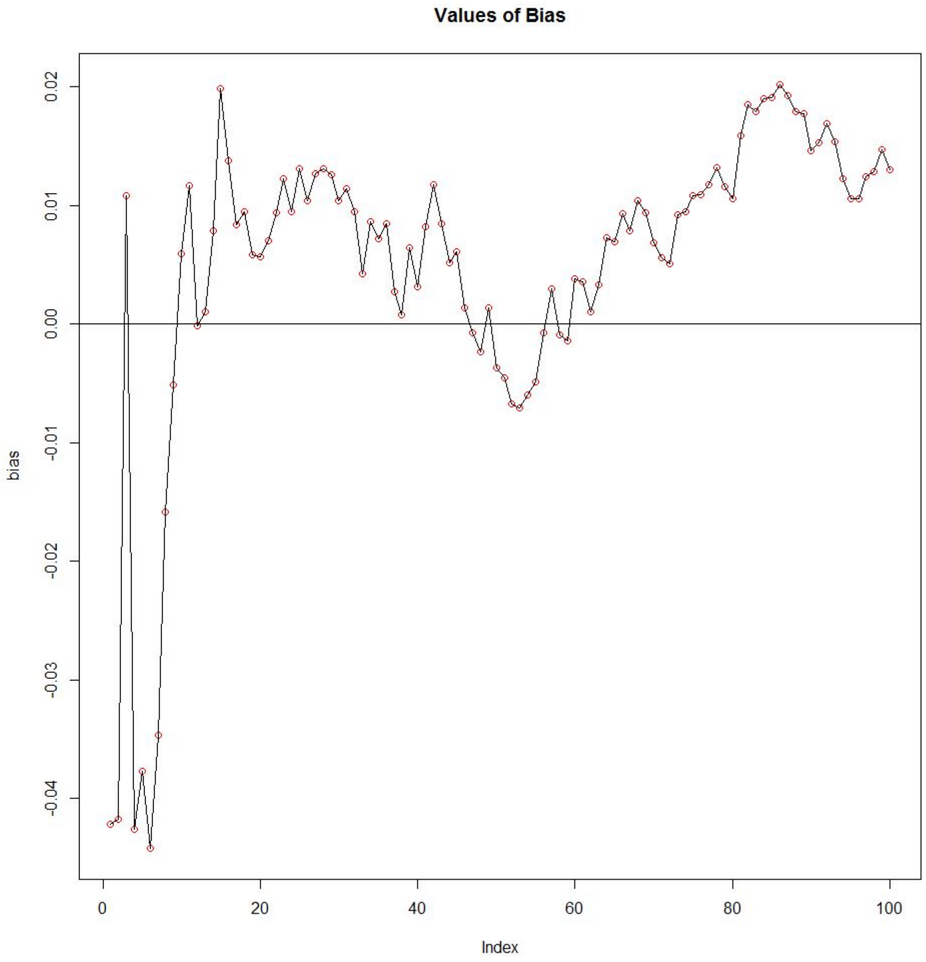
$$\sigma^2_{expected} = \frac{(n-1)}{n}$$

Progression of the bias of the estimators generated earlier

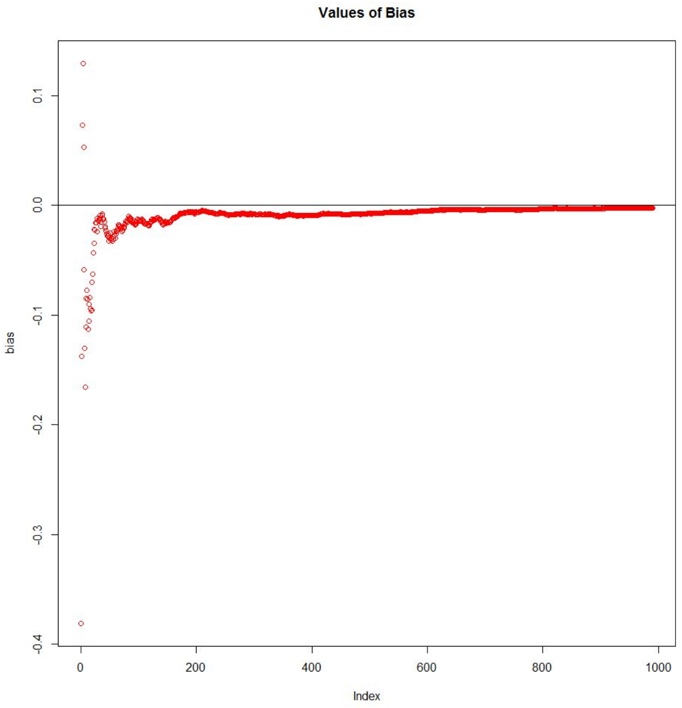
Each point is generated from the equation below as ‘n’ increase

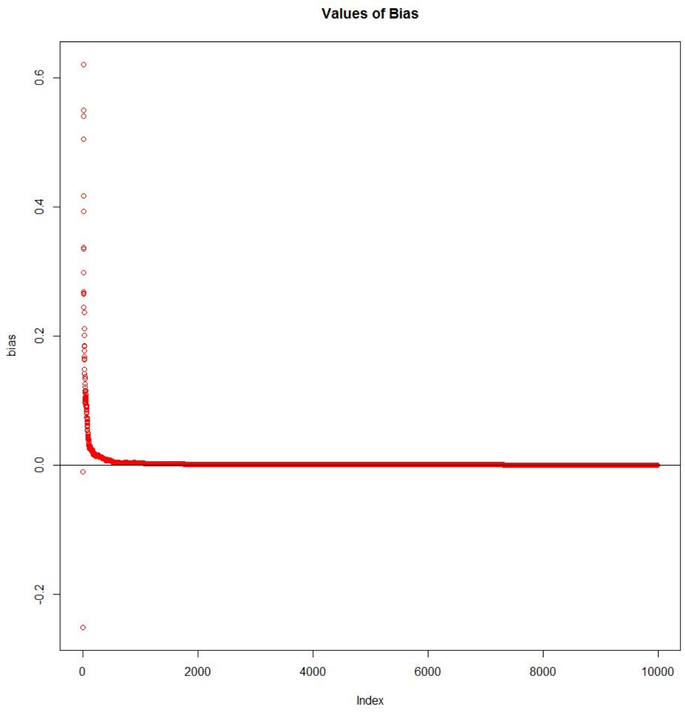
$$bias = E(\hat{\sigma}^2_n) - \sigma^2_{expected}$$

The estimators would be unbiased if they stayed near or around the zero-line



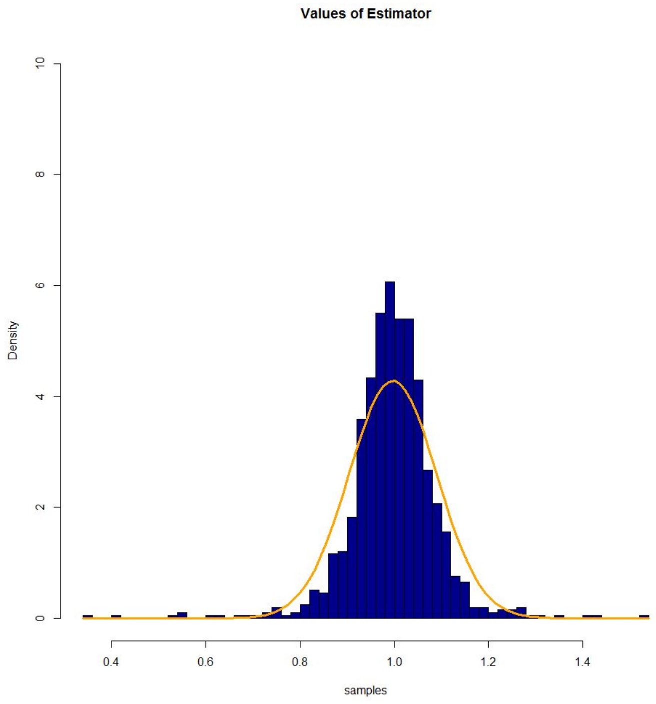
Asymptotically Unbiased with n = 1000 and n = 10000

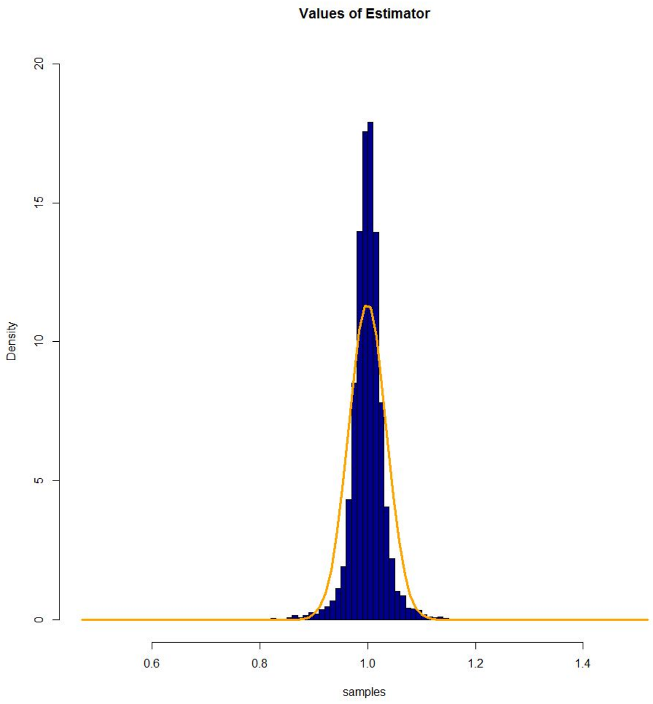




Asymptotically, the estimator is unbiased because the value of the estimator approaches 0 as n approaches infinity

Asymptotically Normal with n = 1000 and n = 10000





Asymptotically, the estimator becomes more and more normally distributed as the number of samples increases

Conclusion for Normal Distribution

Sample Size	Expected Value of Estimator	Observered Value of Estimator	Difference	Percent Error
100	0.99	1.003	.013	1.32%
1000	0.999	.997	.002	0.22%
10000	0.9999	1.000	.000	0.02%

It is obvious to see as N increases, the difference in the Expected Value of the Estimator and the actual observed value of the estimator decreases.

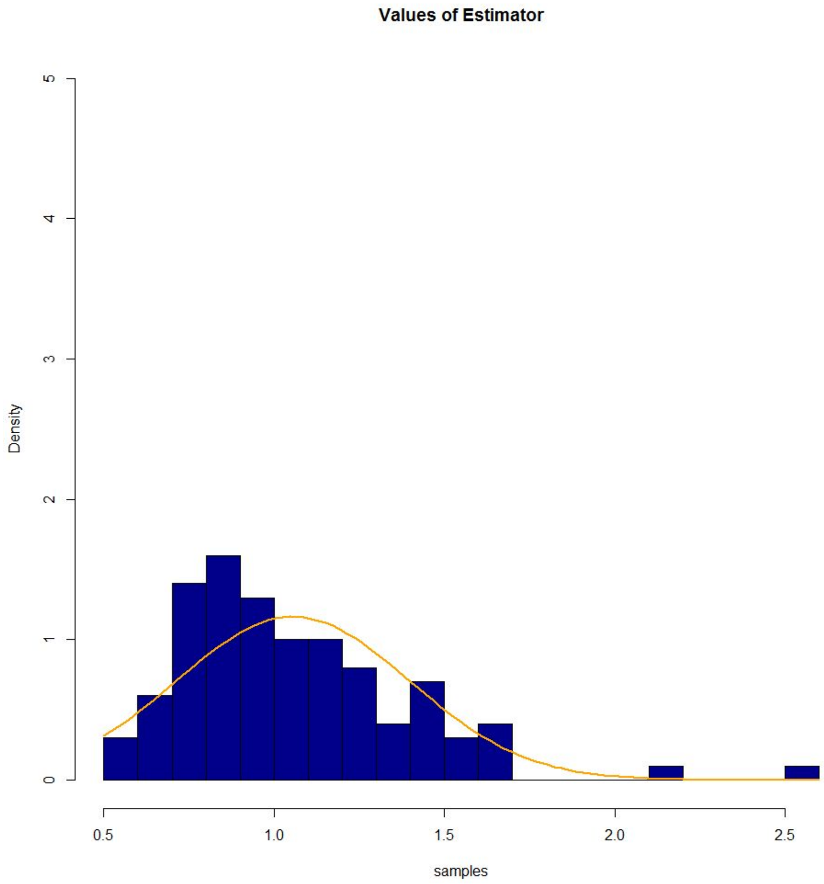
This means that in theory, if N could be infinite, the values would be exactly the same, and the estimator would be unbiased.

Generated Estimators

$$\hat{\sigma}_n^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x}_n)^2}{n}$$

100 estimators of an exponential distribution with 100 random values

The histogram of these values more closely resembles an exponential distribution which is to be expected when N is small



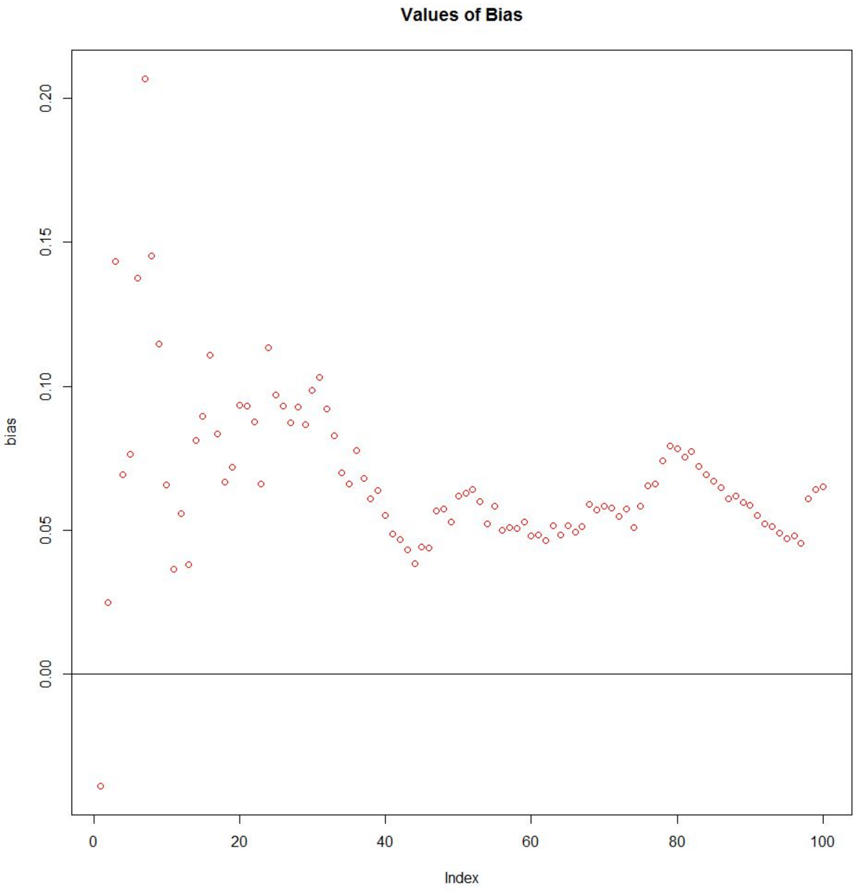
Biased Estimators

$$\sigma^2_{expected} = \frac{(n-1)}{n}$$

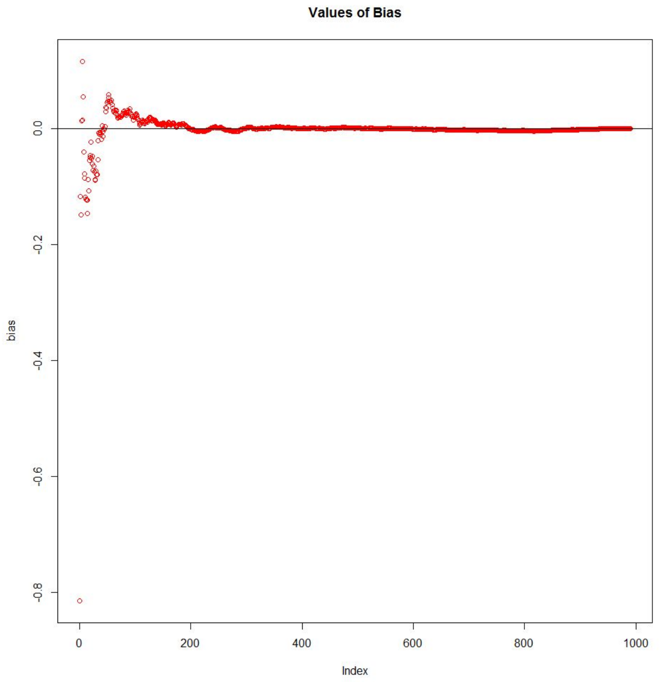
Progression of the bias of the estimators generated from the equation below as 'n' increase

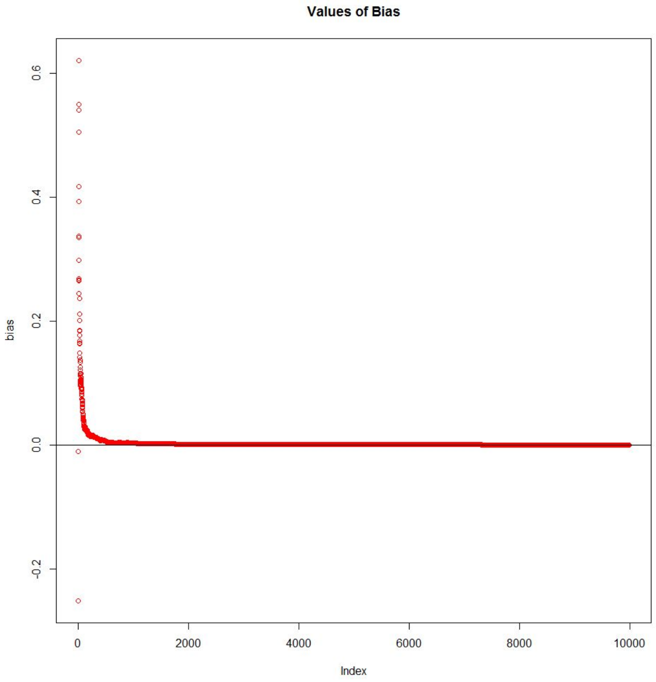
$$bias = E(\hat{\sigma}^2_n) - \sigma^2_{expected}$$

The estimators would be unbiased if they stayed near or around the zero-line and this simulation is easy to see that the estimator is biased as the difference is nowhere near the zero-line



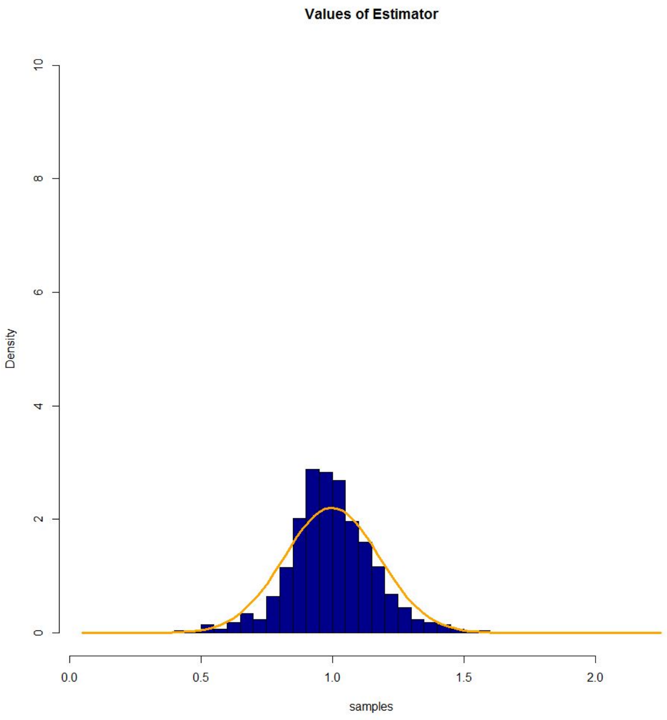
Asymptotically Unbiased with n = 1000 and n = 10000

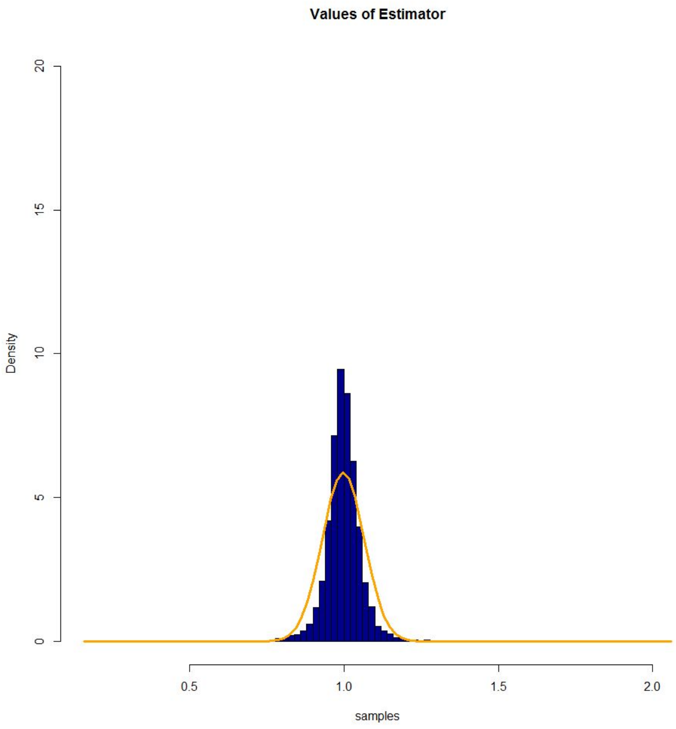




Asymptotically, the estimator is unbiased because the value of the estimator approaches 0 as n approaches infinity. This is very easy to see for the jump from N = 1000 to N = 10000

Asymptotically Normal with n = 1000 and n = 10000





Asymptotically, the estimator becomes more and more normally distributed as the number of samples increases. Both of these simulations show that the distribution is much more normally distributed than it was for N = 100

Conclusion for Exponential Distribution

Sample Size	Expected Value of Estimator	Observered Value of Estimator	Difference	Percent Error
100	0.99	1.055	.065	6.58%
1000	0.999	.998	.001	0.10%
10000	0.9999	1.000	.000	0.03%

Again, it is clear to see as N increases, the difference in the Expected Value of the Estimator and the actual observed value of the estimator decreases implying that it is asymptotically unbiased.

This simulation made it more obvious that the estimator is biased when N is small though, with a 6.58% error in the difference when N was 100.

Overall Conclusion

An estimator is unbiased if the expected value of the Observed Estimator is equal to the value of the Expected Estimator

Estimators are empirically biased when there is a small sample size of values

As you increase the number of values, the estimators become increasingly unbiased which implies that the estimator is asymptotically unbiased.

Additionally, as N increases, the distribution of the estimators more closely resembles a Normal Distribution implying that the estimator is asymptotically Normal