The Annals of Statistics
1983, Vol. 11, No. 4, 1251–1256

THE ROBUSTNESS OF STEIN'S TWO-STAGE PROCEDURE

BY RAMKARAN

Lucknow University, India

The expressions for coverage probability and ASN of the Stein's two-stage confidence interval procedure for estimating the normal mean have been obtained under the assumption that the underlying distribution is, in fact, different from normal but could be approximated by the first four terms of Edgeworth series. The comparison of coverage probabilities with the corresponding probabilities obtained for the normal distribution shows that the procedure is quite insensitive to moderate departures from normality, and skewness of the parent population has very little effect on the coverage probability of the procedure. The Monte Carlo investigations which involve sampling from the gamma population confirm these conclusions.

1. Introduction. The first attempt to study the robustness of Stein's two-stage procedure against the possible departures from normality was made by Bhattacharjee (1965) who assumed that the parent population could be approximated by the first four terms of the Edgeworth series. Recently, Blumenthal and Govindarajulu (1977) investigated this problem, assuming the parent population to be a mixture of two normal populations differing in location parameters and having the same unknown variance. Both of these studies are concerned with "Criterion—robustness" of the procedure (cf. Box and Tiao, 1973), but give somewhat conflicting conclusions. Whereas Blumenthal and Govindarajulu claim that the procedure is remarkably robust, Bhattacharjee's assertion is surprisingly just the reverse. There seem to be some oversights in Bhattacharjee's work which would account for this discrepancy. The purpose of this article is, therefore, to show that Stein's procedure is, in fact, quite robust even under the Edgeworth series model.

2. The two-stage procedure. Let $Y_1, Y_2, Y_3, \cdots$ be a sequence of iid random observations from $N(\mu, \sigma^2)$. Take an initial sample of size m from this sequence and calculate the unbiased estimate S_Y^2 of σ^2. Then take an additional sample of size $N - m$ from the same sequence where

$$N = \max\{m, [S_Y^2/Z] + 1\}. \tag{2.1}$$

The quantity $Z > 0$ is a preassigned constant and $[q]$ stands for the greatest integer less than q. If $\bar{Y}_N$ is the mean of N observations obtained by pooling the two samples, then the "pivotal-quantity"

$$t = \sqrt{N}\,(\bar{Y}_N - \mu)/S_Y \tag{2.2}$$

could be used for the inferences about μ. Stein has shown that the sampling distribution of t is Student's t with $m - 1$ d.f. and the coverage probability of the interval $(\bar{Y}_N \pm d)$, a function of σ^2, is always greater than or equal to $1 - \alpha$, irrespective of the values of μ and σ^2, provided $Z = d^2/b^2$. The quantity b is the upper 100 $\alpha/2\%$ point of t distribution with $m - 1$ d.f.

Received April 1981; revised April 1982.
AMS 1970 *subject classification.* Primary, 62F35; secondary, 62L12.
Key words and phrases. ASN, confidence interval, coverage probability, Edgeworth series, Monte Carlo, pivotal quantity.

3. Comments on Bhattacharjee's paper. Bhattacharjee claims to base his investigations on the " pivotal-quantity" t defined at (2.2) but, in fact, uses instead the quantity $t' = (\bar{Y}_N - \mu)/\sqrt{Z}$ for computing the power of Student's hypothesis and the coverage probability of the interval described above. But the distribution of t' is Student's t with $m - 1$ d.f. only when $\sigma^2 \to \infty$ [Cf. Ruben, 1961] and not in general. The coverage probability of the interval based on the procedure in question is, in fact, not independent of σ^2 or equal to 1-α for all σ^2 even if the sampling is done from the normal population. Moreover, he has calculated the power and confidence level for the parametric region $(-1.0 \le \lambda_3 \le 1.0; -.5 \le \lambda_4 \le 2.5)$ in which the Edgeworth series may not represent a positive definite and unimodal probability density [Cf. Draper and Tierney, 1972]. The conclusions based on the calculations carried out for the parameter values outside the validity region (ie $-.45 \le \lambda_3 \le .45; 0 \le \lambda_4 \le 2.35$) of the series might be very misleading. This fact becomes quite clear later in our study.

4. The coverage probability and ASN. Let us assume that the probability distribution of Y which has mean μ and variance σ^2 can be approximated by the first four terms of the Edgeworth series. That is

$$f\left(\frac{y-\mu}{\sigma}\right) = f(x) = \phi(x) - \frac{\lambda_3}{6}\frac{d^3\phi(x)}{dx^3} + \frac{\lambda_4}{24}\frac{d^4\phi(x)}{dx^4} + \frac{\lambda_3^2}{72}\frac{d^2\phi(x)}{dx^6}, \tag{4.1}$$

where $\phi(x)$ is the density function of standard normal variate. The parameters λ_3 and λ_4 are the coefficients of skewness and kurtosis respectively. In order to evaluate the expressions for the coverage probability and ASN when sampling is done from (4.1), we need the joint density of $\bar{Y}_N$ and S_Y^2 and the density of S_Y^2 or else the joint density of $\bar{X}_N$ and S_X^2 and the density of S_X^2 where $X = (Y - \mu)/\sigma$. These densities could be worked out using Gayen's (1949) expression for the joint distribution of the sample mean and variance [Cf. Bhattacherjee, 1965]. Let the joint densities of sample mean and variance for the single and two-stage sample be denoted by $g(\bar{X}_m, S_X^2)$ and $f(\bar{X}_N, S_X^2)$ respectively. From (2.1) it easily follows that

$$P(N = n) = \begin{cases} 0 \text{ if } n < m \\ P(0 \le S_X^2 \le (mZ)/\sigma^2) \quad \text{if } n = m \\ P\{Z(n-1)/\sigma^2 < S_X^2 \le (nZ)/\sigma^2\} \quad \text{if } n > m. \end{cases} \tag{4.2}$$

Thus the coverage probability of the interval $(\bar{Y}_N \pm d)$ is given by

$$\begin{aligned} P(\Delta, \lambda_3, \lambda_4) &= P\{\bar{Y}_N - d \le \mu \le \bar{Y}_N + d\} = P(-d/\sigma \le \bar{X}_N \le d/\sigma.) \\ &= \int\int_{R_1} g(\bar{X}_m, S_X^2)\, d\bar{X}_m\, dS_X^2 + \textstyle\sum_{n=m+1}^{\infty} \int\int_{R_2} f(\bar{X}_n, S_X^2)\, d\bar{X}_n\, dS_X^2, \end{aligned} \tag{4.3}$$

where

$$R_1: \{-\Delta \le \bar{X}_m \le \Delta;\ 0 \le S_X^2 \le (mZ)/\sigma^2\},$$

$$R_2: \{-\Delta \le \bar{X}_n \le \Delta;\ Z(n-1)/\sigma^2 < S_X^2 \le (nZ)/\sigma^2\}$$

and $\Delta = d/\sigma$. After massive calculations and algebraic simplifications the above coverage probability could be expressed in terms of incomplete gamma integrals and incomplete moments of the standard normal distribution as follows:

$$P(\Delta, \lambda_3, \lambda_4) = u_0 V_0 + \frac{\lambda_4}{24} [\{m^{-1}u_4 - 6u_2 + 3mu_0\} V_0 + 6(m-1)(m^{-1}u_2 - u_0) V_1 + 3m^{-1}(m-1)^2 u_0 V_2]$$
$$+ \frac{\lambda_3^2}{72} [6m^{-1}(m-1)(m-2)u_0 V_3 + \{m^{-1}u_6 - 3m^{-1}(2m+3)u_4 + 9(m+4)u_2 - 15mu_0\} V_0 + 6(m-1)\{m^{-1}u_4 - 3m^{-1}(m+3)u_2 + 6u_0\} V_1 + 9m^{-1}(m-1)\{(m+1)u_2 - 3(m-1)u_0\} V_2]$$
$$+ \textstyle\sum_{n=m+1}^{\infty} \Big[W_0 Z_0 + \frac{\lambda_4}{24} \Big[\{n^{-1}W_4 - 6n^{-1}mW_2 + 3(m + (nm)^{-1}(n-m)(1-2m)) W_0\} Z_0 + 6(m-1)\Big\{n^{-1}W_2 + \Big(\frac{n-m}{mn} - 1\Big) W_0\Big\} Z_1 + 3m^{-1}(m-1)^2 W_0 Z_2 \Big]$$
$$(4.4)\qquad + \frac{\lambda_3^2}{72} \Big[\{W_6 - 3(2m+3) W_4 + 9m(m+4) W_2 - (15m^2 + 6m^{-1}(m-1)(m-2)(n-m)) W_0\} (Z_0/n) + 6(m-1)\{W_4 - 3(m+3) W_2 + 3(2m + m^{-1}(n-m)(m-2)) W_0\} (Z_1/n) + 9(m-1)\{n^{-1}(m+1) W_2 + (-3(m-1) + n^{-1}(m+1)(n-m))(W_0/m)\} Z_2 + 6m^{-1}(m-1)(m-2) W_0 Z_3] \Big]\Big]$$

where

$$(4.5)\qquad x = m\,D, \quad D = (2b^2)^{-1}(m-1)\Delta^2; \quad V_i = I_x\Big(\frac{m-1}{2} + i\Big), \quad i = 0, 1, 2, 3,$$

$I_x(p)$ is the incomplete gamma integral (normalised) and

$$(4.6)\qquad u_i = \mu_i(\sqrt{m}\,\Delta) = \int_{-\Delta\sqrt{m}}^{\Delta\sqrt{m}} y^i \phi(y)\,dy, \quad i = 0, 2, 4, 6$$

$$(4.7)\qquad Z_i = I_U\Big(\frac{m-1}{2} + i\Big) - I_L\Big(\frac{m-1}{2} + i\Big), \quad i = 0, 1, 2, 3$$

$$(4.8)\qquad U = nD, \quad L = (n-1)D, \quad W_i = \mu_i(\sqrt{n}\,\Delta), \quad i = 0, 2, 4, 6.$$

It is evident that the coverage probability is symmetric in λ_3. It could be easily verified that $P(\Delta, \lambda_3, \lambda_4)$ tends to unity as $\Delta \to \infty$ for any given pair (λ_3, λ_4). In order to derive the expression for ASN, we need probabilities $P(N = n)$, $n = m, m+1, \cdots$. These probabilities depend only on the distribution of S_X^2 and could be worked out using (4.2) and density $g(S_X^2)$. Thus, ASN is given by

$$
\begin{aligned}
E(N) &= mP(0 \le S_X^2 \le (Zm)/\sigma^2) \\
&\quad + \textstyle\sum_{n=m+1}^{\infty} nP\{Z(n-1)/\sigma^2 < S_X^2 \le (nZ)/\sigma^2\} \\
&= m[V_0 + \lambda_4(m-1)^2(V_0 - 2V_1 + V_2)/(8m) \\
&\qquad + \lambda_3^2(m-1)(m-2)(-V_0 + 3V_1 - 3V_2 + V_3)/(12m)] \\
&\quad + \textstyle\sum_{n=m+1}^{\infty} n[Z_0 + \lambda_4(m-1)^2(Z_0 - 2Z_1 + Z_2)/(8m) \\
&\qquad + \lambda_3^2(m-1)(m-2)(-Z_0 + 3Z_1 - 3Z_2 + Z_3)/(12m)]
\end{aligned} \tag{4.9}
$$

where all terms involved are the same as before. The ASN also is a function of Δ, λ_3 and λ_4 and is symmetric in λ_3. It is easy to show that ASN approaches m as $\Delta \to \infty$.

5. Discussion of the results. To study the behaviour of Stein's procedure under non-normal populations, the coverage probabilities and ASN have been computed for various values of parameters λ_3, λ_4 and Δ. The desired coverage probability $1 - \alpha$ has been set at .95 and calculations for other parameters have been done for two different first stage sample sizes $m = 5$ and $m = 11$. The various values of the coefficient of skewness and kurtosis for which the calculations have been done are $\lambda_3 = 0.0$, .2, .3, .4 and $\lambda_4 = -.5$, 0.0, .5, 1.0, 2.0, 2.4. These values except $\lambda_4 = -.5$ lie in the validity region of the Edgeworth series. The value $\lambda_4 = -.5$ has been deliberately included to show the misleading nature of the series outside its validity region. Since the expression of the coverage probability is symmetric in λ_3, the cases concerning only positive skewness have been considered. The various values of parameter Δ which have been considered are .15, .20, .25, .40, .50, .80, 1.00, 1.25, 1.50 and 2.00. We have presented here the tables for only a few values of λ_3, λ_4 and Δ, though our conclusions are based on the whole range of calculations.

As expected, the coverage probability approaches unity and ASN tends to m as Δ gets large, for the normal parent coverage probability of the procedure tends to $1 - \alpha$, .95 in our case, when $\Delta \to 0$ [Cf. Ruben, 1961]. The pattern of probabilities corresponding to $\lambda_3 = \lambda_4 = 0$ in Table 1 is consistent with this theoretical result. However, the asymptotic coverage probability in non-normal populations falls short of the required one, ie., $1 - \alpha$. Fortunately, this deficit does not seem to be very serious for practical purposes. It is clear from Table 1 that the maximum deviation of the actual coverage probability from the corresponding normal values is .014 when $\lambda_3 = 0.0$, $\lambda_4 = 2.4$ and $\Delta = .20$. The deviations in other cases are even smaller. Thus, the procedure is quite robust against the mild departures from normality. One very interesting feature which has been noticed in this investigation is that the effect of kurtosis on the coverage probability, contrary to student's t case, is more dominant than that of the skewness which seems to be rather negligible in most of the cases. Luckily, the values of ASN for Stein's procedure under non-normal populations in question are very close to those for the normal population and as such they do not pose any problem in the interpretation of the results. For $\lambda_3 = .4$ and $\lambda_4 = -.5$, although the procedure is robust, there appear some odd phenomena because of the inclusion of $\lambda_4 = -.5$. Firstly, the coverage probabilities fluctuate near $\Delta = 1.25$ and exceed unity (the coverage probabilities for $\Delta = 1.00$, 1.25 and 1.50 are .99971, 1.00003 and .99999 respectively), and secondly, the ASN is smaller than even the minimum possible value, i.e. 11 for $\Delta = 1.0$ (the ASN for $\Delta = 1.0$ is 10.99). These features, as will be clear in the following text, could only be attributed to the failure of the Edgeworth series to represent a genuine p.d.f. outside its validity region.

To confirm the above findings for small λ_3 and to study the effect of large values of λ_3 which lie outside the validity region, we have done Monte Carlo study using the gamma distribution $G(H, R)$, given by $f(x) = (H^R\,\Gamma R)^{-1}x^{R-1}\exp(-\,x/H)$; $x \ge 0$, R, $H > 0$. The

TABLE 1
The coverage probabilities and ASN *(within parentheses) for* $\alpha = .05$ *and* $m = 11$.

(λ_3, λ_4) \ Δ	.20	.40	.80	1.25
(0, 0) Normal	.951 (120.76)	.954 (30.64)	.993 (11.56)	1.000 (11.00)
(0, .5)	.948 (120.65)	.955 (30.64)	.993 (11.65)	1.000 (11.00)
(0, 2.4)	.937 (120.24)	.958 (30.61)	.994 (11.90)	1.000 (11.00)
(.2, 0)	.950 (120.76)	.953 (30.64)	.993 (11.56)	1.000 (11.00)
(.4, 0)	.950 (120.76)	.951 (30.64)	.993 (11.56)	1.000 (11.00)
(.2, .5)	.948 (120.65)	.954 (30.64)	.993 (11.61)	1.000 (11.00)
(.4, 2.4)	.936 (120.25)	.955 (30.61)	.995 (11.53)	1.000 (11.00)
(.4, −.5)	.953 (120.87)	.951 (30.65)	.994 (12.12)	1.000 (11.00)

TABLE 2
The estimates of coverage probabilities and ASN *based on simulations.*

Δ	$R = 1.0, H = 1.0$ $\lambda_3 = 2.0, \lambda_4 = 6.0$		$R = 5.0, H = 1.0$ $\lambda_3 = .894, \lambda_4 = 1.20$		$R = 40.0, H = 1.0$ $\lambda_3 = .316, \lambda_4 = .15$	
	CP	ASN	CP	ASN	CP	ASN
.20	.900 (.691)	125.71 (55.42)	.940 (.940)	124.37 (124.60)	.949 (.949)	123.67 (124.60)
.40	.927 (.679)	32.64 (14.38)	.948 (.941)	32.20 (31.62)	.949 (.952)	31.61 (31.59)
.80	.997 (.745)	12.42 (7.99)	.996 (.996)	11.81 (11.84)	.992 (.993)	11.55 (11.56)
1.25	1.000 (.999)	11.09 (11.01)	1.000 (1.000)	11.01 (11.00)	1.000 (1.000)	11.00 (11.00)

various degrees of skewness could be represented by considering different values of R. For large values of R, this distribution tends to normal distribution and, therefore, we expect that the results for such values of R should be very close to those obtained for the normal population. For estimating the coverage probabilities and ASN for different values of the width ($2d$) of the interval ($\bar{Y}_N \pm d$), we have simulated 5,000 two stage samples, using the procedure described in Section 2, from $G(H, R)$ on the M7600 machine of London University. We have used G05DGF sub-routine of NAG Library [MARK 7] for the purpose. Taking the initial sample of the size $m = 11$, we have drawn the two-stage samples from $G(1.0, 1.0)$, $G(1.0, 5.0)$ and $G(1.0, 40.0)$. For each sample, the stopping time N and the pooled mean $\bar{Y}_N$ are obtained and it is checked if the population mean (ie. RH) is covered by the interval ($\bar{Y}_N \pm d$). The relative frequency of this event in the repeated experiments gives the estimates of coverage probability. The equation (2.1) is used to calculate N and the average of all stopping times is taken as the estimate of ASN. In order to make the coverage probabilities and ASN comparable to those obtained for the normal case, we have obtained the estimates corresponding to those values of d for which Δ[i.e. $d/N(RH^2)$] = .20, .40, .80 and 1.25. Table 2 displays the estimates of coverage probabilities and ASN for the three cases mentioned above. The coverage probabilities and ASN calculated for the Edgeworth series model from formulas (4.4) and (4.9) respectively, taking same values of λ_3 and λ_4 as for the given gamma distribution, are written within parentheses. The comparison of the values with corresponding values for the normal distribution given in the first row of Table 1 shows that Stein's procedure is, in fact, robust for the values of R as low as 5. This confirms our conclusion made earlier that the skewness has very little effect on the coverage probability. The estimates are quite close to corresponding normal values for $R = 40.0$. Moreover, the difference between the coverage probabilities and ASN for the gamma distribution and those calculated for the comparable Edgeworth population are negligible provided values of λ_3 and λ_4 lie in the validity region (see case $R = 40.0$). How badly the series could behave outside this region is shown in case $R = 1.0$. Here the coverage probability falls as low as .679 (see $\Delta = .40$) whereas the corresponding probability for the gamma distribution is .927. For $\Delta = .8$, ASN is 7.99 which is smaller than even the minimum possible value, ie., 11, whereas for the corresponding gamma distribution it is 12.42.

Thus, we conclude, contrary to Bhattacharjee's assertion, that Stein's procedure for estimating the mean is remarkably robust against the moderate departures from normality. By virtue of the relationship between the confidence interval estimation and hypothesis testing, similar conclusions can be drawn for the latter problem.

Acknowledgements. This work was done while I was a Commonwealth Research Fellow at the University of Kent, Canterbury, England. I wish to thank the Association of Commonwealth Universities, London, for the financial support. I am grateful to Prof. G. B. Wetherill for the supervision and to Dr. J. Köllerström for useful discussions.

REFERENCES

BHATTACHARJEE, G. P. (1965). Effect of non-normality on Stein's two sample test. *Ann. Math. Statist.* **36** 651–663.

BOX, G. E. P. and TIAO, G. C. (1973). *Bayesian Inference in Statistical Analysis.* Addison-Wesley, Reading, Massachusetts.

BLUMENTHAL, S. and GOVINDARAJULU, Z. (1977). Robustness of Stein's two-stage procedures for mixture of normal populations. *J. Amer. Statist. Assoc.* **72** 192–196.

DRAPER, N. R. and TIERNEY, D. E. (1972). Regions of positive and unimodel series expansion of the Edgeworth and Gram-Charlier approximations. *Biometrika* **49** 463–465.

GAYEN, A. K. (1949). The distribution of Student's t in random samples of any size drawn from non-normal universe. *Biometrika* **36** 353–369.

RUBEN, H. (1961). Studentization of two-stage sample means from normal populations with unknown common variance. *Sankhya Ser. A* **23** 231–250.

STEIN, C. (1945). A two sample test for a linear hypothesis whose power is independent of the variance. *Ann. Math. Statist.* **16** 243–258.

DEPARTMENT OF STATISTICS
LUCKNOW UNIVERSITY
LUCKNOW -226 007, INDIA.