HALF-SAMPLE ESTIMATION OF COVERAGE PROBABILITY[1]

by

Jun Shao

Purdue University
Technical Report #88-4

Department of Statistics
Purdue University

February 1988

[1] This research was supported in part by the NSF-AFOSR grant ISSA-860068.

o

HALF-SAMPLE ESTIMATION OF COVERAGE PROBABILITY [1]

by

Jun Shao

Purdue University

Abstract

A half-sample estimator is proposed for estimating the coverage probability of a fixed-width interval $|\hat{\theta}-\theta|\leq\delta$, where θ is a parameter of interest and $\hat{\theta}$ is its point estimator. Theoretical finite sample properties of the half-sample estimator are obtained. Some extensions and the robustness of the method are also studied. Numerical results show that the half-sample estimator works very well in general situations.

KEY WORDS: Half-sampling; Accuracy measure; Unbiasedness; Conservative estimators; Incomplete U-statistics.

[1] This research was supported in part by the NSF-AFOSR grant ISSA-860068.

1. INTRODUCTION

When presenting a point estimate in statistical applications, it is usually necessary to indicate the accuracy of the estimate. Suppose that the parameter of interest is θ, a characteristic of an unknown population distribution F, and an estimate of θ based on n samples is $\hat{\theta}$. Then any measure of the accuracy of $\hat{\theta}$ depends on the sampling distribution of $\hat{\theta}$ and has to be estimated from the samples. For example, a commonly used measure of the accuracy of $\hat{\theta}$ is the mean squared error (MSE) of $\hat{\theta}$, which is the second order moment of $\hat{\theta}-\theta$ under the sampling distribution of $\hat{\theta}$, i.e.,

$$MSE = MSE(\hat{\theta}) = E(\hat{\theta}-\theta)^2.$$

Another measure of the accuracy of $\hat{\theta}$, which is suggested and discussed by Kiefer (1977), is the coverage probability of a fixed-width interval:

$$p(\delta) = p(\delta,\hat{\theta}) = P(\ |\hat{\theta}-\theta| \le \delta\ ), \tag{1.1}$$

where $\delta>0$ is a fixed constant. Both MSE and $p(\delta)$ are characteristics of the sampling distribution of $\hat{\theta}$. MSE and $p(\delta)$ are related, i.e., $MSE = \int_0^\infty [1-p(\delta^{1/2})]d\delta$. An advantage of using $p(\delta)$ as an accuracy measure is that $p(\delta)$ is always finite whereas the finiteness of the MSE of $\hat{\theta}$ requires some moment conditions on F.

When the sample size n is large, a standard technique is to apply asymptotic theory. That is, after obtaining the limit of the sampling distribution (assume it exists), we approximate the characteristics of the sampling distribution by the corresponding characteristics of the limiting distribution. Even if the MSE of $\hat{\theta}$ does not exist, we can still use the variance of the limiting distribution as a measure of the accuracy of $\hat{\theta}$. Setting confidence interval for θ, a closely related problem, can also be done by using the limiting sampling distribution.

The limitations of this approach are: (1) obviously, n needs to be large; (2) the existence of the limit of the sampling distribution requires some condition on the tails of the population distribution F (e.g., the second moment of F exists if $\hat{\theta}$ is the sample mean).

Resampling methods such as the jackknife (Quenouille, 1956; Tukey, 1958) and the bootstrap (Efron, 1979) are convenient nonparametric methods for estimating MSE or $p(\delta)$. They can be used in *both* large sample and small sample situations. Large sample (asymptotic) properties of the resampling methods have attracted a lot of attention in the statistical literature in recent years. In particular, the resampling methods are proved to be superior to the classical asymptotic methods for their accuracy (Abramovitch and Singh, 1985; Hall, 1986; Beran, 1987) and robustness (Hinkley, 1977; Shao and Wu, 1987). However, much less attention has been focused on finite sample (exact) properties of the resampling methods. The study of the finite sample properties of the resampling methods is perhaps more important since few methods are available for estimating the statistical accuracy when the sample size is not large. Except for the bias of the jackknife variance estimator (Efron and Stein, 1981; Bhargava, 1983; Rao and Wu, 1985; Wu, 1986), not much has been known yet about the finite sample properties of the resampling methods.

In this paper, we propose a resampling estimator of $p(\delta)$: the *half-sample estimator* (defined in Section 2), and study its finite sample properties under weak or no assumption on F. The half-sampling method was used for estimating standard error in the literature of sample surveys (McCarthy, 1969).

Descriptions of the half-sample estimator and its theoretical properties are given in Section 2. We discuss the robustness of the half-sample estimator and some extensions in Section 3. Section 4 contains a method of alleviating computational burden of the half-sample estimator when n is large and Section 5 contains a brief discussion of the asymptotic theory. In Section 6, we study by simulation the performances of the half-sample estimators of $p(\delta)$ when $\hat{\theta}$ is the sample mean or other types of estimators such as the sample median or the reciprocal of the sample mean. The numerical results indicate that the half-sample estimator works very well in general situations.

Since the estimation of $p(\delta)$ is closely related to the estimation of the sampling distribution of $\hat{\theta}-\theta$, our results can be applied to the latter case with minor modifications.

2. THE HALF-SAMPLE ESTIMATOR

Suppose that $X_1,\ldots, X_n$ are independent and identically distributed (i.i.d.) samples from the population distribution F. Let the parameter of interest be θ and its point estimator be $\hat{\theta}=T_n(X_1,\ldots, X_n)$, where T_n is symmetric. Suppose that the sample size n is even. Let $r=n/2$ and $\mathbf{S}_r$ be the collection of subsets of $\{ 1,\ldots,n \}$ which have size r. For $s=\{ i_1,\ldots,i_r \}\in \mathbf{S}_r$, let $\hat{\theta}_s=T_r(X_{i_1},\ldots, X_{i_r})$. The half-sample estimator of $p(\delta)$ defined in (1.1) is

$$\hat{p}(\delta) = \frac{1}{N}\sum_s I(\frac{1}{2}|\hat{\theta}_s-\hat{\theta}_{s^c}|\leq\delta), \tag{2.1}$$

where $N=\binom{n}{r}$, $I(A)$ is the indicator function of the set A, s^c is the complement of s and $\sum_s$ is the summation over all subsets in $\mathbf{S}_r$. $\hat{p}(\delta)$ is related to the delete-r jackknife histogram (Wu, 1987). See also Section 5 (c).

For s, s_1 and $s_2\in\mathbf{S}_r$, let

$$a(\delta) = P(\frac{1}{2}|\hat{\theta}_s-\hat{\theta}_{s^c}|\leq\delta), \tag{2.2}$$

k be the number of integers common to s_1 and s_2, and

$$\zeta_k= Cov[\, I(\frac{1}{2}|\hat{\theta}_{s_1}-\hat{\theta}_{s_1^c}|\leq\delta),\, I(\frac{1}{2}|\hat{\theta}_{s_2}-\hat{\theta}_{s_2^c}|\leq\delta)\,]. \tag{2.3}$$

$a(\delta)$ and ζ_k are independent of the choices of the subsets since $X_1,\ldots, X_n$ are i.i.d.

Theorem 1. The half-sample estimator $\hat{p}(\delta)$ has the following properties:

$$E[\hat{p}(\delta)] = a(\delta); \tag{2.4}$$

$$Var[\hat{p}(\delta)] = \frac{1}{N}\sum_{k=1}^{r}\binom{r}{k}^2\zeta_k; \tag{2.5}$$

and

$$Var[\hat{p}(\delta)] \le \frac{1}{2}a(\delta)[1-a(\delta)]. \tag{2.6}$$

Proof. Since $X_1,\ldots,X_n$ are i.i.d.,

$$E[\hat{p}(\delta)] = P(\frac{1}{2}|\hat{\theta}_s - \hat{\theta}_{s^c}| \le \delta) = a(\delta).$$

Note that $\hat{p}(\delta)$ is a U-statistic with a kernel $I(\frac{1}{2}|\hat{\theta}_{s_o} - \hat{\theta}_{s_o^c}| \le \delta)$, $s_o = \{1,\ldots,r\}$. Therefore (2.5) and (2.6) follow from standard results on U-statistics (Hoeffding, 1948) and

$$Var[I(\frac{1}{2}|\hat{\theta}_{s_o} - \hat{\theta}_{s_o^c}| \le \delta)] = a(\delta)[1-a(\delta)]. \ \square$$

From (2.4), the bias of $\hat{p}(\delta)$ is $a(\delta)-p(\delta)$. The exact expression of $Var[\hat{p}(\delta)]$ in (2.5) may be hard to evaluate. The upper bound for $Var[\hat{p}(\delta)]$ in (2.6) is quite efficient in many cases. This upper bound achieves its maximum 0.125 when $a(\delta)=0.5$. For the accuracy of $\hat{p}(\delta)$, sometimes it is appropriate to look at a relative measure

$$\frac{\{Var[\hat{p}(\delta)]\}^{1/2}}{p(\delta)} \le \frac{\{a(\delta)[1-a(\delta)]/2\}^{1/2}}{p(\delta)}. \tag{2.7}$$

For the purpose of estimating the accuracy $p(\delta)$, it is desirable to have an unbiased estimator, i.e., $a(\delta)=p(\delta)$ for all F, or a conservative estimator, i.e., $a(\delta) \le p(\delta)$ for all F. We show in the following that we do achieve the unbiasedness of $\hat{p}(\delta)$ in the situation where θ is the population mean (assume it exists) and $\hat{\theta}$ is the sample mean $\bar{X}=n^{-1}\sum_{i=1}^n X_i$. Even for this simple case $(\hat{\theta}=\bar{X})$ there is no satisfactory method of estimating the accuracy of $\bar{X}$ available when the sample size n is not large and/or the population distribution F has heavy tails. For instance, the sample variance, a customary estimator of the population variance, is highly fluctuating when n is small and the population has a large variance and becomes meaningless when the population variance does not exist. In fact, if the population variance is nearly infinity or very large, it is not appropriate to use it as an accuracy measure. The following properties of $\hat{p}(\delta)$ are established under very weak assumption on F.

When $\hat{\theta}=\bar{X}$, $\hat{\theta}_s=\bar{X}_s=r^{-1}\sum_{i\in s}X_i$ for $s\in \mathbf{S}_r$. From (2.1), the half-sample estimator of $p(\delta)=P(\,|\bar{X}-\theta|\le\delta\,)$ reduces to:

$$\hat{p}(\delta)=\frac{1}{N}\sum_s I\left(\frac{1}{2}|\bar{X}_s-\bar{X}_{s^c}|\le\delta\right). \tag{2.8}$$

Let H_s be the distribution function of $\bar{X}_s$. Note that $\bar{X}_s$ have the same distribution for all s since $X_1,\ldots,X_n$ are i.i.d.

Theorem 2. Suppose that H_s is symmetric about θ. Then in the case of $\hat{\theta}=\bar{X}$, (2.4)-(2.6) hold with $a(\delta)$ replaced by $p(\delta)$.

Proof. We only need to show $a(\delta)=p(\delta)$. This is a direct consequence of Lemma A in the Appendix. □

Hence $\hat{p}(\delta)$ is unbiased. The relative measure defined in (2.7) is now bounded by $(0.5)^{1/2}$ times the square root of the odds ratio $[1-p(\delta)]/p(\delta)$, which is a decreasing function of $p(\delta)$ and tends to zero (infinity) as $p(\delta)$ tends to one (zero).

The symmetry assumption of H_s is not serious. It is implied by the symmetry of the population distribution F. H_s is approximately symmetric if n is large and $a_n(\bar{X}-\theta)$ is asymptotically normal, where a_n is a sequence of positive constants. Data transformation can also be used to achieve the symmetry. The robustness of $\hat{p}(\delta)$ against asymmetry will be discussed in Section 6.

3. SOME EXTENSIONS AND ROBUSTNESS OF THE HALF-SAMPLE ESTIMATOR

We discuss in this section the robustness of the half-sample estimator $\hat{p}(\delta)$ and some extensions. Some other extensions will be discussed in Section 6.

(a) *Robustness against heavy-tail distributions.* The results in Theorem 2 hold even if the mean of F does not exist. For example, F is Cauchy with density $\pi^{-1}\sigma[\sigma^2+(x-\theta)^2]^{-1}$. In this case θ is the center of the symmetric distribution F. In fact, we have

Theorem 3. The results in Theorem 2 are true without any moment condition on F.

Proof. From the proofs of Theorems 1 and 2, the results follow from

$$\frac{1}{2}(\bar{X}_s-\bar{X}_{s^c}) \text{ and } \bar{X}-\theta \text{ has the same distribution.} \tag{3.1}$$

From Lemma A in the Appendix, (3.1) is true without any moment condition on F. □

(b) *Robustness against non-identical distributions.* When $\hat{\theta}=\bar{X}$, another interesting robustness property that $\hat{p}(\delta)$ has is that $\hat{p}(\delta)$ is unbiased even if X_i are not identically distributed. This is certainly a very nice property of an estimator since in practice often we are not so sure about whether the X_i have exactly the same distribution. An example is

$$X_i = \theta+\varepsilon_i,$$

where ε_i are independently distributed as $F(x/\sigma_i)$, $F(x)$ is symmetric about zero, and the scale parameters σ_i may or may not be equal.

Theorem 4. For $\hat{\theta}=\bar{X}$, $\hat{p}(\delta)$ is unbiased even if the X_i are not identically distributed.

Proof. From Lemma A in the Appendix, (3.1) still holds if X_i are not identically distributed. This proves the result. □

(c) *The case that n is not even.* When n is odd, the estimator $\hat{p}(\delta)$ defined in (2.1) can still be used with $r=(n-1)/2$. This is motivated by the fact that in the situation where $\hat{\theta}=\bar{X}$ and the variance of X_1 exists and is equal to σ^2, we have

$$Var[\frac{1}{2}(\bar{X}_s-\bar{X}_{s^c})] = \frac{n}{(n+1)(n-1)}\sigma^2 \geq \frac{1}{n}\sigma^2.$$

Thus $\hat{p}(\delta)$ will usually be a conservative estimator of $p(\delta)$. Two examples are given below.

For the variance of $\hat{p}(\delta)$ when n is odd, we have

$$Var[\hat{p}(\delta)] = \frac{1}{N}\sum_{k=1}^{r}\binom{r}{k}\binom{n-r}{r-k}\zeta_k \leq \frac{n-1}{2n}a(\delta)[1-a(\delta)],$$

where $a(\delta)$ and ζ_k are defined in (2.2) and (2.3), respectively.

Example 1. Suppose that F belongs to a family $\mathbf{F}$ satisfying the following properties:

(A) The variance of F exists for any $F \in \mathbf{F}$;

(B) If $X_1,\ldots, X_n$ are i.i.d. with $F \in \mathbf{F}$, $\theta=EX_1$ and c_i, d_i, $i=1,\ldots,n$, are constants, then

$$Var(\sum_{i=1}^{n} c_iX_i) \geq Var(\sum_{i=1}^{n} d_iX_i)$$

implies

$$P(\ |\sum_{i=1}^{n} c_i(X_i-\theta)|\leq\delta\ ) \leq P(\ |\sum_{i=1}^{n} d_i(X_i-\theta)|\leq\delta\ ).$$

Then $\hat{p}(\delta)$ defined in (2.8) with $r=(n-1)/2$ is a conservative estimator of $p(\delta)$, i.e.,

$$E[\hat{p}(\delta)] \leq P(\ |\bar{X}-\theta|\leq\delta\ ).$$

An example of a family of distributions having properties (A) and (B) is the normal distribution family $\{\ N(\theta, \sigma^2), -\infty<\theta<\infty, \sigma^2>0\ \}$.

Example 2. Suppose that F belongs to the Cauchy distribution family with location parameter θ and scale parameter σ. Thus F has a density

$$f(x, \theta, \sigma) = \pi^{-1}\sigma[\sigma^2+(x-\theta)^2]^{-1}, \quad -\infty<\theta<\infty, \ \sigma>0.$$

Note that if $X_1,\ldots, X_n$ are i.i.d. as F, then $\bar{X}-\theta$ and $\frac{1}{2}(\bar{X}_s-\bar{X}_{s^c})$ have the same distribution with density $f(x, 0, \sigma)$. Hence $\hat{p}(\delta)$ is still unbiased for $p(\delta)$.

(d) *Extension to regression.* Another non-i.i.d. situation is the regression problem

$$y_i = x_i'\beta + e_i, \quad i=1,\ldots,n, \tag{3.2}$$

where y_i are observations, x_i are known p-vectors, β is an unknown p-vector, and the random errors e_i are i.i.d. as F. We want to estimate the statistical accuracy of the least squares estimator of $\theta=c'\beta$, where c is a known p-vector. Let $r=n/2$ if n is even and $r=(n-1)/2$ if n is odd, $\mathbf{S}_r$ be defined as before, $A=\sum_{i=1}^n x_ix_i'$ and $A_s=\sum_{i\in s} x_ix_i'$ for $s\in\mathbf{S}_r$. Define

$$\mathbf{S}_r^T=\{\ s\in\mathbf{S}_r : A_s \text{ and } A_{s^c} \text{ are nonsingular } \}.$$

Assume the number of subsets in $\mathbf{S}_r^T$ is $T>0$. The least squares estimators of β under model (3.2) and the submodel

$$y_i = x_i'\beta + e_i, \quad i\in s, \quad s\in\mathbf{S}_r^T$$

are $\hat{\theta}=c'(A^{-1}\sum_{i=1}^n x_iy_i)$ and $\hat{\theta}_s=c'(A_s^{-1}\sum_{i\in s} x_iy_i)$, respectively. Denote

$$p(\delta) = P(\ |\hat{\theta}-\theta|\le\delta\ ).$$

Then $\hat{p}(\delta)$ defined in (2.1) can be extended to

$$\hat{p}(\delta) = \frac{1}{T}\sum_{s\in\mathbf{S}_r^T} I(\tfrac{1}{2}\,|\hat{\theta}_s-\hat{\theta}_{s^c}|\le\delta\ ). \tag{3.3}$$

Theorem 5. If $F \in \mathbf{F}$ satisfying the properties (A) and (B) stated in Example 1, then the estimator defined in (3.3) is a conservative estimator of $p(\delta)$.

The proof is given in the Appendix.

4. RANDOM HALF-SAMPLE ESTIMATOR

The computation of $\hat{p}(\delta)$ requires $N=\binom{n}{r}$ evaluations of $\hat{\theta}_s$. Table 1 gives the values of N for even n ranging from 6 to 20. For n larger than 20, it may be hard to evaluate $\hat{p}(\delta)$. However, it is not necessary to compute all N terms in (2.1) in order to obtain a good estimator of $p(\delta)$. Shortcuts can be taken by applying various techniques with $\hat{\theta}_s$ evaluated only for s in a subclass $\mathbf{S}_r^{(1)}$ of $\mathbf{S}_r$. If the size of $\mathbf{S}_r^{(1)}$ is T, then

$$\hat{p}_1(\delta) = \frac{1}{T}\sum_{s \in \mathbf{S}_r^{(1)}} I(\frac{1}{2}|\hat{\theta}_s - \hat{\theta}_{s^c}| \leq \delta)$$

is an incomplete U-statistic. $\hat{p}_1(\delta)$ has the same expectation as $\hat{p}(\delta)$ for any fixed $\mathbf{S}_r^{(1)}$. Hence $\hat{p}_1(\delta)$ is still an unbiased estimator when the conditions in Theorem 2 are satisfied.

Obviously $\hat{p}_1(\delta)$ has larger variance than $\hat{p}(\delta)$ and we would like to choose a subclass from $\mathbf{S}_r$ such that the increase in variance is as small as possible. There are some results (e.g., Blom, 1976; Brown and Kildea, 1978) of choosing a suitable subclass to obtain an efficient incomplete U-statistic when the size of the kernel of the U-statistic is much smaller than n. Since in our case the kernel size is $r=n/2$, these results can not be applied directly.

The problem of choosing a fixed subclass to obtained an efficient approximation to $\hat{p}(\delta)$, such as the construction of a "balanced" subclass (see Blom (1976) and Brown and Kildea (1978)), is interesting and needs further theoretical investigation. In the following we will focus on a simple but useful method: *random sampling*. That is, we select a simple random sample (srs) of size m (with or without replacement) from $\mathbf{S}_r$. This technique is not new and is applied in many statistical applications. Advantages of using random sampling are (1) the random sampling is simple and easy to use; and (2) the loss of efficiency of the resulting approximation to $\hat{p}(\delta)$ can be controlled by selecting a suitable m, as the following discussion indicates.

Suppose that $\mathbf{S}_r^{srs}$ of size m is an srs from $\mathbf{S}_r$. The approximation to $\hat{p}(\delta)$ is then

$$\hat{p}_{srs}(\delta) = \frac{1}{m}\sum_{s\in \mathbf{S}_r^{srs}} I\left(\frac{1}{2}\,|\hat{\theta}_s - \hat{\theta}_{s^c}| \le \delta \right). \tag{4.1}$$

There are actually $N/2$ terms on the right hand side of (2.1). Hence $m \le N/2$. Let E_* and Var_* be the expectation and variance taken under the second stage random sampling for given $X_1,\ldots, X_n$. Then

$$E[\hat{p}_{srs}(\delta)] = E\{E_*[\hat{p}_{srs}(\delta)]\} = E[\hat{p}(\delta)].$$

Hence $\hat{p}_{srs}(\delta)$ is unbiased (or conservative) if $\hat{p}(\delta)$ is. Note that

$$Var[\hat{p}_{srs}(\delta)] = E\{Var_*[\hat{p}_{srs}(\delta)]\} + Var\{E_*[\hat{p}_{srs}(\delta)]\}$$

$$= E\{Var_*[\hat{p}_{srs}(\delta)]\} + Var[\hat{p}(\delta)].$$

Hence the increase in variance by using random sampling is $E\{Var_*[\hat{p}_{srs}(\delta)]\}$. From sampling theory (Cochran, 1977),

$$Var_*[\hat{p}_{srs}(\delta)] = \frac{1-f}{m(N-1)}\sum_s [I(\frac{1}{2}\,|\hat{\theta}_s - \hat{\theta}_{s^c}| \le \delta) - \hat{p}(\delta)]^2,$$

where $f = m/N$ if the samples are taken without replacement and $f = 1/N$ otherwise. Therefore

$$E\{Var_*[\hat{p}_{srs}(\delta)]\} \le \frac{1}{m} a(\delta) \le \frac{1}{m}, \tag{4.2}$$

where $a(\delta)$ is defined in (2.2). Define the following relative measure of efficiency loss

$$\Delta = \{[Var(\hat{p}_{srs}(\delta))]^{\frac{1}{2}} - [Var(\hat{p}(\delta))]^{\frac{1}{2}}\} / [Var(\hat{p}(\delta))]^{\frac{1}{2}},$$

which is the relative increase in root mean squared error if $\hat{p}(\delta)$ is unbiased. It is not possible to control Δ under a fixed level by just choosing an m (unless $\mathbf{S}_r^{srs} = \mathbf{S}_r$), since $Var[\hat{p}(\delta)]$ can be arbitrary small. However, when $Var[\hat{p}(\delta)]$ is very small, say $< \varepsilon$, $Var[\hat{p}_{srs}(\delta)] < 2\varepsilon$ is plenty enough for practical use, although Δ may be large. According to this idea, we have the following working rule (for choosing m).

Theorem 6. Let ε, τ and ρ be given positive small constants, $\tau > \varepsilon$. If we choose an m according to

$$m = 1 + \text{the integer part of } \min\{N/2, \max[\varepsilon^{-1}((\rho+1)^2-1)^{-1}, (\tau-\varepsilon)^{-1}]\}, \qquad (4.3)$$

then we have either

$$\Delta \leq \rho$$

or

$$Var[\hat{p}(\delta)] \leq \varepsilon \quad \text{and} \quad Var[\hat{p}_{srs}(\delta)] \leq \tau.$$

The proof of this result is given in the Appendix. ε and τ are predescribed accuracy of $\hat{p}(\delta)$ and $\hat{p}_{srs}(\delta)$, respectively, and ρ is a desirable control level of efficiency loss. We can choose a small ε (e.g., ε=0.01) and a τ=2ε or 1.5ε. Then if $Var[\hat{p}(\delta)]$ is smaller than ε, (4.3) ensures that $Var[\hat{p}_{srs}(\delta)]$ does not exceed τ. On the other hand if $Var[\hat{p}(\delta)]$ is not small, the relative efficiency loss Δ is controlled to be smaller than ρ under (4.3). Table 2 gives some values of m (n is assumed to be larger than 14) obtained from (4.3) for some combinations of ε, τ and ρ.

Table 1

n	6	8	10	12	14	16	18	20
N	20	70	252	924	3,432	12,870	48,620	184,756

Table 2

ε	.05	.05	.05	.01	.01	.01
τ	.06	.06	.06	.02	.02	.02
ρ	.10	.05	.01	.10	.05	.01
m	100	196	996	477	976	4,976

5. ASYMPTOTICS

Although we are mainly focusing on the finite sample properties of $\hat{p}(\delta)$, it is also interesting to know some asymptotic properties of $\hat{p}(\delta)$. We glance over the consistency of $\hat{p}(\delta)$ for the case $\hat{\theta}=\bar{X}$ in this section.

(a) *Limiting behavior of* $\hat{p}(\delta)$ *when n is fixed.* When n is fixed, $p(\delta)$ tends to one (zero) as δ tends to infinity (zero). From (2.6) and $a(\delta)=p(\delta)$, the variance of $\hat{p}(\delta)$ tends to zero if δ tends to infinity or zero. Hence $\hat{p}(\delta)-p(\delta)\rightarrow 0$ in L_2 space as δ tends to infinity or zero.

(b) *Limiting behavior of* $\hat{p}(\delta)$ *when* δ *is fixed.* If the mean of the population distribution exists, then as $n\rightarrow\infty$, $\bar{X}\rightarrow\theta$ *a.s.* according to the law of large numbers. Hence $\lim_{n\rightarrow\infty} p(\delta)=1$ and therefore $\lim_{n\rightarrow\infty} Var[\hat{p}(\delta)]=0$ and $\hat{p}(\delta)-p(\delta)\rightarrow 0$ in L_2 space.

(c) *The case* δ *is a function of n.* In practice, δ can be taken as a function of the sample size n. For example, if we take $\delta=\delta_n=\delta_0 n^{-1/2}$ for some $\delta_0>0$, then $\hat{p}(\delta_n)$ is the same as the jack-knife histogram (Wu, 1987). If we assume that the variance of the population distribution exists and is equal to σ^2, then by the central limit theorem,

$$p(\delta_n) = P(n^{1/2}|\bar{X}-\theta| \le \delta_0) \to \Phi(\delta_0/\sigma) - \Phi(-\delta_0/\sigma),$$

where $\Phi(t)$ is the standard normal distribution function. In this case, the upper bound of $Var[\hat{p}(\delta_n)]$ in (2.6) does not tends to zero. However, $\hat{p}(\delta_n)$ is still consistent, i.e.,

$$\hat{p}(\delta_n) - p(\delta_n) \to 0 \ \ a.s.$$

according to Theorem 1 of Wu (1987).

(d) *The random half-sample estimator.* The assertions in (a)-(c) are true for $\hat{p}_{srs}(\delta)$ if $\min(n, m) \to \infty$.

6. DISCUSSIONS, EXTENSIONS AND NUMERICAL RESULTS

Theorem 1 in Section 2 provides some ideas about the bias and the variance of $\hat{p}(\delta)$. Theorem 2 gives more precise results when $\hat{\theta}$ is the sample mean and H_s is symmetric. Extensions of the results in Theorem 2 to the following cases are of both theoretical and practical interests: (a) the situation where $\hat{\theta}=\bar{X}$ but H_s is not symmetric; and (b) the general case where the point estimator $\hat{\theta}$ is arbitrary. Further theoretical studies are called for. In this section we give a discussion with some numerical results.

(a) *Numerical results for symmetric distributions and* $\hat{\theta}=\bar{X}$. We first present some simulation results for the simple case where $\hat{\theta}$ is the sample mean and the population distribution is symmetric. Two symmetric distributions are considered: normal distribution with mean 1 and variance 0.25 and Cauchy distribution with median 1 and shape parameter 0.2. Thus θ=1. The sample size is n=16. The estimator of $p(\delta)$ is $\hat{p}_{srs}(\delta)$ defined in (4.1) with m=200 (see Table 2). All the results presented below are based on 3000 simulations on a VAX 11/780 at the Purdue University.

The biases (BIAS) and root mean squared errors (RMSE) of $\hat{p}_{srs}(\delta)$ are shown in Table 3 for two choices of δ: 0.1 and 0.2. The values of $p(\delta)$ are also included. The results show that $\hat{p}_{srs}(\delta)$ performs very well. The biases are all relatively negligible. In the normal population case, the RMSE of $\hat{p}_{srs}(\delta)$ is much smaller than the bound given in (2.6).

(b) *Numerical results for asymmetric distributions and* $\hat{\theta}=\bar{X}$. To see the effect of asymmetry, we also study the BIAS and RMSE of $\hat{p}_{srs}(\delta)$ for the cases where the distributions are Gamma with mean 1 and variance 0.5 and scaled Poisson with parameter 5. θ in both cases are 1. Note that H_s are asymmetric for both cases and is discrete in the Poisson case.

The results are included in Table 3. It is clear that the half-sample estimator works well in these asymmetric situations.

Table 3. BIAS and RMSE of $\hat{p}_{srs}(\delta)$ ($\hat{\theta}$ = the sample mean)

	δ=0.1			δ=0.2		
Distribution	$p(\delta)$	BIAS	RMSE	$p(\delta)$	BIAS	RMSE
Normal	.569	.007	.096	.889	.001	.070
Cauchy	.292	−.005	.272	.491	−.000	.372
Gamma	.417	.022	.126	.748	.002	.137
Poisson	.306	−.002	.081	.518	.014	.104

(c) *The case of sample median.* We now focus on other types of point estimators. In some situations we are interested in the population median instead of the population mean. The point estimator is then the sample median defined to be $(Y_{n/2}+Y_{(n+2)/2})/2$ if n is even and $Y_{(n+1)/2}$ if n is odd, where $Y_1,\ldots, Y_n$ are the order statistics of the samples $X_1,\ldots, X_n$. We study the performances of $\hat{p}_{srs}(\delta)$ given in (a) when the population distributions are normal, Cauchy and Gamma described in (a)-(b). The results are reported in Table 4. The half-sample estimator still works well. The biases of $\hat{p}_{srs}(\delta)$ are slightly larger.

Table 4. BIAS and RMSE of $\hat{p}_{srs}(\delta)$ ($\hat{\theta}$ = the sample median)

	δ=0.1			δ=0.2		
Distribution	$p(\delta)$	BIAS	RMSE	$p(\delta)$	BIAS	RMSE
Normal	.487	.026	.180	.818	.018	.146
Cauchy	.784	−.038	.189	.971	−.015	.081
Gamma	.398	.013	.175	.699	.017	.184

(d) *The reciprocal of the sample mean.* Another type of estimator considered is the reciprocal of the sample mean (as an estimator of the reciprocal of the mean θ). The same distributions as in (c) are considered and the results are given in Table 5. The performances of $\hat{p}_{srs}(\delta)$ are again very good. The biases are all relatively negligible.

Table 5. BIAS and RMSE of $\hat{p}_{srs}(\delta)$ ($\hat{\theta}$ = the reciprocal of sample mean)

	δ=0.1			δ=0.2		
Distribution	$p(\delta)$	BIAS	RMSE	$p(\delta)$	BIAS	RMSE
Normal	.574	−.006	.143	.884	−.028	.123
Cauchy	.294	.004	.282	.502	.001	.374
Gamma	.420	.002	.118	.738	−.020	.138

(e) *Concluding remarks.* The half-sample estimator of $p(\delta)$ works very well in all the situations under consideration. The results in (d) indicate that the half-sample estimator performs well if $\hat{\theta}=g(\bar{X})$ and g is smooth at θ. Another observation from Tables 3 and 4 is that for the estimation of the center of a symmetric distribution, the sample median is better than the sample mean when the population distribution has heavy tails (such as Cauchy) whereas the sample mean is preferred when the population distribution has second order moment (such as normal). Thus, good estimates of the coverage probabilities $p(\delta)$ enable us to compare the efficiency of the point estimators.

ACKNOWLEDGEMENTS

I would like to thank Professor Leon Jay Gleser for drawing my attention to this problem.

APPENDIX

Lemma A. Suppose that W_i is distributed as F_i, $i=1,2$, W_i are independent and F_i are symmetric about 0. Then W_1-W_2 and W_1+W_2 have the same distribution.

Proof. Let ϕ_i be the characteristic function of F_i, $i=1,2$. Then $\phi_i(t)=\phi_i(-t)$ since F_i is symmetric. The characteristic function of W_1-W_2 is

$$\phi_1(t)\phi_2(-t) = \phi_1(t)\phi_2(t),$$

which is the characteristic function of W_1+W_2. This proves the result. □

Proof of Theorem 5. We only need to show that when $s \in \mathbf{S}_r^T$,

$$Var[\frac{1}{2}(\hat{\theta}_s - \hat{\theta}_{s^c})] \geq Var\hat{\theta}. \tag{A1}$$

Denote $Var(e_i)$ by σ^2. Then $Var\hat{\theta}=\sigma^2 c'A^{-1}c$ and $Var[\frac{1}{2}(\hat{\theta}_s-\hat{\theta}_{s^c})]=\sigma^2 c'(A_s^{-1}+A_{s^c}^{-1})c/4$. Thus (A1) follows from

$$A_s^{-1}+A_{s^c}^{-1} \geq 4A^{-1},$$

which is simply the matrix version of Jensen's inequality. □

Proof of Theorem 6. If $N/2 \leq \max[\varepsilon^{-1}((\rho+1)^2-1)^{-1}, (\tau-\varepsilon)^{-1}]$, then $m=N/2$ and $\hat{p}_{srs}(\delta)=\hat{p}(\delta)$. Suppose that $N/2 > \max[\varepsilon^{-1}((\rho+1)^2-1)^{-1}, (\tau-\varepsilon)^{-1}]$. If $Var[\hat{p}(\delta)] \leq \varepsilon$, then $Var[\hat{p}_{srs}(\delta)] \leq \varepsilon + m^{-1}$ by (4.2). Hence $Var[\hat{p}_{srs}(\delta)] \leq \tau$ since $m \geq (\tau-\varepsilon)^{-1}$ from (4.3). If $Var[\hat{p}(\delta)] > \varepsilon$, then

$$\Delta \le \{m^{-1}[Var(\hat{p}(\delta))]^{-1}+1\}^{½}-1 \le (m^{-1}\varepsilon^{-1}+1)^{½}-1 \le \rho,$$

since m is chosen to be not smaller than $\varepsilon^{-1}[(\rho+1)^2-1]^{-1}$. □

REFERENCES

Abramovitch, L. and Singh, K. (1985). Edgeworth corrected pivotal statistics and the bootstrap. *Ann. Statist. 13,* 116-132.

Beran, R. (1987). Prepivoting to reduce level error of confidence sets. *Biometrika 74,* 457-468.

Bhargava, R. P. (1983). A property of the jackknife estimation of the variance when more than one observation is omitted. *Sankhya A 45,* 112-119.

Blom, G. (1976). Some properties of incomplete U-statistics. *Biometrika 63,* 573-580.

Brown, B. M. and Kildea, D. G. (1978). Reduced U-statistics and the Hodged-Lehmann estimator. *Ann. Statist. 6,* 828-835.

Cochran, G. (1977). *Sampling Techniques.* 3rd edition. Wiley, New York.

Efron, B. (1979). Bootstrap methods: another look at the jackknife. *Ann. Statist. 7,* 1-26.

Efron, B. and Stein, C. (1978). The jackknife estimator of variance. *Ann. Statist. 9,* 586-596.

Hall, P. (1986) On the bootstrap and confidence intervals. *Ann. Statist. 14,* 1431-1452.

Hinkley, D. V. (1977). Jackknifing in unbalanced situations. *Technometrics 19,* 285-292.

Kiefer, J. (1977). Conditional confidence statements and confidence estimators (with discussion). *J. Amer. Statist. Assoc. 72,* 789-827.

McCarthy, P. J. (1969). Pseudo-replication: half-samples. *Rev. ISI 37,* 239-263.

Quenouille, M. (1956). Notes on bias in estimation. *Biometrika 43,* 353-360.

Rao, J. N. K. and Wu, C. F. J. (1985). Inference from stratified samples: second order analysis of three methods for nonlinear statistics. *J. Amer. Statist. Assoc. 80,* 620-630.

Shao, J. and Wu, C. F. J. (1987). Heteroscedasticity-robustness of jackknife variance estimators in linear models. *Ann. Statist. 15,* 1563-1579.

Tukey, J. (1958). Bias and confidence in not quite large samples. *Ann. Math. Statist. 29,* 614.

Wu, C. F. J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis (with discussion). *Ann. Statist. 14,* 1261-1350.

Wu, C. F. J. (1987). On the asymptotic properties of the jackknife histograms. Preprint.