HODGES-LEHMANN ESTIMATE OF THE LOCATION PARAMETER IN CENSORED SAMPLES*

A. K. MD. EHSANES SALEH

(Received Dec. 7, 1972; revised Apr. 30, 1976)

1. Introduction

Consider a random sample of size n from a distribution with absolutely continuous distribution function $F(x-\theta)$, where θ is unknown and F is symmetric about zero. Suppose the sample has been censored symmetrically under a type II scheme with a fixed proportion α of censoring from each end $(0<\alpha<1/2)$. For distributions with heavier tails than the normal distribution i.e. contaminated by gross-errors, Tukey and the Statistical Research group at Princeton (1949) proposed the α-Censored mean and α-Winsorized mean as estimates of θ, the location parameter. A survey paper on the properties of these estimates has been published by Tukey [17]. Hodges and Lehmann [10] proposed the *median* of the $\binom{n+1}{2}$ averages of pairs of observations as the estimate of θ. This estimate is based on the study of Wilcoxon's one-sample statistic. Huber [11] considers the class of maximum likelihood estimates of θ and found those members of this class which minimizes the maximum variance over various classes of contaminated distributions. Rothenberg et al. [13] show that the 0.375-censored best linear unbiased estimate of location parameter is most efficient relative to the maximum likelihood estimate for the Cauchy-family. It is also well-known that the maximum likelihood estimate of θ for the double exponential distribution is the "median" of the sample. Recently, Bickel [4] studied the asymptotic properties of the α-censored mean, α-Winsorized mean and Huber's estimate and he compared them with the Hodges-Lehmann estimate and remarked "unless the computation involved is prohibitive, the Hodges-Lehmann estimate is to be preferred in any situation where degree of contamination and type of distribution is not known with great precision." A general survey and advances on

* Research Supported by N.R.C. grant no. A3088.

AMS *Subject Classification*: Primary 62G05 Secondary 62G20, 62G35.
Key Words and Phrases: Censored Sample, Hodges and Lehmann estimate, Robustness, Wilcoxon Statistic, Efficiency.

robust estimate of location may be found in Andrews et al. [2].

In this paper, we consider the Wilcoxon's one-sample statistic in an α-censored sample and study its efficiency property as a test statistic for the hypothesis $\theta=0$. Next, we consider the estimation of θ based on this statistic and it is shown that the estimate of θ is the "median of $\binom{n-2k+1}{2}$ averages of pairs of observations in the uncensored portion of the sample," ($k=[n\alpha]$). This may be termed as the Hodges-Lehmann estimate in an α-censored sample. Finally, we compare the estimate with an α-censored mean, the α-Winsorized mean and the Huber's estimate and it turns out that the α-censored Hodges-Lehmann estimate reduces considerably the effect of contamination by gross-errors (see Section 5). Further, if the distribution is known, the efficiency of the estimate may be improved by properly choosing the proportion α of censoring.

Rank statistics in censored samples have already been discussed by Basu [3], Gastwirths [7], Hettmansperger [8], Sobel [15] and Tamura [16]. Hettmansperger [8] and Tamura [16] demonstrates that efficiency of test statistics may be improved by proper censoring of the sample.

In Section 2, we define the Wilcoxon's one-sample statistic in an α-censored sample and study its asymptotic properties. In Section 3, we propose an estimate of θ based on this statistic and study the asymptotic properties of the estimate. In Section 4, asymptotic relative efficiencies (ARE) of the estimate are given for various distributions. In Section 5, we consider the lower bound of the efficiency of the proposed estimate with respect to the families of all symmetric and symmetric unimodal distributions, the α-censored mean, the α-Winsorized mean and the principal estimate proposed by Huber [11] (proposal 2). This section reveals the reduction of the effect of contamination by the use of the proposed α-censored Hodges-Lehmann estimate.

2. The Wilcoxon's one-sample statistic in a α-censored sample

Let $X_1, X_2, \cdots, X_n$ be n independent random variables from the c.d.f. $F(x-\theta)$ where θ is unknown and F is **symmetric** about zero. Further, F is strictly increasing c.d.f. and absolutely continuous with respect to Lebesgue measure possessing the density $f(x-\theta)$ which is continuous and strictly positive on its convex support $C=\{x: 0<F(x)<1\}$. Let U_α denote the α-censored Wilcoxon's one-sample statistic.

$$U_\alpha(X)=\binom{n-2k+1}{2}^{-1} \sum_{k\leq i\leq j\leq n-k} \phi(X_{(i)}, X_{(j)}) \tag{2.1}$$

where $X=(X_{(k)}, X_{(k+1)}, \cdots, X_{(n-k)})$ and $k=[n\alpha]$ denoting the largest integer contained in [ ]. Further,

$$\phi(X_{(i)}, X_{(j)})=\begin{cases} 1 & \text{if } X_{(i)}+X_{(j)}>0 \\ 0 & \text{otherwise}. \end{cases} \tag{2.2}$$

In this section we first study the asymptotic property of U_α defined above and propose an estimate of θ based on U_α in Section 3.

Let $X_{(1)}\leqq X_{(2)}\leqq\cdots\leqq X_{(n)}$ be the order statistics of the random variables $X_1, X_2,\cdots, X_n$. Assume $\alpha\in(0, 1/2)$ and the density f continuously differentiable in the neighbourhood of the population quantiles ξ_α and $\xi_{1-\alpha}$ of orders α and $1-\alpha$ respectively. Let X_α and $X_{1-\alpha}$ be the sample quantiles corresponding to ξ_α and $\xi_{1-\alpha}$. Then, conditionally on the vector statistic

$$Z=\{\sqrt{n}(X_\alpha-\xi_\alpha), \sqrt{n}(X_{1-\alpha}-\xi_{1-\alpha})\}, \tag{2.3}$$

U_α is Wilcoxon's one-sample statistic based on a sample of size $n-2k$, $(k=[n\alpha])$ from the distributions C_α with density

$$g_\alpha(x\mid Z)=\begin{cases} \dfrac{f(x-\theta)}{F(X_{1-\alpha}-\theta)-F(X_\alpha-\theta)}; & X_\alpha\leqq x\leqq X_{1-\alpha}, \\ 0 & \text{otherwise}. \end{cases} \tag{2.4}$$

Now, define the statistic

$$R_\alpha=\binom{n-2k+1}{2}^{-1}\sum_{k\leqq i\leqq j\leqq n-k}\{\phi(X_{(i)}, X_{(j)})-\mathrm{E}(\phi\mid Z)\}, \tag{2.5}$$

where $\mathrm{E}(\phi\mid Z)$ denotes the conditional expectation of ϕ in (2.2) given the vector Z in (2.3). Lehmann [12] has shown that the conditional distribution of U_α given Z is asymptotically normal for all θ and Cramer [6] shows that the density of Z converges point-wise to the bivariate normal density. Further, it follows from Theorem 2 of Sethuraman [14] that the random vector

$$\{\sqrt{n}R_\alpha, \sqrt{n}(X_\alpha-\xi_\alpha), \sqrt{n}(X_{1-\alpha}-\xi_{1-\alpha})\} \tag{2.6}$$

is asymptotically distributed as a trivariate normal distribution for all θ with the elements of the covariance matrix as follows:

$$\sigma_{11}=\frac{1}{3(1-2\alpha)}, \quad \sigma_{22}=\frac{\alpha(1-\alpha)}{f^2(\xi_\alpha)}, \quad \sigma_{33}=\frac{\alpha(1-\alpha)}{f^2(\xi_{1-\alpha})}, \quad \sigma_{23}=\frac{\alpha^2}{f^2(\xi_\alpha)}, \tag{2.7}$$
$$\sigma_{1j}=0, \qquad j=2, 3.$$

Now,

$$\sqrt{n}\{U_\alpha-\mathrm{E}(U_\alpha)\}=\sqrt{n}R_\alpha+\sqrt{n}\{\mathrm{E}_\theta(\phi\mid Z)-\mathrm{E}(U_\alpha)\} \tag{2.8}$$

where

$$\text{(2.9)} \qquad \begin{aligned} E_\theta(\phi \mid Z) &= P(X_{(i)} + X_{(j)} > 0) \\ &= \int_{X_\alpha}^{X_{1-\alpha}} \int_{-y}^{X_{1-\alpha}} \frac{f(x-\theta)f(y-\theta)dxdy}{[F(X_{1-\alpha}-\theta)-F(X_\alpha-\theta)]^2} \, . \end{aligned}$$

An application of Theorem 4.2.5 of Anderson [1] shows that $\sqrt{n}(U_\alpha - E(U_\alpha))$ is asymptotically normally distributed with zero mean and variance

$$\text{(2.10)} \qquad \frac{1+4\alpha}{3(1-2\alpha)^2} \, .$$

The above conclusion follows from the following considerations: consider the asymptotic variance of U_α under the hypothesis $\theta=0$. In this case,

$$\text{(2.11a)} \qquad E_0(U_\alpha) = 1/2$$

and

$$\text{(2.11b)} \quad \begin{aligned} \operatorname{Var}\left\{n^{1/2}\left(U_\alpha - \frac{1}{2}\right)\right\} &= \operatorname{Var}(n^{1/2}R_\alpha) + \operatorname{Var}\left\{n^{1/2}\left(E_0(\phi \mid Z) - \frac{1}{2}\right)\right\} \\ &\quad + 2\operatorname{Cov}\{n^{1/2}R_\alpha, n^{1/2}(E_0(\phi \mid Z))\} \end{aligned}$$

Now,

$$\text{(2.12)} \qquad E_0(\phi \mid Z) = \frac{3F(X_{1-\alpha}) + F(X_\alpha) - 2}{2[F(X_{1-\alpha}) - F(X_\alpha)]} \, .$$

By Taylor's expansion of (2.12) about ξ_α and $\xi_{1-\alpha}$ we have

$$\text{(2.13)} \quad E_0(\phi \mid Z) = \frac{1}{2} + (X_\alpha - \xi_\alpha)\frac{f(\xi_\alpha)}{(1-2\alpha)} + (X_{1-\alpha} - \xi_{1-\alpha})\frac{f(\xi_{1-\alpha})}{(1-2\alpha)} + O(n^{-1}) \, .$$

This results in

$$\text{(2.14)} \qquad \begin{aligned} \operatorname{Var}\left\{n^{1/2}\left(U_\alpha - \frac{1}{2}\right)\right\} &= \frac{1}{3(1-2\alpha)} + \frac{2\alpha}{(1-2\alpha)^2} + O(n^{-1}) \\ &= \frac{1+4\alpha}{3(1-2\alpha)^2} + O(n^{-1}) \, . \end{aligned}$$

Hence,

$$\text{(2.15)} \qquad \lim_{n\to\infty} \operatorname{Var}\left\{n^{1/2}\left(U_\alpha - \frac{1}{2}\right)\right\} = \frac{1+4\alpha}{3(1-2\alpha)^2} \, .$$

Also, for any θ, the asymptotic mean of U_α is given by

$$\text{(2.16)} \qquad 12(1-2\alpha)^{-2} \int_{\xi_\alpha}^{\xi_{1-\alpha}} \int_{-y}^{\xi_{1-\alpha}} f(x-\theta)f(y-\theta)dxdy$$

and the derivative of (2.16) evaluated at $\theta=0$ is

$$12(1-2\alpha)^{-2}\int_{\xi_\alpha}^{\xi_{1-\alpha}} f^2(x)dx\ . \tag{2.17}$$

Note that for $\alpha=0$, the corresponding result coincides with the uncensored Wilcoxon's statistic. Therefore, it follows that

THEOREM 2.1. *For the class C_α of distributions $0<\alpha<1/2$, the asymptotic efficiency of U_α relative to U_0 is*

$$\mathrm{ARE}\,(U_\alpha : U_0)=\frac{\left(\int_{\xi_\alpha}^{\xi_{1-\alpha}} f^2(x)dx\right)^2}{(1+4\alpha)(1-2\alpha)^2\left(\int_{-\infty}^{\infty} f^2(x)dx\right)^2}\ . \tag{2.18}$$

Moreover, for C_α with unimodal densities

$$\frac{1}{1+4\alpha}\leqq \mathrm{ARE}\,(U_\alpha : U_0)\leqq \frac{1}{(1+4\alpha)(1-2\alpha)^2}\ . \tag{2.19}$$

PROOF. First part follows from the definition of asymptotic relative efficiency (ARE) and the result (2.19) follows from Hettmansperger [8]. The lower bound is attained by the rectangular distribution. Bickel [4] has shown that this lower bound is the same as the efficiency of α-censored mean relative to the mean.

3. Estimation of location parameter

From Section 2, we consider two immediate properties of U_α conditional on the vector random variable Z:

(a) U_α is symmetrically distributed about 1/2 independent of F when $\theta=0$,

(b) $U_\alpha(X_{(k)}+a, X_{(k+1)}+a,\cdots, X_{(n-k)}+a)$ is non-decreasing in a for all $(X_{(k)},\cdots, X_{(n-k)})$.

The possible values of $\binom{n-2k+1}{2}U_\alpha$ are $0, 1, 2, \cdots, M$ where $M=\binom{n-2k+1}{2}$. M may be odd or even. Let $M=2l+1$, then we have

$$\begin{aligned}\theta_\alpha^{**}&=\sup_\theta\,\{\theta : U_\alpha(X-\theta)>1/2\}=W_{(l+1)}\\ \theta_\alpha^{*}&=\inf_\theta\,\{\theta : U_\alpha(X-\theta)<1/2\}=W_{(l+1)}\end{aligned} \tag{3.1}$$

where $W_{(1)}<W_{(2)}<\cdots<W_{(M)}$ are the ordered averages $(X_{(i)}+X_{(j)})/2=W_{ij}$, $k\leqq i\leqq j\leqq n-k$. Similarly, for $M=2l$,

$$(3.2)\qquad \begin{aligned} \theta_\alpha^{**} &= \sup_\theta \{\theta : U_\alpha(X-\theta) > 1/2\} = W_{(l+1)} \\ \theta_\alpha^{*} &= \inf_\theta \{\theta : U_\alpha(X-\theta) > 1/2\} = W_{(l)} \,. \end{aligned}$$

In both cases, we have the α-censored estimate of θ as

$$(3.3)\qquad \hat{\theta}_\alpha = \operatorname*{med}_{k \leq i \leq j \leq n-k} \frac{\{X_{(i)} + X_{(j)}\}}{2} \,.$$

We now consider the asymptotic properties of $\hat{\theta}_\alpha$. Following Lehmann [10], it may be proved that $\hat{\theta}_\alpha$ has an absolutely continuous distribution conditionally on Z. Further, under the same condition $\hat{\theta}_\alpha$ is translation invariant i.e.

$$(3.4)\qquad \hat{\theta}_\alpha(X+a) = \hat{\theta}_\alpha(X) + a \,.$$

Thus, we have

$$(3.5)\qquad \mathrm{P}_\theta \{(\hat{\theta}_\alpha - \theta) \leqq u \,|\, Z\} = \mathrm{P}_0 (\hat{\theta}_\alpha \leqq u \,|\, Z)$$

where P_θ denotes that the probability has been computed assuming θ to be true value of the parameter. The distribution of $\hat{\theta}_\alpha$ is conditionally symmetric about θ by Theorem 3 of Hodges and Lehmann [10]. Therefore, $\hat{\theta}_\alpha$ is an unbiased estimate, (conditionally on Z) follows from the fact that $n^{1/2}\{U_\alpha(X-\theta-a/\sqrt{n}) - 1/2\}$ is asymptotically normal with zero mean and variance $1/3(1-2\alpha)$ for all θ using the theory of U-statistics due to Hoeffding [5] and Lehmann [12]. Thus, by Hodges and Lehmann [10],

$$(3.6)\qquad \begin{aligned} G(a) &= \lim_{n\to\infty} \mathrm{P}_0 \{n^{1/2}(\hat{\theta}_\alpha - \theta) \leqq a\} \\ &= \lim_{n\to\infty} \mathrm{P}_\theta [n^{1/2}\{U_\alpha(X-\theta-a/\sqrt{n}) - \mu_n\}/\sigma \leqq n^{1/2}(1/2 - \mu_n)/\sigma] \end{aligned}$$

where

$$(3.7a)\qquad \mu_n = \frac{1}{2} - \frac{2a}{n^{1/2}(1-2\alpha)^2} \int_{\xi_\alpha}^{\xi_{1-\alpha}} f^2(x)dx + O(n^{-1})$$

$$(3.7b)\qquad \sigma^2 = \frac{1+4\alpha}{3(1-2\alpha)^2}$$

and as $n \to \infty$

$$n^{1/2}\left(\frac{1}{2} - \mu_n\right) \to \frac{2a}{(1-2\alpha)^2} \int_{\xi_\alpha}^{\xi_{1-\alpha}} f^2(x)dx \,.$$

Thus, as $n \to \infty$

$$G(a)=\Phi(aB/A) \tag{3.8}$$

where

$$B^2=4(1-2\alpha)^{-4}K^2 \qquad \text{and} \qquad A^2=\frac{1+4\alpha}{3(1-2\alpha)^2}$$

$$K=\int_{\xi_\alpha}^{\xi_{1-\alpha}} f^2(x)dx$$

and Φ is the standard normal c.d.f. Therefore, it follows that $n^{1/2}(\hat{\theta}_\alpha-\theta)$ is asymptotically normal with zero mean and variance

$$\frac{(1+4\alpha)(1-2\alpha)^2}{12\left(\int_{\xi_\alpha}^{\xi_{1-\alpha}} f^2(x)dx\right)^2}\,. \tag{3.9}$$

The asymptotic variance of the *uncensored* Hodges and Lehmann estimate of θ is given by

$$1/12\left(\int_{-\infty}^{\infty} f^2(x)dx\right)^2.$$

Therefore, the asymptotic relative efficiency (ARE) of $\hat{\theta}_\alpha$ may be obtained as

$$\text{ARE}\,(\hat{\theta}_\alpha:\hat{\theta}_\alpha)=\frac{\left(\int_{\xi_\alpha}^{\xi_{1-\alpha}} f^2(x)dx\right)^2}{(1+4\alpha)(1-2\alpha)^2\left(\int_{-\infty}^{\infty} f^2(x)dx\right)^2}\,, \tag{3.10}$$

which is the same as the result given in Theorem 2.1.

The following table of efficiency of the Hodges and Lehmann estimate of θ in censored sample relative to uncensored sample gives a good idea of the performance of the estimate we have proposed.

Table 1

α	.10	.20	.25	.30	.35	.40	.45
Logistic	.995	.968	.945	.917	.882	.842	.798
Double-exponential	1.029	1.089	1.125	1.164	1.204	1.246	1.289
Cauchy	1.089	1.258	1.339	1.403	1.435	1.441	1.405

These tabulers values have been quoted from Tamura [16]. For the normal case the efficiency is near optimal for $\alpha=.05$. Higher values of α reduces the efficiency for other distributions listed.

However, it is useful to compare our estimate to the optimal linear α-censored estimate as presented in Chernoff et al. [5]. The asymptotic variance of the optimal linear α-censored estimate of the location parameter is given by

(3.11) $$\sigma^2_{\theta_\alpha^{\rm opt}} = \left\{ \int_{\xi_\alpha}^{\xi_{1-\alpha}} x^2 f(x)dx + \frac{2f^2(\xi_\alpha)}{\alpha} \right\} .$$

Therefore, the ARE $(\hat{\theta}_\alpha : \theta_\alpha^{\rm opt})$ is given by

$$\text{ARE}\,(\hat{\theta}_\alpha : \theta_\alpha^{\rm opt}) = \frac{12\left(\int_{\xi_\alpha}^{\xi_{1-\alpha}} f^2(x)dx\right)^2}{(1+4\alpha)(1-2\alpha)^2\left\{\int_{\xi_\alpha}^{\xi_{1-\alpha}} x^2 f(x)dx + \frac{2f^2(\xi_\alpha)}{\alpha}\right\}} .$$

4. ARE examples

If $\phi(x) = (2\pi)^{-1/2} e^{-x^2/2}$, $-\infty \leqq x \leqq \infty$, the standard normal density and $\Phi(x)$ the corresponding c.d.f., then the ARE is given by

(4.1) $$\frac{3}{\pi} \frac{[2\Phi(2^{1/2}\xi_{1-\alpha})-1]^2}{(1+4\alpha)(1-2\alpha)^2\left[1-2\alpha+2\Phi^{-1}(\alpha)\phi(\Phi^{-1}(\alpha))+2\frac{\phi^2(\Phi^{-1}(\alpha))}{\alpha}\right]}$$

for $0<\alpha<1/2$. In this case, ARE $(\hat{\theta}_\alpha : \theta_\alpha^{\rm opt}) \leqq .955$, for all $0<\alpha<1/2$ and decreases to 0, as $\alpha \to 1/2$. If $f(x) = (1/2)e^{-|x|}$, $-\infty \leqq x \leqq \infty$ (the double exponential density), then

(4.2) $$\text{ARE}\,(\hat{\theta}_\alpha : \theta_\alpha^{\rm opt}) = \frac{3(1+2\alpha)^2}{8(1+4\alpha)\{1-\alpha(1+\ln 2\alpha)^2\}} .$$

(4.2) is always $\leqq 1$ for all $0<\alpha<1/2$ and increasing in α and

$$\lim_{\alpha\to 1/2} \text{ARE}\,(\hat{\theta}_\alpha : \theta_\alpha^{\rm opt}) = 1 .$$

If $f(x) = \pi^{-1}(1+x^2)^{-1}$, $-\infty \leqq x \leqq \infty$. The ARE becomes

(4.3)
$$\text{ARE}\,(\hat{\theta}_\alpha : \theta_\alpha^{\rm opt}) = \left(\frac{3}{\pi}\right) \frac{\{(1/\pi)\sin \pi(1-2\alpha)+(1-2\alpha)\}^2}{2(1+4\alpha)(1-2\alpha)^2\left\{\tan\frac{\pi}{2}(1-2\alpha)+\frac{\cos^4 \pi(1-2\alpha)/2}{\pi_\alpha}-\frac{\pi}{2}(1-2\alpha)\right\}}$$

for all $0<\alpha<1/2$. This ARE is a convex function of α and has a maximum at $\alpha=.375$ further as $\alpha\to 0$ or $\alpha \to 1/2$, ARE $(\hat{\theta}_\alpha : \theta_\alpha^{\rm opt}) \to 0$. Thus, the middle 25 percent observations gives most efficient α-censored Hodges and Lehmann estimate of the location parameter θ.

5. Comparison with α-censored mean, α-Winsorized mean and Huber's estimate

Let $\mathcal{F}$ be the family of all symmetric distributions possessing the regularity conditions of Section 2 and let $\mathcal{G}$ be the family of all symmetric unimodal distributions which possess the same conditions as outlined in Section 2. In this section, we compare the α-censored Hodges and Lehmann estimate with α-censored mean, α-Winsorized mean and Huber-estimates.

5a. *Comparison with α-censored mean*

Consider the α-censored mean defined by

$$\bar{X}_\alpha = \frac{1}{n-2[n\alpha]} \sum_{i=k}^{n-k} X_{(i)} , \qquad k=[n\alpha] . \tag{5.1}$$

It has been shown by Huber [11] that it is the maximum likelihood estimate of θ when the density is given by

$$f(x) = \begin{cases} \dfrac{1-\varepsilon}{\sqrt{2\pi}} \exp\left(-\dfrac{x^2}{2}\right) & \text{for } |x|<a \\ \dfrac{1-\varepsilon}{\sqrt{2\pi}} \exp\left\{-a|x|-\dfrac{1}{2}a^2\right\} & \text{for } |x|\geqq a \end{cases} \tag{5.2}$$

which $\alpha=F(-a)$ and F is the c.d.f. corresponding to $f(x)$ and $0<\varepsilon<1$ while a depends on ε. Bickel [4] has shown that $n^{1/2}(\bar{X}_\alpha-\theta)$ is asymptotically normal with zero mean and variance

$$(1-2\alpha)^{-2}\left\{\int_{\xi_\alpha}^{\xi_{1-\alpha}} x^2 f(x)dx + 2\alpha\xi_\alpha^2\right\} \tag{5.3}$$

for *any* symmetric $f(x)$. We may compare our estimate with (5.1) and obtain the ARE as

$$\text{ARE}\,(\hat{\theta}_\alpha : \bar{X}_\alpha) = \frac{12}{(1+4\alpha)(1-2\alpha)^4}\left(\int_{\xi_\alpha}^{\xi_{1-\alpha}} f^2(x)dx\right)^2 \cdot \left\{\int_{\xi_\alpha}^{\xi_{1-\alpha}} x^2 f(x)dx + 2\alpha\xi_\alpha^2\right\} . \tag{5.4}$$

(i) If $f(x)$ is normal we obtain

$$\frac{3}{\pi}\,\frac{[2\Phi(2^{1/2}\Phi^{-1}(\alpha))-1]^2}{(1+4\alpha)(1-2\alpha)^2}\{1-2\alpha+2\Phi^{-1}(\alpha)\phi(\Phi^{-1}(\alpha))+2\alpha[\Phi^{-1}(\alpha)]^2\} . \tag{5.5}$$

(ii) If $f(x)$ is double-exponential distribution we have

$$\frac{3}{2}\,\frac{(1+2\alpha)^2}{(1+4\alpha)(1-2\alpha)^2}[1-2\alpha(1-\ln 2\alpha)] . \tag{5.6}$$

(iii) If $f(x)$ is Cauchy distribution, we have

$$\frac{6\{(1/\pi)\sin\pi(1-2\alpha)+(1-2\alpha)\}^2}{\pi^3(1+4\alpha)(1-2\alpha)^4}\left\{\pi\alpha\tan^2\frac{\pi}{2}(1-2\alpha)\right. \tag{5.7}$$

$$+\tan\frac{\pi}{2}(1-2\alpha)-\frac{\pi}{2}(1-2\alpha)\Big\} .$$

Now, we present the following theorem which gives the lower bound of the efficiency defined at (5.4).

THEOREM 5.1. *Let* ARE $(\hat{\theta}_\alpha : \bar{X}_\alpha)$ *be the efficiency of the Hodges and Lehmann estimate* $\hat{\theta}_\alpha$ *and the mean* $\bar{X}_\alpha$ *in an* α*-censored sample. Then*

$$\inf_{F\in\mathcal{F}} \text{ARE}\,(\hat{\theta}_\alpha : \bar{X}_\alpha)=\frac{1}{(1+4\alpha)(1-2\alpha)^2}\frac{27}{2000}\{(1-2\alpha)c(\alpha)+10\alpha\} \cdot \{3c^2(\alpha)-10c(\alpha)+15\}^2 \tag{5.8}$$

where

$$c(\alpha)=1+\frac{4}{3}[\{3(\alpha+\alpha^2)\}^{1/2}-3\alpha](1-2\alpha)^{-1} \qquad \textit{and}$$

$$\sup_{F\in\mathcal{F}} \text{ARE}\,(\hat{\theta}_\alpha : \bar{X}_\alpha)=\infty .$$

PROOF. Follows from Bickel [4] Theorem 4.2.

The following tabular values are given for the lower bound which are quite high and we conclude that the Hodges and Lehmann estimate in α-censored sample is to be preferred unless very precise knowledge of the required α is available. These values are higher than the Bickel's [4] result.

Table 2

α	.01	.02	.03	.04	.05	.06	.07	.08	.10
$\inf_{F\in\mathcal{F}}$ ARE $(\hat{\theta}_\alpha : \bar{X}_\alpha)$	.8911	.9042	.9195	.9268	.9362	.9477	.9507	.9556	.9654

5b. *Comparison with* α*-Winsorized mean*

Let $\bar{X}_\alpha^*$ denote the α-Winsorized mean defined by

$$\bar{X}_\alpha^*=n^{-1}\Big\{kX_{(k)}+\sum_{i=k+1}^{n-k}X_{(i)}+kX_{(n-k+1)}\Big\} . \tag{5.9}$$

It has been shown by Bickel [4] that $n^{1/2}(\bar{X}_\alpha^*-\theta)$ is asymptotically normal with zero mean and variance

$$\int_{\xi_\alpha}^{\xi_{1-\alpha}} t^2dF(t)+2\alpha\{\xi_{1-\alpha}+\alpha/f(\xi_\alpha)\}^2 \tag{5.10}$$

for any symmetric or symmetric unimodal densities $f(x)$.

We may compare our estimate with $\bar{X}_\alpha^*$ and obtain the ARE as

$$(5.11)\qquad \text{ARE}\,(\hat{\theta}_\alpha : \bar{X}_\alpha) = \frac{12}{(1+4\alpha)(1-2\alpha)^2}\left(\int_{\xi_\alpha}^{\xi_{1-\alpha}} f^2(x)dx\right)^2 \cdot \left\{\int_{\xi_\alpha}^{\xi_{1-\alpha}} t^2 dF(t) + 2\alpha(\xi_{1-\alpha} + \alpha/f(\xi_\alpha))^2\right\} .$$

The following theorem gives the lower bound of ARE $(\hat{\theta}_\alpha : \bar{X}_\alpha^*)$ for $F \in \mathcal{F}$.

THEOREM 5.2.

$$(5.12)\qquad \inf_{F\in\mathcal{F}} \text{ARE}\,(\hat{\theta}_\alpha : \bar{X}_\alpha^*) = \frac{1}{1+4\alpha}\,\frac{27}{2000}\{(1-2\alpha)c(\alpha)+10\alpha\} \cdot \{3c^2(\alpha)-10c(\alpha)+15\}^2$$

where $c(\alpha)$ *is defined in Theorem* 5.1.

PROOF. Follows from Bickel [4] using the expression (4.24) on page 854 and Theorem 5.1. For $F \in \mathcal{G}$, where $\mathcal{G}$ is the family of symmetric unimodal distributions, the approximate lower bound may be obtained following Bickel [4] as

$$(5.13)\qquad \inf_{F\in\mathcal{G}} \text{ARE}\,(\hat{\theta}_\alpha : \bar{X}_\alpha^*) \geqq \inf_{c\geqq 1} (1+4\alpha)^{-1}\frac{27}{2000}[(1-2\alpha)^3 c + 10\alpha] \cdot (3c^2-10c+15)^2 .$$

Table 3 gives the values of lower bounds for chosen values α as in Bickel [4] to show the improvement by censoring when $F \in \mathcal{F}$ and $F \in \mathcal{G}$.

Table 3

α	.01	.02	.03	.04	.05	.06	.07	.08	.10
$\inf_{F\in\mathcal{F}}$ ARE $(\hat{\theta}_\alpha : \bar{X}_\alpha^*)$	.8558	.8333	.8125	.7845	.7583	.7339	.7031	.6742	.6179
Lower bound of ARE $(\hat{\theta}_\alpha : \bar{X}_\alpha^*)$ for $F \in \mathcal{G}$	.8510	.8439	.8286	.8148	.8127	.8019	.7922	.7837	.7813

5c.

Let $H(a)$ be the estimate of θ considered by Huber [11] in proposal 2 which is obtained as a unique solution of the system of equations

$$\sum_{i=1}^n \psi(a, (x_i - T)/s) = 0 , \qquad \sum \psi^2(a, (x_i - T)/s) = 0 .$$

Bickel [4] has shown that $\sqrt{n}(H(a)-\theta)$ is asymptotically normal with zero mean and variance

$$(5.14)\qquad \left(\int_{-q}^{q} dF(t)\right)^{-2}\left(\int_{-q}^{q} t^2 dF(t) + 2q^2\int_q^\infty dF(t)\right)$$

and q satisfies the condition

$$q^2\beta(a)/a^2=\int_{-q}^{q} t^2 dF(t)+2q^2\int_q^\infty dF(t) \tag{5.15}$$

and

$$\beta(a)=\int_{-a}^{a} t^2 d\Phi(t)+2a^2\int_a^\infty d\Phi(t) \tag{5.16}$$

where Φ is the standard normal distribution. The efficiency of $\hat\theta_\alpha$ relative to $H(a)$ is then given by

$$\text{ARE}\,(\hat\theta_\alpha : H(a))=\frac{12}{(1+4\alpha)(1-2\alpha)^2}\left(\int_{\xi_\alpha}^{\xi_{1-\alpha}} f^2(x)dx\right)^2\left(\int_{-q}^{q} dF(t)\right)^{-2} \cdot\left(\int_{-q}^{q} t^2dF(t)+2q^2\int_q^\infty dF(t)\right). \tag{5.17}$$

The following theorem gives the lower bound of the efficiency of the ARE $(\hat\theta_\alpha : H(a))$. Let $\tau=2a^2/\beta(a)$. Then τ is the monotonically increasing function of a. The lower bounds are given in terms of τ.

THEOREM 5.3.

(a) $\sup_{F\in\mathcal{F}} \text{ARE}\,(\hat\theta_\alpha : H(a))=\infty$

(b) $$\inf_{F\in\mathcal{F}} \text{ARE}\,(\hat\theta_\alpha : H(a))=\frac{1}{(1+4\alpha)(1-2\alpha)^4}\,\frac{27}{8} \cdot \tau^{-5}\{32\alpha_\tau^2\tau^2+(8\tau^2-80\tau)\alpha_\tau+(3\tau^2-20\tau-60)\}^2 \quad \textit{for } 2\leqq\tau\leqq 7.415$$

$\alpha_\tau=1/2-.685(1-2/\tau)$,

(c) $$\inf_{F\in\mathcal{F}} \text{ARE}\,(\hat\theta_\alpha : H(a))=\frac{1}{(1+4\alpha)(1-2\alpha)^2}\,\frac{27}{8}\{\tau^{-5}(3\tau^2-20\tau+60)^2\} \quad \textit{for } 7.415\leqq\tau\leqq 10$$

$$=\frac{.864}{(1+4\alpha)(1-2\alpha)^2} \quad \textit{for } \tau\geqq 10\,.$$

PROOF. Follows from Bickel's Theorems 5.1 and 5.2 [4].

We now present Table 4 which gives the ARE $(\hat\theta_\alpha : \bar X_\alpha)$ and ARE $(\hat\theta_\alpha : \bar X_\alpha^*)$ for normal, rectangular, Double-exponential and Cauchy distributions. This table has been computed using $\alpha=.01$ and $\alpha=.05$.

Table 4

	Normal		Rectangular		Double-exponential		Cauchy	
ARE $(\hat\theta_\alpha : \bar X_\alpha)$	.9611	1.0134	1.0412	1.2346	1.4117	1.2757	6.7280	2.7469
ARE $(\hat\theta_\alpha : \bar X_\alpha^*)$	.9580	.9981	.9712	.8583	1.4805	1.4369	8.3597	6.5777

The first column under distribution is for $\alpha=.01$ and the second is for $\alpha=.05$.

The conclusion of this investigation is that the α-censored Hodges and Lehmann estimate is robust and reduces the effect of contamination of distributions.

Acknowledgement

The author wishes to thank the referee for his valuable comments which have improved the paper to the present form. This research is a part of the project under National Research Council Grant A-3088.

CARLETON UNIVERSITY

REFERENCES

[1] Anderson, T. W. (1958). *Introduction to Multivariate Analysis*, Wiley, New York.

[2] Andrews, D. F., Bickel, P. J., Hampel, F. R., Huber, P. J., Rogers, W. H. and Tukey, J. W. (1972). *Robust Estimates of Location*, Princeton University Press.

[3] Basu, A. P. (1967). On the large sample properties of generalised, Wilcoxon-Mann-Whitney Statistic, *Ann. Math. Statist.*, **38**, 905-915.

[4] Bickel, P. J. (1965). On some robust estimates of location, *Ann. Math. Statist.*, **36**, 847-858.

[5] Chernoff, H., Gastwirth, J. L. and Jones, M. V. (1967). Asymptotic distributions of linear combinations of functions of order statistics with application to estimation, *Ann. Math. Statist.*, **38**, 52-74.

[6] Cramer, H. (1946). *Mathematical Methods of Statistics*, Princeton University Press.

[7] Gastwirth, J. L. (1965). Asymptotically most powerful rank tests for the two sample problem with censored data, *Ann. Math. Statist.*, **36**, 1243-1247.

[8] Hettmansperger, T. P. (1968). On trimmed Mann-Whitney test, *Ann. Math. Statist.*, **39**, 1610-1614.

[9] Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution, *Ann. Math. Statist.*, **19**, 293-325.

[10] Hodges, J. L. and Lehmann, E. L. (1963). Estimates of location parameter based on rank tests, *Ann. Math. Statist.*, **34**, 598-611.

[11] Huber, P. J. (1964). Robust estimation of location parameter, *Ann. Math. Statist.*, **35**, 73-102.

[12] Lehmann, E. L. (1951). Consistency and unbiasedness of certain non-parametric tests, *Ann. Math. Statist.*, **22**, 165-179.

[13] Rothenberg, T., Fisher, F. and Tilanus, C. (1964). A note on estimates from a Cauchy sample, *J. Amer. Statist. Ass.*, **59**, 460-463.

[14] Sethuraman, J. (1961). Some limit theorems for joint distributions, *Sankhya*, 10, 383-386.

[15] Sobel, M. (1965). On a generalisation of Wilcoxon's rank sum test for censored samples, *Technical Report*, No. 69, University of Minnesota.

[16] Tamura, T. (1971). On a *c*-sample test based on samples, *Ann. Math. Statist.*, **42**, 1455-1460.

[17] Tukey, J. (1960). A Survey of Sampling from Contaminated Distributions, *Contribution to Probability and Statistics*, Stanford University Press, 448-486.