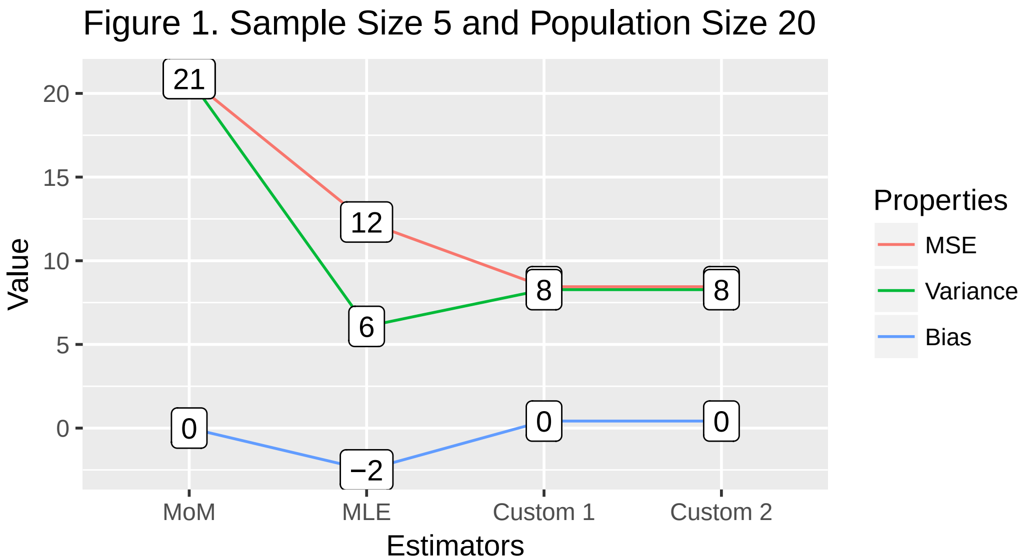
Results

James Yang

2/19/2018

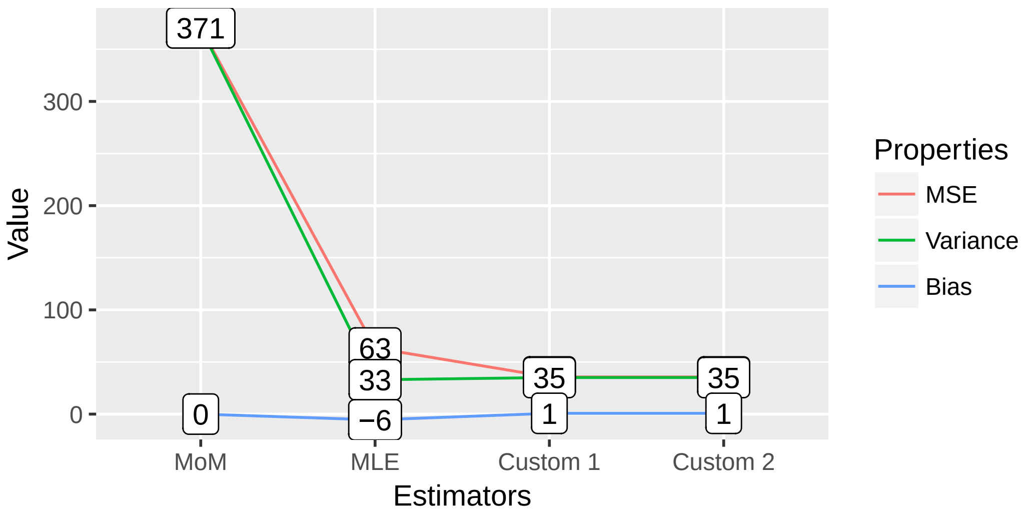
Results

In this section, we examine the four estimators (Method of Moment, Maximum Likelihood, and two of our custom estimators) by looking at the results of our simulation study. We have computed the mean squared error, variance as well bias for each of our estimator, given different sample sizes and population sizes as the input. Different inputs of sample size and population size are used to test if the performances of the estimators are consistent. We plot the properties of the estimators in the line charts displayed below.



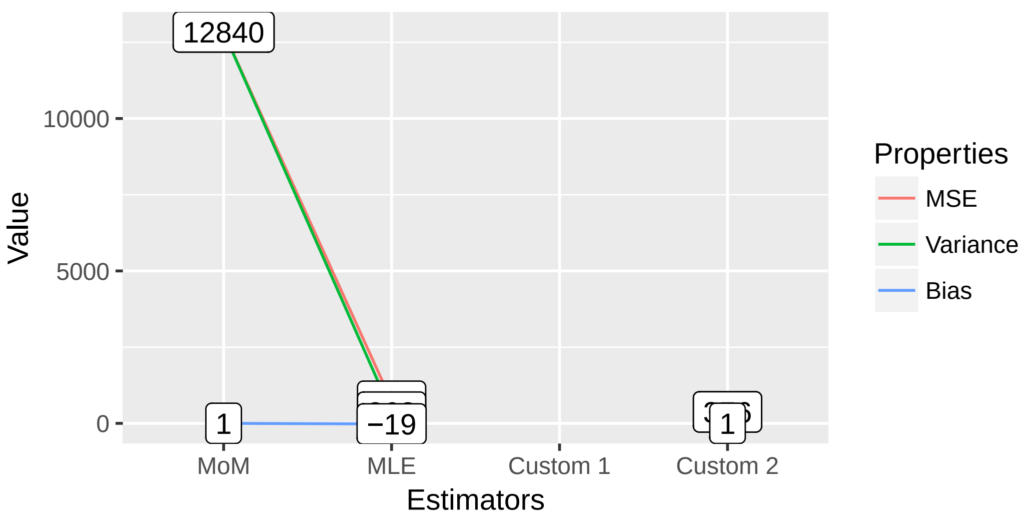
From `Figure 1`, we see that our first and second custom estimators are equally the best estimators in this case, since they both have the least mean squared errors and bias of 0. However, MLE has the least variance thus most efficient, but since its bias is significantly larger than that of the other three, its mean squared error is not optimal. In comparison, the method of moment estimator is much worse, given its much higher variance.

Figure 2. Sample Size 30 and Population Size 200



In `Figure 2`, after we increase the sample size and population size, we see that the trend between the four estimators is roughly the same as `Figure 1`, except that the mean squared error and variance of the method of moment estimator is much larger than those of the rest, which increases the slope of the line chart. Our two custom estimators still have the best performances among the four, in terms of MSE.

Figure 3. Sample Size 100 and Population Size 2000



Estimators	MSE	Variance	Bias
MoM	12840.76	12840.41	0.59
MLE	724.92	368.30	−18.88
Custom 1	NA	NA	NA
Custom 2	376.17	375.63	0.73

In `Figure 3`, we find that as sizes increase, the MSE of MoM estimator becomes significantly worse. Our MLE and second estimator are similar in terms of efficiency, but MLE has a worse bias thus MSE. Unfortunately, we get `NA` values for our first custom estimator, because in order to get the estimate, we need to first derive $max(sample\_values)^{(k+1)}$, which exceeds the limit of R's calculation. Given that our two custom estimators have similar performances given smaller sample sizes, and our second estimator has a consistently great performance, unlike our first one, which crashes under large sample/population sizes, we decide to use our second custom estimator given its low mean squared error compared to MoM and MLE, and its consistency in calculations compared to our first custom estimator.

After deciding the estimator, we use it to estimate N based on a random sample of 20 objects[1]. Our estimate for N is 2003 after rounding. We also perform bootstrap on the random sample, and compute the 95% bootstrap confidence interval for our estimate. We are 95% confident that the true N is between 1334.67 and 2003.05.

[1]The random sample used above consists of 922, 299, 1106, 121, 1621, 1164, 1912, 937, 914, 593, 85, 1090, 1004, 139, 1451, 24, 267, 1045, 1062, 1274.