Journal of Mathematics and Statistics 6 (3): 347-349, 2010
ISSN 1549-3644

Subsample Goal Model for Multihalver on Outliers

B. Onoghojobi
Department of Statistics, Federal University of Technology, Owerri, Nigeria

Abstract: Problem statement: In this study, a delete-half jackknife problem reformulated as a subsample multihalver was presented. **Approach:** In this respect, exploiting outlier nomination and estimation, since considering all possible half-sample is unpractical and unfeasible were considered. **Results:** We derived subsample algorithm which is unbiased multihalver and the performance of the model in formulating the subsample multihalver was shown. **Conclusion:** The result of subsample multihalver method of nomination and estimation is better way of resolving large population.

Key words: Subsample multihalver, unbiased multihalver, outlier nomination, delete-half

INTRODUCTION

The Robust version of the Jackknife using the trimmed mean of the pseudovalues and a general M-estimator base on the pseudovalues offer moderately robust alternatives to Jackknife with goal asymptotic properties, but quite limited small sample results (Fernholz *et al*., 2004). However, if the population is large but uniformly contaminated with outlier, the effective of the leave out half Jackknife is over shadowed by the size.

The moderate discussion of trimmed and M-estimator has been on Jackknife (Hinkley and Wang, 1980; Cheng, 1991). The original delete-one Jackknife use the delete-k resampling method (Turkey, 1979). Multihalver was used to detect and estimate outlier Fernholz *et al*., 2004). A robust rank-based nonparameteric spectral estimation was introduced for detecting periodicity in nonideal dataset (Pearson *et al*., 2003).

The main stream approach for connecting outliers had been executed with caution (Grane and Veiga, 2009). It is an established fact that neglecting the existence of some outlier during the estimation phase of the detection methodology may entail to end up with biased parameter (Van Dijk *et al*., 1999).

In this present study, we attempt to oversome the difficulties associated with large population but uniformly contaminated with outlier, by reformulating the Jackknife robust version using trimmed and M-estimator as a subsample goal technique and solving same by the usual procedure.

MATERIALS AND METHODS

The tools and logic for this research are similar to the research of (Fernholz *et al*., 2004) when dealing with multihalver and subample multihalver.

The subsample muiltihalver: For a sample of size n and, a T statistic, the subsample multihalver use the resample of size n/k *(*where k is a multiple of n) with a reasonable number of subsample halving which are splits of data in subgroup as required. The number of all possible subsample halving is $\binom{n}{n/k}/2 \sim 2^n/\sqrt{2\pi_n}$, which grow to n quickly is an improvement on (Fernholz *et al*., 2004). It is desirable to have intersecting subsample approach orthogonality, wherever possible and convenient to have them reasonably similarly related. For the sample size of k, this can be achieved by repeated Hadamard matrices. The case when n is a multiple of four is an instant (Plackett and Burman, 1946) which is obtained using the sequence from Plackett-Burman. Halving can be described in two ways, either as a pairing of observations leading to a split of the data in half, or as two complementary half sample. Hence, subsample multivalve can be describe as a sampling of observation leading to a slit of data into k delete subsample or a k complementary delete subsample. If subsampling is described as sampling by pairing and repairing the existing pairs. Then, it is understood that we select the left-sample L by taking the first element in each pair and the right half-sample R by taking the second element in each pair. Thus, the pairing:

$$\omega = \{(y_1, y_2), (y_3, y_4), \cdots, (y_{n-1}, y_n)\} \tag{1}$$

which correspond to the halving with two complementary half samples:

$$L = \{y_1, y_3, \cdots, y_{n-1}\} \text{ and } R = \{y_2, y_4, \ldots, y_n\} \tag{2}$$

Hence, we can further obtained halving from L and R in (2) by repeating the pair procedures above, therefore we have:

$$L_L = L = \{y_1, y_5, \cdots, y_{n-3}\} \text{and} \, L_R = \{y_3, y_n, \ldots, y_{n-1}\} \qquad (3)$$

Similarly:

$$R_L = \{y_2, y_6, \cdots, y_{n-2}\} \text{and} \, R_R = \{y_4, y_8, \ldots, y_n\} \qquad (4)$$

This procedure continues until a k subsample is obtained recursively. This k subsample is the proposed subsampling goal model. Thus, we have $s_1, s_2, \ldots, s_k$ subsamples established:

Hadamard matrices and associated sub sampling result: Let:

$$H_1 = \begin{bmatrix} +1 & +1 \\ +1 & -1 \end{bmatrix} \qquad (5)$$

be a Hadamard matrix of order 2. Where H_1 is an array of data set and $+1_{s \text{ and}}$ -1 are subdivision of H_1 i.e., the data set H_1 has been divided into four groups. Hence, the matrix is with entries equal ±1 and orthogonal rows and columns. We can construct Hadamard matrices of higher order recursively by the operation:

$$H_k = \begin{bmatrix} H_{k-1} & H_{k-1} \\ H_{k-1} & -H_{k-1} \end{bmatrix} \text{for } k \geq 2 \qquad (6)$$

The details of Hadamard matrices for pairing and having had be shown (Fernholz *et al*., 2004) and it is based on the principle of pairing and halving, we derived the subsampling Multrihalver Method; which can be stated as a process of continuous fitting of Hadamard matrices and it is associated pairing until the kth order Hadamard matrices is obtained. As shown below. Let:

$$H_2 = \begin{Bmatrix} H_1 & H_1 \\ H_1 & -H_1 \end{Bmatrix} \qquad (7)$$

where, H_1 is as defined in (5), therefore

$$H_2 = \begin{bmatrix} H_1 & H_1 & H_1 & H_1 \\ H_1 & -H_1 & H_1 & -H_1 \\ H_1 & H_1 & -H_1 & -H_1 \\ H_1 & -H_1 & -H_1 & H_1 \end{bmatrix} \qquad (8)$$

Hence:

$$H_3 = \begin{bmatrix} H_2 & H_2 \\ H_2 & -H_2 \end{bmatrix} \qquad (9)$$

$$H_k = \begin{bmatrix} H_{k-1} & H_{k-1} \\ H_{k-1} & -H_{k-1} \end{bmatrix} \qquad (10)$$

The resultant imagination of the abstract version of H_k by substituting $H_{k-1}\ldots$, H_1 gives the proposed the matrix of sub sampling goal model. This is also a good example of the kth delete subsample or k complementary delete subsample.

RESULTS AND DISCUSSION

Estimation based on the sub sampling multivalve: Let γ be the set of subsamples obtained by Hadamard matrice method described above. For each subsamples $s_i \in \gamma$ we compute the difference L-R or h_L-h_R of the statistic T, where h_L and h_R are the values of h on the left and right subsamples of the halver. For each j, we furthermore defined a pseudovalue for the subsample using the extensions ideas as:

$$T_j^* = iT - \left[\frac{h_L + h_R}{i}\right] i, j = 1, \ldots, n \qquad (11)$$

The subsample estimate associated to T is then:

$$T_{ss} = \frac{1}{H} \sum_{j \in x} T_j^n \qquad (12)$$

This shows that the estimate for subsample and multihalver are alike, expect for the fact that subsample reduce the analytical and computation process before applying multipltihalver method. Hence, it is obvious that the subsample estimate the standard deviation of T as:

$$S_{e_{ss}} = \frac{1}{H} \Sigma \left[\frac{h_L - h_r}{N^2}\right]^{1/2} \qquad (13)$$

where, H is equal to total number of Subsample Hadamard metrics.

CONCLUSION

It is obvious that outlier nomination base on multihalver is applied to a subsample multihalver. This subsample result will then act as an unbias solution to the entire problem, thereby reducing the stress of working with the whole population.

The subsample goal technique has be shown to provide a better outlier nomination and estimation when the population is very large than the conventional Jackknife Robust version using trimmed mean of the pseudovalue and m-estimator based on pseudovalues.

REFERENCES

Cheng, K.F., 1991. M-estimator using Jackknife psedovalues. Scand. J. Stat., 18: 51-61.

Fernholz, L.T., S. Morgenthaler and J.W. Tukey, 2004. An outlier nomination method based on the multihalver. J. Stat. Plan. Inference, 122: 125-127. DOI: 10.1016/j.jspi.2003.06.008

Grane, A. and H. Veiga, 2009. Wavelet-based detection of outliers in models. University of Carles. http://e-archivo.uc3m.es/bitstream/10016/3507/5/ws090403.pdf

Hinkley, D. and H.L. Wang, 1980. A trimmed Jackknife. J. R. Stat. Soc. Ser. B, 42: 347-356. http://www.jstor.org/stable/2985171

Plackett, R.L. and J.P. Burman, 1946. The design of optimum multifactorial experiments. Biometrika, 33: 305-325. DOI: 10.1093/biomet/33.4.305

Pearson, R.K., H. Lahdesmaki, H. Huttumen and O. Yli-Harja, 2003. Detecting periodicity in nonideal dataset. Proceeding of the SIAM International Conference on Data on Data Mining, (DDM'03), Cathedral Hill Hotel, San Francisco, CA., pp: 1-5. http://www.cs.tut.fi/~harrila/research/SIAM2003Spectral.pdf

Turkey, J.W., 1979. Trendless Jackknifing of leter-values (and functions thereof). Princeton University.

Van Dijk, D., P.H. Frances and A. Lucas, 1999. Testing for ARCH in the presence of Additive outliers. J. Applied Econ., 14: 539-562. http://ideas.repec.org/a/jae/japmet/v14y1999i5p539-62.html