statistics 19 (1988) 4, 493–501

# On the COHEN-SACKROWITZ Estimator of a Common Mean

C. G. BHATTACHARYA
Indian Statistical Institute

**Summary.** The paper reconsiders certain estimators proposed by COHEN and SACKROWITZ [*Ann. Statist.* (1974) 2, 1274–1282, correction: *Ann. Statist.* 4, 1294] for the common mean of two normal distributions on the basis of independent samples of equal size from the two populations. It derives the necessary and sufficient condition for improvement over the first sample mean, under squared error loss, for any member of a class containing these. It shows that the estimator proposed by them for simultaneous improvement over both sample means has the desired property if and only if the common size of the samples is at least nine. The requirement is milder than that for any other estimator at the present state of knowledge and may be contrasted with their result which implies the desired property of the estimator only if the common size of the samples is at least fifteen. Upper bounds for variances if the estimators derived by them are also improved.

*AMS* 1970 *subject classifications:* Primary 62F10; secondary 62C15.

*Key words:* Common mean, unbiased estimators, minimax estimators.

## 1. Introduction

COHEN and SACKROWITZ [1974; correction (1976)] proposed some estimators for the common mean of two normal distributions when the samples available from the two populations are of equal size. They claimed the interesting result that under squared error loss, one of these would offer simultaneous improvement over both sample means when the common size of the samples is at least ten. This is in contrast to certain estimators belonging to a class proposed by KHATRI and SHAH (1974), including the GRAYBILL-DEAL estimator [GRAYBILL and DEAL (1959)], which have such a property if and only if the common size of the samples is at least eleven and the fact that no other estimator is known to have such a property. Unfortunately, an error in their paper corrected in COHEN and SACKROWITZ (1976) vitiated this result and it turned out that the desired property of their estimator could be established only if the common size of the samples is at least fifteen.

However, the technique employed by them did not yield exact condition on the common size of the samples for which their estimator has the property in question and hence the true potentialities of the estimator remained an open problem for further study. As such, we consider here a class of estimators containing the estimator discussed above and obtain the necessary and sufficient condition for improvement over the first sample mean for any member of the class. This class is implicit in the work of COHEN and SACKROWITZ and contains another estimator which they proposed for the less stringent problem of improvement over the first sample mean. We show that improvement over the first sample mean by some members of the class would be possible if and only if the common size of the samples is at least six. This is in contrast to the fact that the particular member of the class proposed by COHEN and SACKROWITZ for this purpose has the desired property if and only if the common size of the samples is at least seven and not 'at least six' as stated by them. We show furthermore, that the estimator proposed by them for simultaneous improvement over both sample means has the desired property if and only if the common size of the samples is at least nine. Lastly, we improve the upper bounds for the variances of the estimators derived by them. Section 2 gives notations and results which are basic but not explicit. The intricate problem of obtaining explicit results are tackled in Section 3.

## 2. Preliminaries

To begin with we list below some well known properties of the hypergeometric function [see e.g. LEBEDEV (1972)], in the real case, which we shall use:

$$_2F_1(\alpha, \beta; \lambda; x) = {}_2F_1(\beta, \alpha; \lambda; x) \tag{2.1}$$

$$_2F_1(0, \beta; \lambda; x) = 1 \tag{2.2}$$

$$_2F_1(\alpha-1, \beta+1; \lambda; x) - {}_2F_1(\alpha, \beta; \lambda; x) = \lambda^{-1}(\alpha-\beta-1)\, x\, {}_2F_1(\alpha, \beta+1; \lambda+1; x) \tag{2.3}$$

$$_2F_1(\alpha, \beta; \lambda; x) = \frac{\Gamma(\lambda)}{\Gamma(\beta)\,\Gamma(\lambda-\beta)} \int_0^1 t^{\beta-1}(1-t)^{\lambda-\beta-1}(1-tx)^{-\alpha}\, dt \tag{2.4}$$

$$\lambda > \beta > 0, \quad |x| < 1$$

Let $X_1$, $X_2$, $S_1$, $S_2$ be independent random variables such that

$$X_i \sim N(\mu, \eta_i), \quad S_i/\eta_i \sim \chi^2_m, \quad \eta_i > 0, \quad i = 1, 2 \tag{2.5}$$

where $\mu$, $\eta_1$, $\eta_2$ are unknown. Consider the problem of estimating $\mu$ under squared error loss, on the basis of $(X_1, X_2, S_1, S_2)$. The setting includes as a special case that for estimating the common mean of two normal distributions on the basis of independent samples of the same size from each after reduction to minimal sufficient statistics with the following correspondence: $X_1, X_2$ correspond to sample means; $S_1, S_2$ correspond to sample variances not corrected for bias; $n = m+1$ corresponds to the common size of the samples; $\eta_1, \eta_2$ correspond to the variances of the sample means. Let

$$Z = S_2/S_1; \quad \gamma = \eta_1/(\eta_1+\eta_2) \tag{2.6}$$

and note that $\gamma$ lies in the open interval (0,1).

The COHEN-SACKROWITZ estimator under consideration is given by

$$\hat{\mu}_a = X_1 + a\,(X_2 - X_1)\,W \tag{2.7}$$

where $W$ stands for the unique unbiased estimator of $\gamma$ based on $(S_1, S_2)$ and '$a$' is a constant to be suitably chosen. From COHEN and SACKROWITZ (1974), who write $G_*(z)$ for our $W$, we have (with slight modification to ensure unambiguity)

$$\begin{aligned} W &= {}_2F_1\,(1, 1-m/2; m/2; Z) \quad \text{if} \quad 0 \leqq Z < 1 \\ &= 1/2 \quad \text{if} \quad Z = 1 \\ &= (m-2)\,m^{-1} Z^{-1}\,{}_2F_1\,(1, 2-m/2; 1+m/2; Z^{-1}) \quad \text{if} \quad Z > 1\,. \end{aligned} \tag{2.8}$$

A close look at (2.8) shows that it is valid for $m \geqq 2$ but not so far $m = 1$. We suspect that the stipulated $W$ does not exist for $m = 1$. Hence, the reader should note from the beginning that in all our discussions $m$ is subject to the natural restriction that the stipulated $W$ exists, although this may not be stated explicitly. It can be easily seen that $\hat{\mu}_a$ is unbiased for $\mu$ and that

$$\mathsf{V}(\hat{\mu}_a) = \eta_1\,[1 - \mathsf{E}\,(2aW - a^2 W^2/\gamma)]\,. \tag{2.9}$$

Let

$$W_* = W/\gamma\,; \quad \Psi(\gamma) = \mathsf{E}\,W_*^2 \tag{2.10}$$

Note that since $W$ is unbiased for $\gamma$, we have,

$$\mathsf{E}\,W_* = 1 \quad \forall \gamma \in (0,1)\,. \tag{2.11}$$

Then (2.9) can be written as

$$\mathsf{V}(\hat{\mu}_a) = \eta_1\,[1 - \gamma a\,\{2 - a\Psi(\gamma)\}]\,. \tag{2.12}$$

Let,

$$\nu = \operatorname{Sup}\,\{\Psi(\gamma) : 0 < \gamma < 1\}\,; \quad A = 2/\nu\,. \tag{2.13}$$

Note that variance of $W$ is positive for every $\gamma \in (0, 1)$ and hence using (2.10) and (2.11), we have

$$\Psi(\gamma) = \mathsf{E}\,W_*^2 > (\mathsf{E}\,W_*)^2 = 1 \quad \forall \gamma \in (0,1)\,. \tag{2.14}$$

Next observe that $Z$ can be expressed as

$$Z = V\,(1-\gamma)/\gamma \tag{2.15}$$

where

$$V = (S_2/\eta_2)/(S_1/\eta_1) \sim F_{m,m} \tag{2.16}$$

and that for $m \geqq 2$, $W$ as a function of $Z$ is (i) bounded above in absolute value by 1 (ii) continuous every where except on the null event $[Z = 1]$ and (iii) satisfies: $W\,(0+) = 1$. Hence, using LEBESGUE Dominated Convergence Theorem, note also that $\Psi(\gamma)$ is continuous for every $\gamma \in (0, 1)$ and satisfies: $\Psi\,(1-) = 1$. It then follows that for $m \geqq 2$, we have.

$$\Psi(\gamma) \leqq \nu \tag{2.17}$$

with strict inequality for every $\gamma \in (\gamma_0, 1)$ for some $\gamma_0 \in (0, 1)$. Hence (2.12) implies

**Theorem 2.1.** *A necessary condition for $\hat{\mu}_a$ to be better than $X_1$ is that $a \in (0, A]$. The same condition is both necessary and sufficient, provided $m \geqq 2$. Furthermore,*

$$V(\hat{\mu}_a) \leqq \eta_1 [1 - \gamma a (2 - a\nu)] .$$

It is easy to see that for $m \geqq 2$, $W$ as a function of $Z$ satisfies the equation:

$$W(Z^{-1}) = 1 - W(Z) \quad \forall Z > 0 . \tag{2.18}$$

In view of this, theorem 2.1 implies:

**Theorem 2.2.** *A necessary condition for $\hat{\mu}_1$ to be simultaneously better than both $X_1$ and $X_2$ is that $A \geqq 1$. The same condition is both necessary and sufficient, provided $m \geqq 2$. Furthermore*

$$V(\hat{\mu}_1) \leqq \text{Min}\,\{\eta_1 [1 - \gamma (2 - \nu)],\ \eta_2 [1 - (1 - \gamma)(2 - \nu)]\}$$

**Remark 2.1.** It will be seen in the next section that the necessary conditions of Theorem 2.1–2.2 fail to hold unless $m \geqq 5$. Since these conditions are also sufficient for $m \geqq 2$, these are, in fact both necessary and sufficient for any $m$.

## 3. Optimality properties of the proposed estimators

In view of Theorems 2.1–2.2 and Remark 2.1, the problem of investigating the desired optimality properties of the proposed estimators reduces to that of evaluation of $\nu$. This turns out to be a very intricate job except when $m \leqq 4$. Lemmas 3.2–3.3 which we shall shortly prove, would be our main tools. We first prove an inequality which we shall need for proving Lemma 3.2.

**Lemma 3.1.** *For $x \in (0, 1)$, let*

$$F_i(x) = {}_2F_1\,(1, \beta + i; \lambda + i; x), \quad i = 0, 1 , \tag{3.1}$$

*Assume that*

$$\beta < 0; \quad \lambda > 1 . \tag{3.2}$$

*Then*

$$(\lambda - 1)/\lambda < F_0(x)/F_1(x) < 1 \quad \forall x \in (0, 1) \tag{3.3}$$

Proof. Let $T$ be a random variable such that

$$T \sim \text{Beta}\,(1, \lambda - 1) . \tag{3.4}$$

Let,

$$f_x(t) = (1 - tx)^{-\beta}; \quad g_x(t) = (1 - t)/(1 - tx); \quad h(t) = 1 - t . \tag{3.5}$$

Then in view of (2.1), (2.4) and the assumption on $\lambda$ in (3.2), we can rewrite (3.3) as

$$1 < \mathsf{E} f_x(T)/\mathsf{E}[f_x(T)\, g_x(T)] < 1/\mathsf{E} h(T) \quad \forall x \in (0, 1) . \tag{3.6}$$

Note that, we have

$$0 < f_x(t) < \infty; \quad h(t) < g_x(t) < 1, \quad \forall (x, t) \in (0, 1) \times (0, 1). \tag{3.7}$$

Hence,

$$\mathsf{E}[f_x(T)\, h(T)] < \mathsf{E}[f_x(T)\, g_x(T)] < \mathsf{E} f_x(T) \tag{3.8}$$

Note also that, in view of the assumption on $\beta$ in (3.2), $h(t)$ and $f_x(t)$, $x \in (0, 1)$, are all strictly monotonic decreasing functions of $t$ for $t \in (0, 1)$. Hence using a well-known inequality [see e.g. Remark 2.1 on page 131 in BHATTACHARYA (1984)], we have

$$\mathsf{E}[f_x(T)\, h(T)] > \mathsf{E} f_x(T)\, \mathsf{E} h(T) \quad \forall x \in (0, 1). \tag{3.9}$$

(3.8) and (3.9) imply (3.6), which is equivalent to (3.3).

We now prove the two main Lemmas:

**Lemma 3.2.** *Assume that* $m \geqq 5$. *Let*

$$W_i(Z) = (a_i + bZ)^{-1}; \quad i = 1, 2 \tag{3.10}$$

*where*

$$a_1 = 1; \quad a_2 = m\,(m-4)/(m^2-4); \quad b = m/(m-2) \tag{3.11}$$

*Then*

$$W_1(Z) < W(Z) < W_2(Z) \quad \forall Z > 0. \tag{3.12}$$

Proof. Let

$$f_i(Z) = (a_i + bZ)\, W(Z) - 1, \quad i = 1, 2 \tag{3.13}$$

From (3.11), note that $a_1$, $a_2$, $b$ are all positive and finite in view of the assumption on $m$. Hence, (3.12) is equivalent to

$$f_1(Z) > 0 \quad \forall Z > 0; \quad f_2(Z) < 0 \quad \forall Z > 0 \tag{3.14}$$

From (2.8), (3.11) and (3.13) it is easily seen that (3.14) holds for $Z = 1$.

For $Z \in (0,1)$, define

$$T_i(Z) = {}_2F_1\,(1, -m/2 - 1 + i; m/2 - 2 + i; Z), \quad i = 1, 2, 3, 4 \tag{3.15}$$

Then from (2.8) we have

$$\begin{aligned} W(Z) &= T_2(Z) \quad \text{if} \quad 0 < Z < 1 \\ &= (Y/b)\, T_3(Y) \quad \text{if} \quad Z > 1 \end{aligned} \tag{3.16}$$

where $Y = 1/Z$. Note that in view of (2.2) and (2.3) $T_i s$ satisfy:

$$T_1(Z) + bZT_2(Z) = 1 \quad \forall Z \in (0,1) \tag{3.17}$$

$$T_3(Y) + (a_2/b)\, YT_4(Y) = 1 \quad \forall Z > 1. \tag{3.18}$$

Using (3.16), (3.17) and (3.18), we can rewrite (3.13) as

$$\begin{aligned} f_i(Z) &= T_2(Z)\,[a_i - T_1(Z)/T_2(Z)] \quad \text{if} \quad 0 < Z < 1 \\ &= T_4(Y)\,(Ya_i/b)\,[T_3(Y)/T_4(Y) - a_2/a_i] \quad \text{if} \quad Z > 1 \end{aligned} \tag{3.19}$$

Note that in view of lemma 3.1, we have

$$(m-4)/(m-2) < T_1(Z)/T_2(Z) < 1 \quad \forall Z \in (0,1) \tag{3.20}$$

$$m/(m+2) < T_3(Y)/T_4(Y) < 1 \quad \forall Z > 1 \tag{3.21}$$

Note also that, in view of (2.1) and (2.4), we have

$$T_2(Z) > 0 \quad \forall Z \in (0,1); \quad T_4(Y) > 0 \quad \forall Z > 1 \tag{3.22}$$

Then (3.14) .follows from (3.19) in view of (3.11) and (3.20) – (3.22).

**Lemma 3.3.** *Assume that $m \geqq 2$, then $\Psi(1-\gamma) < \Psi(\gamma)\ \forall \gamma \in (0, 1/2)$.*

Proof. Let $W_* = g(V, \gamma)$ be the expression of $W_*$ in terms of $V$ and $\gamma$. From (2.10) and (2.15) note that

$$g(V, \gamma) = W\left(V(1-\gamma)/\gamma\right)/\gamma \quad \forall \gamma \in (0, 1) \tag{3.23}$$

where $W(.)$ represent $W$ as a function of $Z$ as defined in (2.8); furthermore,

$$\mathsf{E} g^2(V, \gamma) = \Psi(\gamma) \quad \forall \gamma \in (0,1) \tag{3.24}$$

Using (2.15) and (3.23) we see that

$$W(Z^{-1})/(1-\gamma) = g(V^{-1}, 1-\gamma) \tag{3.25}$$

From (2.16) note that $V$ and $V^{-1}$ have identical distribution which does not depend on $\gamma$. Hence, (3.24) and (3.25) imply

$$\mathsf{E} W^2(Z^{-1})/(1-\gamma)^2 = \Psi(1-\gamma) \tag{3.26}$$

Squaring both sides of (2.18), then taking expectation and using (3.26) along with (2.10) and (2.11), we get

$$(1-\gamma)^2\, \Psi(1-\gamma) = \gamma^2 \Psi(\gamma) - 2\gamma + 1 \quad \forall \gamma \in (0, 1) \tag{3.27}$$

The desired result follows easily from (2.14) and (3.27).

Let

$$\delta = \sup\{\Psi(\gamma) : 0 < \gamma < 1/2\} \tag{3.28}$$

Then lemma 3.3 implies

$$\nu = \delta\ . \tag{3.29}$$

We now begin to evaluate $\nu$. We first prove

**Lemma 3.4.** *$\nu = \infty$ if $m \leqq 4$.*

Proof: Observe that $(S_1, S_2)$ can be regarded as a complete sufficient statistic for $(\eta_1, \eta_2)$ arising from 2 m observations $S_{ij}$, $i = 1, 2$, $j = 1, \ldots, m$ such that $S_{ij}/\eta_i$ are i.i.d. chi-square variables with 1 degree of freedom. Then, using Rao-Blackwell theorem, it is easy to see that $\Psi(\gamma)$ is non-increasing in $m$. Hence, it suffices to prove the lemma for $m = 4$ only. From (2.10), (2.8) and (2.15) it is easy to see that

$$\lim_{\gamma \to 0+} W_* = (m-2)\, m^{-1} V^{-1} \quad \text{a.s.} \tag{3.30}$$

From (2.16) note that $\mathsf{E}V^{-2}=\infty$ if $m=4$. Hence using FATAU's Lemma the desired result is obvious for $m=4$.

Lemma 3.4 implies that $A=0$ when $m\leqq 4$. In view of theorem 2.1 we have thus proved.

**Theorem 3.1.** *None in the family of estimators $\{\hat{\mu}_a\}$ is better than $X_1$ when $m\leqq 4$.*

In view of theorem 3.1, we now assume that $m\geqq 5$. Let $W_i(i=1, 2)$ be as defined in (3.10). Let $W_{i*}$, $\Psi_i(.)$, $\nu_i$, $\delta_i$ be related to $W_i$ $(i=1, 2)$ in the same way as $W_*$, $\Psi(.)$, $\nu$ and $\delta$ are related to $W$ [vide (2.10), (2.13) and (3.28)]. Then using Lemma 3.2, we have

$$\nu_1\leqq\nu; \quad \delta\leqq\delta_2 \tag{3.31}$$

Combining (3.28) and (3.31) we get

$$\nu_1\leqq\nu\leqq\delta_2 . \tag{3.32}$$

In view of (2.10) and (2.15), (3.10) gives

$$W_{i*}^2=[\gamma a_i+(1-\gamma)\, bV]^{-2}, \quad \gamma\in(0, 1), \quad i=1, 2 , \tag{3.33}$$

where $a_1, a_2, b$ are as given in (3.11). It is easy to see that these are convex functions of $\gamma$. It then follows that $\Psi_i(.)$ $(i=1, 2)$ are convex. Hence,

$$\nu_1=\text{Max}\,\{\Psi_1(0), \Psi_1(1)\}; \quad \delta_2=\text{Max}\{\Psi_2(0), \Psi_2(1/2)\} \tag{3.34}$$

From (3.11) and (2.16) note that

$$\mathsf{E}V^{-1}=b; \quad \mathsf{E}V^{-2}=b^2/a_2; \quad 1/a_2>1 \tag{3.35}$$

(3.33) and (3.35) readily give

$$\Psi_1(1)=1 ; \quad \Psi_1(0)=\Psi_2(0)=b^{-2}\mathsf{E}V^{-2}=1/a_2>1 \tag{3.36}$$

Using arithmetic-geometric mean inequality, (3.33) implies

$$W_{2*}^2\leqq 1/(a_2bV) \quad \text{when} \quad \gamma=1/2 \tag{3.37}$$

(3.37) and (3.35) then yield,

$$\Psi_2(1/2)\leqq a_2^{-1}b^{-1}\mathsf{E}V^{-1}=1/a_2 \tag{3.38}$$

From (3.34), (3.36) and (3.38) it follows that

$$\nu_1=\delta_2=1/a_2 \tag{3.39}$$

Recalling the value of $a_2$ from (3.11), (3.32) and (3.39) imply

$$\nu=1/a_2=(m^2-4)/[m\,(m-4)] . \tag{3.40}$$

Therefore,

$$A=2m\,(m-4)/(m^2-4) . \tag{3.41}$$

In view of Theorem 2.1 we have thus proved

**Theorem 3.2.** *Assume that $m\geqq 5$. Then $\hat{\mu}_a$ is better than $X_1$ iff $a\in(0,2\,m\,(m-4)/$*

$(m^2-4)]$. *Furthermore,*

$$\mathsf{V}(\hat{\mu}_a) \leqq \eta_1 \left[1-\gamma a \left\{2-a\,(m^2-4)\,m^{-1}\,(m-4)^{-1}\right\}\right]. \tag{3.42}$$

**Remark 3.1.** *The estimator* $\hat{\mu}_a$ *with 'a' given by*

$a=(n-5)^2/(n-1)^2$ *for n odd i.e.* $(m-4)^2/m^2$ *for m even*
$=(n-4)\,(n-6)/[n\,(n-2)]$ *for n even i.e.* $(m-3)\,(m-5)/(m^2-1)$ *for m odd*

was proposed by COHEN and SACKROWITZ for improvement over $X_1$. An application of our theorem 3.2 shows that the estimator has the desired property if and only if $m \geqq 6$ i.e. $n \geqq 7$. This follows from the COHEN-SACKROWITZ result also noting that they are obviously mistaken in their claim that the estimator has the desired property for $n=6$ i.e. $m=5$. The upper bound for the variance of this estimator as derived by them is : $\eta_1\,(1-\gamma a)$. It is easily seen that for the particular value of '$a$' under consideration the expression within curly brackets in (3.42) exceeds 1 and hence (3.42) constitutes an improvement over the COHEN-SACKROWITZ bound.

From (3.41) we see that $A \geqq 1$ if and only if $m \geqq 8$. In view of theorem 2.2 we have thus proved

**Theorem 3.3.** $\hat{\mu}_1$ *is simultaneously better than both* $X_1$ *and* $X_2$ *iff* $m \geqq 8$. *Furthermore,*

$$\mathsf{V}(\hat{\mu}_1) \leqq \operatorname{Min}\left\{\eta_1\left[1-\gamma\,(m^2-8m+4)\,m^{-1}\,(m-4)^{-1}\right],\ \eta_2\left[1-(1-\gamma)\,(m^2-8m+4)\,m^{-1}\,(m-4)^{-1}\right]\right\} \tag{3.43}$$

**Remark 3.2.** Theorem 3.3 may be contrasted with the COHEN-SACKROWITZ result which implies the property of $\hat{\mu}_1$ in question only when $n \geqq 15$ i.e. $m \geqq 14$. It may also be contrasted with other estimators known to have such a property. All of these have the general form: $\hat{\mu}_{**}=X_1+(X_2-X_1)/(1+cZ)$ where $c$ is a constant to be suitably chosen and all require $m \geqq 10$ [see KHATRI and SHAH (1974)].

### References

[1] BHATTACHARYA, C. G. (1984). Two inequalities with an application. *Ann. Inst. Stat. Math.* **36A**, 129–134.

[2] COHEN, A. and SACKROWITZ, H. B. (1974). On estimating the common mean of two normal distributions. *Ann. Statist.* **2**, 1274–1282.

[3] COHEN, A. and SACKROWITZ, H. B. (1976). Correction to 'On estimating the common mean of two normal distributions. *Ann. Statist.* **4**, 1294.

[4] GRAYBILL, F. A. and DEAL, R. B. (1959). Combining unbiased estimators. *Biometrics* **15**, 543–550.

[5] KHATRI, C. G. and SHAH, K. R. (1974). Estimation of location parameters from two linear models. *Comm. Statist.* **3**, 647–663.
[6] LEBEDEV, N. N. (1972). *Special functions and their applications.* Rev. enl. ed. Translated and edited by RICHARD A. SILVERMAN, New York, Dover Publications.

Received January 1986; revised November 1986.

C. G. BHATTACHARYA
Indian Statistical Institute
7, S. J. S. Sansanwal Marg
New Delhi - 110016, India