Journal of the Korean
Data & Information Science Society
2002, Vol. 13, No.1, pp. 147 ~ 156

On the Use of Winsorized Mean for Truncated Family of Distributions under Type II Censoring[1]

A.Nanthakumar[2], K.Selvavel [3] and M.Masoom Ali [4]

Abstract

In this paper, we study the properties of the modified winsorized mean to estimate the mean of a two-truncation parameter population. Under some mild conditions, the estimator is found to be strongly consistent and asymptotically unbiased even though the sample is doubly type II censored.

Key Words and Phrases: Modified winsorized mean, Truncated distribution.

1 Introduction

The truncated distributions are those distributions for which one or both of the extremities of the range are functions of the unknown parameters. These families were first studied by Davis (1951) and later by many authors including Hogg and Craig (1956), Bar-Lev and Boukai (1985), Rohatgi (1989), Selvavel (1989), Feretinos (1990), Nanthakumar and Selvavel (1994, 1996). In this paper, we consider a population that is truncated between two unknown values (parameters) where a sample from this population is supposedly subject to double type II censoring. We use the modified winsorized mean computed from the censored sample to estimate the population mean.

Tukey (1960) was the first to study the use of robust estimators. The winsorized mean is a robust estimator and it is used to estimate the mean of a symmetric probability distribution. Also, it is found to be more efficient than some other robust

[1]The views expressed are attributible to the authors and do not necessarily reflect the views of the Department of Defense.

[2]Department of Mathematics SUNY-Oswego, Oswego, NY 13126 USA
[3]Department of Defense Arlington, VA 22202, USA
[4]Department of Mathematical Sciences Ball State University Muncie, IN 47306-0490 USA
E-mail: mali@bsu.edu

estimators such as the trimmed mean (see David (1970) for the details). The two-truncation parameter family of distributions is asymmetric and so we use a modified winsorized mean to estimate the mean of the population. We compute the modified winsorized mean by using a sample that is doubly type II censored and study the properties under some reasonable assumptions. The modified winsorized mean is shown to be asymptotically strong consistent and asymptotically normal, as the sample size becomes larger.

We describe the method and the related results in Section 2. The numerical (simulation) results are presented in Section 3.

2 Methodology and Main Results

In this section, we consider the use of modified winsorized mean to estimate the mean of the truncated population having a pdf of the form

$$f(x) = q(\theta_1, \theta_2)h(x),\ \theta_1 < x < \theta_2, \tag{2.1}$$

where θ_1 and θ_2 are unknown truncation parameters, $h(x)$ is an absolutely continuous function and $q(\theta_1, \theta_2)$ is a normalizing constant.

Definition: *By type II censoring, we mean that in a potential sample, a known number of observations are missing at either end (single censoring) or at both ends (double censoring).*

Suppose a situation permits censoring (both right and left) and that in a sample the first $r-1$ smallest observations and the last $n-s$ largest observations are assumed to be missing. The objective is to use the available observations to estimate the population mean. Here, we use the following modified winsorized mean

$$W_n(r, s) = \frac{1}{n}\left(rY_{(r)} + (n-s)Y_{(s)}\right) + \frac{1}{n}\sum_{i=r+1}^{s} Y_{(i)} \tag{2.2}$$

to estimate the population mean.

Theorem 1 *Let $Y_1, Y_2, \ldots, Y_n$ be a random sample from a distribution function $F(x)$ with a pdf $f(x)$ given by (2.1). Then for the ith order statistic $Y_{(i)}$*

$$E(Y_{(i)}) = Q\left(\frac{i}{n+1}\right) + O\left(n^{-2}\right),$$

where $Q(\cdot) = F^{-1}(\cdot)$.

Proof: It is well-known fact that the ith uniform order statistic $U_{(i)}$ can be written as $U_{(i)}) = F(Y_{(i)})$. That is $Y_{(i)} = F^{-1}(U_{(i)}) = Q(U_{(i)})$. Now, by Taylor Series expansion of $Q(U_{(i)})$ about $U_{(i)} = \frac{i}{n+1}$, we get

$$Q(U_{(i)}) = Q\left(\frac{i}{n+1}\right) + \left(U_{(i)} - \frac{i}{n+1}\right) Q'\left(\frac{i}{n+1}\right) + \frac{1}{2}\left(U_{(i)} - \frac{i}{n+1}\right)^2 Q''(Z), \tag{2.3}$$

where $\frac{i}{n+1} < Z < U_{(i)} = F(Y_{(i)}) < F(\theta_2)$.

By taking expectations on both sides of (2.3), we get

$$E[Q(U_{(i)})] = Q\left(\frac{i}{n+1}\right) + E\left[\left(U_{(i)} - \frac{i}{n+1}\right)^2 \frac{Q''(Z)}{2}\right]. \tag{2.4}$$

Since Z is a bounded random variable, we have

$$C_1 \frac{i(n-i+1)}{(n+2)(n+1)^2} < E\left[\left(U_{(i)} - \frac{i}{n+1}\right)^2 Q''(Z)\right] \leq C_2 \frac{i(n-i+1)}{(n+1)(n+2)^2}, \tag{2.5}$$

where C_1 and C_2 are constants.

By combining (2.4) and (2.5), we get

$$E(Y_{(i)}) = Q\left(\frac{i}{n+1}\right) + O(n^{-2}).$$

Corollary 1. *If $\frac{r}{n}$ and $\frac{n-s}{n}$ are of order n^{-a} for some $0 < a < 1$, then $E(W_n(r,s)) = E(Y) + O(n^{-a})$, where $W_n(r,s)$ is the modified winsorized mean and $Y \sim F(\cdot)$.*

Proof: Note that

$$\begin{aligned} E(W_n(r,s)) &= \frac{r}{n}E(Y_{(r)}) + \frac{n-s}{n}E(Y_{(s)}) + \frac{1}{n}\sum_{i=r+1}^{s} E(Y_{(i)}) \qquad (2.6) \\ &= \frac{r}{n}Q\left(\frac{r}{n+1}\right) + \frac{n-s}{n}Q\left(\frac{s}{n+1}\right) + \frac{1}{n}\sum_{i=r+1}^{s} Q\left(\frac{i}{n+1}\right) + O(n^{-2}) \\ &= \frac{r}{n}Q\left(\frac{r}{n}\right) + \frac{n-s}{n}Q\left(\frac{s}{n}\right) + \frac{1}{n}\sum_{i=r+1}^{s} Q\left(\frac{i}{n}\right) + O(n^{-2}) \\ &= \frac{r}{n}Q\left(\frac{r}{n}\right) + \frac{n-s}{n}Q\left(\frac{s}{n}\right) + \int_{\frac{r}{n}}^{\frac{s}{n}} Q(y)dy + O(n^{-2}) \\ &= \frac{r}{n}Q\left(\frac{r}{n}\right) + \frac{n-s}{n}Q\left(\frac{s}{n}\right) + \frac{s}{n}Q\left(\frac{s}{n}\right) - \frac{r}{n}Q\left(\frac{r}{n}\right) - \int_{\frac{r}{n}}^{\frac{s}{n}} ydQ(y) \\ &\quad +O(n^{-2}). \end{aligned}$$

This implies

$$\begin{aligned}
E(W_n(r,s)) &= Q\left(\frac{s}{n}\right) - \int_{\frac{r}{n}}^{\frac{s}{n}} y dQ(y) + O(n^{-2}) \qquad (2.7)\\
&= Q(1) - \left(1-\frac{s}{n}\right)Q'(1) - \int_0^1 y dQ(y) + \int_{\frac{s}{n}}^1 y dQ(y) + \int_0^{\frac{r}{n}} y dQ(y)\\
&\quad + O(n^{-2}) \qquad (2.8)\\
&= Q(1) - \int_0^1 y dQ(y) + O(n^{-2}) + O(n^{-a})\\
&= \int_0^1 Q(y)dy + O(n^{-a}) = \int_{\theta_1}^{\theta_2} y dF(y) + O(n^{-a}) = E(Y) + O(n^{-a}).
\end{aligned}$$

Note that the modified winsorized mean is asymptotically unbiased in estimating the mean of the two-truncation parameter distribution.

Next we investigate the strong consistency of the modified winsorized mean.

Theorem 2 *If $\frac{r}{n}$ and $\frac{n-s}{n}$ are of order n^{-a} for $0 < a < 1$, then $W_n(r,s) \to E(Y)$ as $n \to \infty$, where Y is from a population having pdf of the form given in (2.1).*

Proof: Consider the modified winsorized mean

$$\begin{aligned}
W_n(r,s) &= \frac{1}{n}(rY_{(r)} + (n-s)Y_{(s)}) + \frac{1}{n}\sum_{i=s+1}^{s} Y_{(i)} \qquad (2.9)\\
&= \frac{r}{n}Y_{(r)} + \frac{n-s}{n}Y_{(s)} + \frac{1}{n}\sum_{i=1}^{n} Y_{(i)} - \frac{1}{n}\sum_{i=1}^{r} Y_{(i)} - \frac{1}{n}\sum_{i=s+1}^{n} Y_{(i)}.
\end{aligned}$$

This implies

$$\begin{aligned}
|W_n(r,s) - \bar{Y}| &= \left|\frac{r}{n}Y_{(r)} + \frac{n-s}{n}Y_{(s)} - \frac{1}{n}\sum_{i=1}^{r} Y_{(i)} - \frac{1}{n}\sum_{i=s+1}^{n} Y_{(i)}\right| \qquad (2.10)\\
&\le \frac{r}{n}Y_{(r)} + \frac{n-s}{n}Y_{(s)} + \frac{1}{n}\sum_{i=1}^{r} Y_{(i)} + \frac{1}{n}(n-s)Y_{(s)}.
\end{aligned}$$

Therefore, $|W_n(r,s) - \bar{Y}| \le 2\left(\frac{r+n-s}{n}\right)\max(|\theta_1|, |\theta_2|) \to 0$ as $n \to \infty$. But due to the strong law of large numbers $\bar{Y} \stackrel{\text{a.s.}}{\to} E(Y)$ as $n \to \infty$. Hence $W_n(r,s) \stackrel{\text{a.s.}}{\to} E(Y)$ as $n \to \infty$.

The following Theorem establishes the asymptotic normality of the modified winsorized mean.

Theorem 3 $Y_{(i)} - E(Y_{(i)}) \stackrel{\text{a.s.}}{\to} 0$ *as $n \to \infty$.*

Proof: By Kolmogorov's inequality

$$\begin{aligned} P(\max|Y_{(i)} - E(Y_{(i)})| > \epsilon) &< \sum_{i=1}^{n} \frac{VarY_{(i)}}{\epsilon^2} \qquad (2.11) \\ &< \sum_{i=1}^{n} \frac{r(n-r)}{n^3} \left[\left(F^{-1}\left(\frac{r}{n}\right)\right)' \right]^2 + O(n^{-5}) \\ &< O(n^{-a}) \overset{\text{a.s.}}{\to} 0 \text{ as } n \to \infty. \end{aligned}$$

Theorem 4 *If* $\frac{r}{n}$ *and* $\frac{n-s}{n}$ *are of* $O(n^{-a})$ *for* $0 < a < 1$*, then (i)* $Var(Y_{(r)}) = O(n^{-(1+a)})$*, (ii)* $Var(Y_{(s)}) = O(n^{-(1+a)})$.

Proof: To prove (i), consider

$$\begin{aligned} Var(Y_{(r)}) &= \left(\frac{r}{n+1}\right)\left(\frac{n-r+1}{n+1}\right)\left(\frac{1}{n+2}\right)\left[Q'\left(\frac{r}{n}\right)\right]^2 + O(n^{-4}) \qquad (2.12) \\ &= \left(\frac{1}{n+2}\right)\left(\frac{r}{n}\right)\left(1-\frac{r}{n}\right)\left[f\left(Q\left(\frac{r}{n}\right)\right)\right]^{-2} + O(n^{-4}) \\ &= O(n^{-(1+a)}) + O(n^{-4}) = O(n^{-(1+a)}), \end{aligned}$$

where $f = \frac{dF}{dx}$, $Q(\cdot) = F^{-1}(\cdot)$, and $F(\cdot)$ is the cdf of Y.

We can prove (ii) along the same lines.

Theorem 5 $Var(W_n(r,s)) = \begin{cases} O(n^{-1}) & \text{if } \ a \geq \frac{1}{3}, \\ O(n^{-3a}) & \text{if } \ 0 < a < \frac{1}{3}. \end{cases}$

Proof: By using the Cauchy-Schwarz inequality, we obtain

$$Var(W_n(r,s)) \leq \frac{1}{n}\left(r^2 Var(Y_{(r)}) + (n-s)^2 Var(Y_{(s)})\right) + \frac{1}{n}\sum_{i=r+1}^{s} Var(Y_{(i)}). \qquad (2.13)$$

Since we are dealing with bounded order statistics, we have the result as the variance terms of $Y_{(i)}$ are of order n^{-1}.

Theorem 6 *If* $\frac{r}{n}$ *and* $\frac{n-s}{n}$ *are of order* n^{-a} *for some* $a \geq \frac{1}{3}$*, then* $\frac{W_n(r,s)-E(W_n(r,s))}{\sigma_{W_n(r,s)}} \xrightarrow{\mathcal{L}} N(0,1)$ *as* $n \to \infty$.

Proof: One can write

$$\begin{aligned} W_n(r,s) - E(W_n(r,s)) &= \bar{Y} - E(\bar{Y}) + \frac{r}{n}\left(Y_{(r)} - E(Y_{(r)})\right) + \frac{n-s}{n}\left(Y_{(s)} - E(Y_{(s)})\right) \\ &\quad -\frac{1}{n}\sum_{i=s+1}^{n}\left(Y_{(i)} - E(Y_{(i)})\right) - \frac{1}{n}\sum_{i=1}^{r}\left(Y_{(i)} - E(Y_{(i)})\right). \qquad (2.14) \end{aligned}$$

Dividing both sides by $\sigma_{W_n(r,s)}$ we have

$$\begin{aligned}\frac{W_n(r,s)-E(W_n(r,s))}{\sigma_{W_n(r,s)}} &= \frac{(\bar{Y}-E(\bar{Y}))}{\sigma_{W_n(r,s)}}\frac{\sigma_{\bar{Y}}}{\sigma_{\bar{Y}}}+\frac{r}{n}\frac{\left(Y_{(r)}-E(Y_{(r)})\right)}{\sigma_{W_n(r,s)}} \\ &+\frac{n-s}{n}\frac{\left(Y_{(s)}-E(Y_{(s)})\right)}{\sigma_{W_n(r,s)}}-\frac{1}{n}\sum_{i=s+1}^{n}\frac{\left(Y_{(i)}-E(Y_{(i)})\right)}{\sigma_{W_n(r,s)}} \\ &-\frac{1}{n}\sum_{i=1}^{r}\frac{\left(Y_{(i)}-E(Y_{(i)})\right)}{\sigma_{W_n(r,s)}}. \end{aligned} \tag{2.15}$$

Also $\frac{\sigma_{\bar{Y}}}{\sigma_{W_n(r,s)}} \to 1$ as $n \to \infty$ and

(i) $\frac{r}{n}\frac{\left(Y_{(r)}-E(Y_{(r)})\right)}{\sigma_{W_n(r,s)}} \xrightarrow{\mathcal{L}} 0$,

(ii) $\left(\frac{n-s}{n}\right)\frac{\left(Y_{(s)}-E(Y_{(s)})\right)}{\sigma_{W_n(r,s)}} \xrightarrow{\mathcal{L}} 0$,

(iii) $\frac{1}{n}\sum_{i=s+1}^{n}\frac{\left(Y_{(i)}-E(Y_{(i)})\right)}{\sigma_{W_n(r,s)}} \xrightarrow{\mathcal{L}} 0$,

(iv) $\frac{1}{n}\sum_{i=1}^{r}\frac{\left(Y_{(i)}-E(Y_{(i)})\right)}{\sigma_{W_n(r,s)}} \xrightarrow{\mathcal{L}} 0$, as $n \to \infty$.

Note that $\frac{\bar{Y}-E(\bar{Y})}{\sigma_{\bar{Y}}} \xrightarrow{\mathcal{L}} N(0,1)$ as $n \to \infty$. Hence by the central limit theorem along with Slutsky's theorem, we obtain the desired result.

3 Simulation Results

The following numerical results are based on sample size $n = 100$ for a two-truncation parameter exponential distribution function.

Table 1: $\theta_1 = 2$ and $\theta_2 = 5$.

r	$n-s$	Modified Winsorized Mean	Sample Mean	Actual Mean
20	10	2.703280	2.7202	2.84
20	10	2.790200	2.8139	2.84
20	10	2.855850	2.8830	2.84
20	10	2.871780	2.8923	2.84
20	10	3.027670	2.8900	2.84
20	10	2.866580	2.8772	2.84
20	10	2.757480	2.7799	2.84
20	10	2.846800	2.8594	2.84

(Table 1 continued)

r	$n-s$	Modified Winsorized Mean	Sample Mean	Actual Mean
20	10	2.828390	2.8537	2.84
20	10	2.950428	2.9581	2.84
20	10	2.826960	2.8467	2.84
20	10	2.730641	2.7554	2.84
20	10	2.935785	2.9503	2.84
20	10	2.842774	2.8571	2.84
20	10	2.709596	2.7241	2.84
20	10	2.829064	2.8320	2.84
20	10	2.851910	2.8588	2.84
20	10	2.827685	2.8317	2.84
20	10	2.915902	2.9227	2.84
20	10	2.732881	2.7526	2.84
20	10	2.813960	2.8332	2.84
20	10	2.941625	2.9334	2.84
20	10	2.772804	2.7751	2.84
20	10	2.792100	2.8227	2.84
20	10	2.683484	2.7264	2.84
20	10	2.906450	2.9094	2.84
20	10	2.844848	2.8687	2.84
20	10	2.856460	2.8616	2.84
20	10	2.821648	2.8573	2.84
30	20	2.814200	2.8713	2.84
30	20	2.708800	2.7884	2.84
30	20	2.706000	2.8095	2.84
30	20	2.713800	2.8051	2.84
30	20	2.720700	2.8155	2.84
30	20	2.759900	2.8452	2.84
30	20	2.752800	2.8231	2.84
30	20	2.816300	2.8432	2.84
30	20	2.907300	2.9321	2.84
30	20	2.774800	2.8107	2.84
30	20	2.738800	2.7947	2.84
30	20	2.949700	3.0005	2.84
30	20	2.615500	2.6991	2.84
30	20	2.727700	2.7901	2.84
30	20	2.731600	2.7876	2.84

Table 2: $\theta_1 = 1$ and $\theta_2 = 3$.

r	$n-s$	Modified Winsorized Mean	Sample Mean	Actual Mean
20	10	1.6981	1.6902	1.68
20	10	1.6240	1.6337	1.68
20	10	1.7654	1.7622	1.68
20	10	1.7197	1.7207	1.68
20	10	1.7610	1.7585	1.68
20	10	1.6845	1.6910	1.68
20	10	1.6981	1.6971	1.68
20	10	1.6193	1.6247	1.68
20	10	1.6738	1.6699	1.68
20	10	1.7454	1.7404	1.68
20	10	1.7120	1.7091	1.68
20	10	1.7031	1.6986	1.68
20	10	1.6276	1.6348	1.68
30	20	1.6819	1.7059	1.68
30	20	1.6333	1.6572	1.68
30	20	1.7032	1.7223	1.68
30	20	1.7294	1.7519	1.68
30	20	1.6532	1.6945	1.68
30	20	1.6330	1.6660	1.68
30	20	1.7327	1.7406	1.68
30	20	1.5786	1.6375	1.68
30	20	1.7035	1.7164	1.68
30	20	1.6096	1.6298	1.68
30	20	1.7186	1.7129	1.68
30	20	1.6170	1.6191	1.68
30	20	1.6415	1.6750	1.68
30	20	1.6509	1.6815	1.68
30	20	1.5933	1.6181	1.68
30	20	1.6785	1.6793	1.68
30	20	1.5779	1.6208	1.68
30	20	1.6812	1.6908	1.68
30	20	1.6545	1.6741	1.68
30	20	1.6424	1.6525	1.68

4 Discussion and Conclusion

In this paper, a modified winsorized mean is used to estimate the mean of a two-truncation parameter population. Under a very mild condition that r/n and $(n-s)/n$ are of order n^{-a} for $0 < a < 1$, it is shown that the modified winsorized mean, not only is a robust estimator, but also is a strongly consistent and asymptotically unbiased estimator for the mean of the two-truncation parameter population. Moreover, the standardized modified winsorized mean is found to be asymptotically normal when r/n and $(n-s)/n$ are of order n^{-a} for $a \geq 1/3$.

Acknowledgement: We thank the referees for their comments and suggestions that helped to improve the paper.

References

1. Bar-Lev, S.K. and Boukai, B. (1985). Minimum Variance Unbiased Estimation for the Families of Distributions Involving Two Truncations Parameters. *J. Stat. Plan. and Inf.*, 12, 379-384.

2. David, H.A. (1970). *Order Statistics*, Wiley& Sons, New York.

3. Davis, R.C. (1951). On minimum variance in nonregular estimation. *Ann. Math. Statist.*, 22, 43-57.

4. Ferentinos, K.K. (1990). Minimum Variance Unbiased Estimation in Doubly Type II Censored Samples from Families of Distributions Involving Two Truncation Parameters. *Commun. Statist. - Theory and Methods*, 19(3), 847 -856.

5. Hogg, R.V. and Craig, A.T. (1956). Sufficient Statistics in Elementary Distribution Theory. *Sankhyā*, 17, 209-216.

6. Nanthakumar, A. and Selvavel, K. (1994). Sequential Estimation in Two truncation Parameter Family of Distributions under Type II Censoring. *J of the Italian Statist. Soc.*, 3, 385-396.

7. Nanthakumar, A. and Selvavel, K. (1996). Sequential Estimation of Parameters in One-truncation Parameter Families Under Type II Censoring. *Sequential Analysis.* 15(4), 271-284.

8. Rohatgi, V.K. (1989). Unbiased Estimation of Parametric Function in Sampling from Two One-truncation Parameter Families. *Aust. J. of Statist.*, 31(2), 327-332.

9. Selvavel, K. (1989). Unbiased Estimation in Sampling from One-truncation Parameter Families when both Samples are Type II Censored. *Commun. Statist. - Theory and Methods*, 18(9), 3519 -3531.

10. Tukey, J.W. (1960). A Survey of Sampling from Contaminated Distributions. *Contributions to Probability and Statistics*, Olkin *et al* (Eds), Stanford University Press, 448-485.