Resampling Assignment-I

Supratim Das

MB2532

Question

Perform a suitable simulation study to evaluate the performance of the *jackknife* method and the bootstrap method for estimation of the variance of the sample median.

1. Perform a simulation study to evaluate the performance of the jackknife for estimation of the variance of the sample median.
2. Perform the same study for the sample mean.
3. Perform the same study for a non-parametric bootstrap.
4. Compare the performance in terms of

$$\frac{\widehat{\mathrm{Var}}(T_n)}{\mathrm{Var}(T_n)}.$$

   Compare your results with the theoretical results based on the Central Limit Theorem (CLT).

Methodology

Here, we will perform a simulation study to analyse the performance of the jackknife method and the boostrap method, and also compare their relative performance with respect to the asymptotic variance obtained by the large sample distribution for mean and median.

Simulation setup:

- At each step, samples of size 30, 70, 100 are taken to analyse the variation of performance of the various resampling estimators to estimate the variance of the relevant statistic. Also, for the purpose of the analysis, samples have been taken from two distributions, $\mathcal{N}(0,1)$ and $\chi^{\in}_{\in}$ to see whether assymetricity in the distributions affects our analysis.

- The performance metric used in our analysis is

$$\frac{\widehat{\mathrm{Var}}(T_n)}{\mathrm{Var}(T_n)}.$$

- We repeatedly take samples and then perform the resampling procedures to obtain multiple values of the performance metric. Then, we observe whether the performance statistic goes to 1 as that would indicate that the variance estimate thus obtained is consistent.

- **Jackknife simulation:** We compute the jackknife pseudovalues as

$$\tilde{T}_{n,i} = n\hat{T}_n - (n-1)\hat{T}_{(i)}$$

For getting the jackknife estimate of the variance of the statistic, we use

$$v_{jack} = \frac{1}{n(n-1)}\sum_{i=1}^{n}(\tilde{T}_{n,i} - \frac{1}{n}\sum_{j=1}^{n}\tilde{T}_{n,i})^2$$

- **Bootstrap simulation:** If our selected sample is of size N, then we perform SRSWR to select new resamples from the selected sample. For this analysis, we keep the size of our resamples to be N. We pick 1000 such resamples and compute the value of the statistic for each of the samples. For getting the bootstrap estimate of the variance of the statistic, we use the formula,

$$v_{boot} = \frac{1}{B}\sum_{b=1}^{B}(T^*_{n,b} - \frac{1}{B}\sum_{l=1}^{B}T^*_{n,l})^2$$

where $B = 1000$.

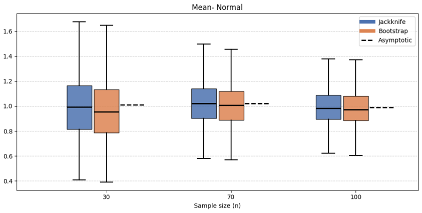
- **Asymptotic variance**:

  1. Mean:
  $$Var_{asymptotic} = \frac{\sigma^2}{n}$$
  2. Median:
  $$Var_{asymptotic} = \frac{\sigma^2}{4nf^2(\theta)}$$

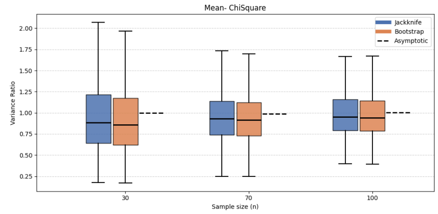
- For getting a good estimate of the true variance of the statistic against which we will benchmark the peroformance of the various resampling and large sample estimators obtained, we repeatedly sample a sample of size n for 10000 times and keep that as the benchamark.

Plots

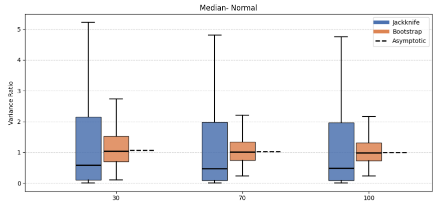
Figure 1: Simulation results under different settings



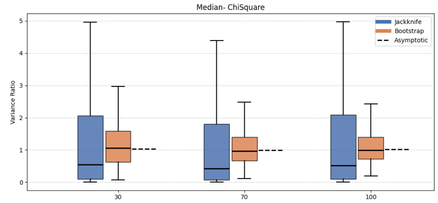
(a) Mean, $N(0,1)$



(b) Mean, χ_2^2



(c) Median, χ_2^2



(d) Median, $N(0,1)$

Tables

Table 1: Performance for estimating mean in $N(0,1)$ Distribution

	Jackknife Avg	Jackknife SD	Boot Avg	Boot SD	Asym
30	1.007000	0.268000	0.973000	0.263000	1.010000
70	1.026000	0.172000	1.010000	0.174000	1.019000
100	0.994000	0.145000	0.986000	0.150000	0.988000

Table 2: Performance for estimating mean in χ_2^2 Distribution

	Jackknife Avg	Jackknife SD	Boot Avg	Boot SD	Asymp
30	0.983000	0.486000	0.951000	0.473000	1.000000
70	0.971000	0.330000	0.957000	0.329000	0.986000
100	1.002000	0.291000	0.990000	0.292000	1.004000

Table 3: Performance for estimating median in $N(0,1)$ Distribution

	Jackknife Avg	Jackknife SD	Boot Avg	Boot SD	Asymp
30	1.993000	3.961000	1.221000	0.805000	1.030000
70	1.906000	5.001000	1.102000	0.616000	0.989000
100	2.005000	3.891000	1.111000	0.557000	1.015000

Table 4: Performance for estimating median in χ^2_2 Distribution

	Jackknife Avg	Jackknife SD	Boot Avg	Boot SD	Asymp
30	1.988000	4.036000	1.186000	0.687000	1.065000
70	1.904000	3.958000	1.107000	0.517000	1.026000
100	1.863000	3.787000	1.074000	0.466000	1.004000

Findings

- From tables 3 and 4, we can see that the performance statistic for the jackknife variance estimator for the median does not converge to 1, and also the standard deviation of the jackknife variance estimator for the median remains considerably large, giving further credence to the claim that the jackknife variance estimator for median is **not consistent**. The same can be seen in 1(c) and 1(d), where the mean of the the jackknife varaince estimates is not near 1 and the box is also spread out indicating the lack of consistency for the estimator.

- On the other hand, for the mean, the Jackknife variance estimator is quite efficient, outperforming the bootstrap variance estimator, though only by a small margin (tables 1 and 2). Both the jackknife and bootstrap estimators are quite efficient at estimating variance of the mean. Also, the standard deviations **decrease** with increase in sample size indicating that the resampling variance estimators are indeed **consistent** for the variance of the mean. Similar observations can be made from plots 1(a) and (b), where the mean of the performance statistics of both the resampling estimators is very close to 1.

- From tables 3 and 4, we can also see that the boostrap variance estimator is **consistent** for estimating the variance as the performance metric is going to 1 on average and the standard deviations are also shrinking. Plots 1(c) and (d) support the same.

- In the case of the mean, the jackknife estimator slightly outperforms the large sample approximation variance estimator obtained from CLT. However the performance gain is not much high. The bootstrap estimator also performs almost on par with the asymptotic variance estimator, albeit with slightly degraded performance.

Conclusion

In this assignment, we explored various resampling estimators like the jackknife estimator and the bootstrap estimator for the mean and median statistics and how good they are for estimating the variance of the statistic. For mean and median, we benchmarked the resampling estimators' performance against the true variance of the statistics, and also compared them with the asymptotic variance obtained from the large sample approximation of mean and median(CLT). We observed that for the mean, both the jackknife estimator and the bootstrap estimator work quite well, approximating the true variance with quite great accuracy despite small sample size. However, in the case of the median, the jackknife variance estimator is found to not be consistent for the variance of the median. However, the bootstrap variance estimator is still consistent in the case of the median (quite close to the true variance) and is very near in performance to the asymptotic variance . The non-consistency of the jackknife variance estimator indicates that delete-1 jackknife mmay not be enough for inference of the statistic, pointing us to delete-d jackknife (a possible improvement).

References

[1] J. Shao and D. Tu, *Jackknife and Bootstrap*, Springer Series in Statistics, Springer, New York, 1995.

[2] GitHub link for code: https://github.com/Supratim2004/Resampling-Project