APPLICATION OF BAYESIAN ESTIMATION IN SOCIAL SCIENCE RESEARCH[1]

by

Mariano E. Garcia[2]

1. Introduction

1.1 *The Problem* – Consider the problem of estimating the mean of a normal distribution. That is, let $\tilde{x}_1, \tilde{x}_2, \ldots \tilde{x}_n$ be a random sample of size n from the normal distribution

$$f(x; \mu, \sigma^2) = \frac{1}{2C} \; exp\left\{ \frac{-1}{2\sigma^2} \; (x - \mu)^2 \right\} \quad (1.1.1)$$

where μ is unknown and σ^2 is known. It is desired to find the best estimator for μ.

The most popular method of dealing with this estimation problem is the maximum likelihood method where μ is estimated by the sample mean $\bar{\tilde{x}}$. This estimator possesses certain desirable properties, and one of them is the property of unbiasedness.

There are, however other estimators of μ, which are not necessarily unbiased but may perform better than $\bar{\tilde{x}}$ in certain situations. For instance, there are occasions where there is a prior information on the unknown parameter. This prior information may be in the form of an initial guessed value, say μ_o or in the form of a prior probability distribution of $\tilde{\mu}$, where $\tilde{\mu}$ is now treated as a random variable.

To this class of estimators belong the shrinkage, Bayesian and compromise estimators discussed in Sections 2.1, 2.2 and 2.3, respectively.

---

1. Paper presented in The First National Convention on Statistics held at the PICC, Dec. 4-5, 1978.

2. Assistant Professor of Statistics, U.P. Statistical Center

1.2 *Basis of comparison* – Basically, these estimators are not comparable because of the difference in underlying concepts. In the classical approach to statistical inference, an estimator is considered desirable if it possesses certain properties such as unbiasedness, efficiency, sufficiency, etc. It is apparent that no single index of "desirability" can be devised to evaluate the superiority of one estimator over the other.

An attempt can be made to make these estimators comparable by using the decision-theoretic approach to the estimation problem. For the particular case of estimating the mean of the normal distribution with known variance, the estimation problem can be formulated from a decision-theoretic viewpoint as follows:

Given (Ω A, L) and a random sample $\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_n$ from the normal distribution given in equation (1.1.1), what estimator or decision function d should the statistician use? Here,

Ω = parameter space or set of all possible states of nature, $\{\mu: -\infty < \mu < \infty\}$

A = action space, set of all actions available to the statistician, $\{\hat{\mu}: -\infty < \hat{\mu} < \infty\}$

L = loss function defined on $A \times \Omega$, where $L(\mu, \hat{\mu}) = (\hat{\mu}\ \mu)^2$

d = a function of the random sample, where the estimator $\tilde{\mu} = d(\tilde{x}_1, \tilde{x}_2,, \ldots, \tilde{x}_n)$.

Given an estimator d, its risk is defined as the expected value of the loss function, that is,

$$R(d, \mu) = E[d(\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_n) - \mu]^2 \qquad (1.2.1)$$
$$= E[\tilde{\hat{\mu}} - \mu]^2$$

assuming a quadratic loss function. This expression is known as the mean squared error or *MSE.* Karlin points out the fact that demanding an estimator to be unbiased in all practical situations is an unjustified restriction and is incorrect from the point of view of admissibility.

1.3 *Assumptions and Notations* – Throughout this paper, unless otherwise explicitly stated, normal distribution with known variance and squared error loss function is assumed.

Although the results presented here are those for the case of known variance, it is speculated that similar results and conclusions will be obtained for the case of unknown variance. This is based on the results of Thompson [13] and Arnold and Al-Bayyatti [1]. The case of unknown variance is not thoroughly explored in statistical literature, owing to the complexity of the risk function especially for the Bayesian case where the joint prior distribution of μ and σ^2 does not have a simple form.

Only for mathematical convenience that in most cases, it is assumed that $\mu_o = 0$ and $\sigma^2 = 1$.

The notations employed in this paper are mostly adapted from Lavalle [8] and Raiffa and Schlaifer [11]. Here, random variables are shown with a tilde (~) to distinguish them from values assumed by the random variable. This notational convenience is especially necessary for the Bayesian case where a parameter can be considered a random variable.

The notations for the risk functions are essentially those of Ferguson [3] where estimators are considered decision functions. For shrinkage estimators, the risk function used is $R(\mu, d)$ which represents the risk of d at μ. In this paper, it is proposed to compare the different estimators using the risk $R(\mu, d)$ or *MSE* as the main criterion.

It is obvious that the *MSE* of the maximum likelihood estimator $\tilde{\bar{x}}$ is equal to its variance and does not depend on μ. The sample mean $\tilde{\bar{x}}$ was shown to be admissible by Hodges and Lehman [5] /and Girschick and Savage [4]. However, it does not necessarily dominate all other estimators in terms of minimum risk, for all $\mu \in O$

The concepts of dominance and admissibility are defined as follows:

An estimator d is said to dominate another estimator d^* if

$R(\mu, d) \leqslant R(\mu, d^*)$ for all $\mu \in \Omega$, and

$R(\mu, d) < R(\mu, d^*)$ for at least one $\mu \in \Omega$.

An estimator d is said to be admissible if there exists no other estimator that dominates it. An estimator is said to be inadmissible if it is not admissible.

It is not only the sample mean which has the property of admissibility. In general, any contraction of $\tilde{\bar{x}}$, say $k\tilde{\bar{x}}$ where k is a constant, $o < k \leqslant_1$ is admissible. This fact was shown by Karlin [6]. The estimator $k\tilde{\bar{x}}$ surely is biased, but is admissible just like $\tilde{\bar{x}}$. Furthermore, $k\tilde{\bar{x}}$ may have a smaller *MSE* than $\tilde{\bar{x}}$ in some interval in Ω though it may have a higher *MSE* outside this interval. This suggests that we can use the biased estimator $k\tilde{\bar{x}}$ in some practical situations. For Bayesian estimates, the risk function used is $R\ (\pi, d)$ which represents the risk of d at the prior distribution π.

In deriving the risk function for the different estimators, the details are ommitted but is contained in the original paper [15].

2. Theory

2.1 *Shrinkage Estimators* – These estimators are applicable in situations where there is a prior knowledge in the form of an initial guessed value μ_o. The presumption is that the researcher, because of his experience and acquaintance with the subject matter, has come up with a single value as his initial guess of the parameter μ . In most cases, such aspects as experience are not translatable into numbers in a direct way and there is no point in asking the experimenter how he came up with that value.

Suppose the researcher knows that the true value μ is near a certain guessed value μ_o. Then he can notify $\tilde{\bar{x}}$ by moving it closer to μ_o. This is done by multiplying it by a shrinkage factor k. This is the process of modifying the estimator "shrinkage" and the resulting estimate "shrinkage estimate". In a similar manner, the estimate $\bar{x}$ can be moved farther from μ_o in which case it is called "expander".

If μ_o happens to be close to μ, then the *MSE* of the shrinkage estimator will be lower than the *MSE* of the unbiased estimator $\tilde{\bar{x}}$, for the interval where μ_o is close to μ. On the other hand, if μ_o happens to be far from μ, the *MSE* of the shrinkage estimator will be higher than that of the unbiased estimator, for the interval where μ_o is far from μ. Hence, the shrinkage process actually buys increased efficiency for μ near μ_o at the expense of lower efficiency of μ far from μ_o.

Shrinkage estimates for the mean of the normal distribution were proposed by Thompson [13] and Mehta and Shrinioisan [10]. The estimator suggested by Thompson is of the form

$$d_{21}\ (x_1, \ldots\ldots x_n) = k\ (\bar{x} - \mu_o) + \mu_o$$

for the case of known variance and k specified by the experimenter in advance. For the case of unknown variance, it is given by

$$d_{31}(x_1, \ldots, x_n) = \frac{(\bar{x} - \mu_o)^3}{(\bar{x} - \mu_o)^2 + \frac{s^2}{n}} + \mu_o \quad (2.1.1.)$$

The estimate proposed by Mehta and Shrinioisan is given by

$$d_{41}(x_1, \ldots, x_n) = \bar{x} - a(\bar{x} - \mu_o) \,[\, exp \frac{b(\bar{x} - \mu_o)^2}{s^2/n} \,], \quad (2.1.2)$$

where a and b are constants specified by the experimenter. The admissibility of these estimators were examined by Strawdermann and Cohen [12].

The general behavior of these estimators is to perform well (in terms of minimum risk) if μ_o is close to μ but to behave badly if μ_o is far from μ. To overcome this limitation. a two-stage process is proposed whereby the assessment of the initial guessed value is made in the first stage and the information obtained in both stages will be pooled together if the result of the initial assessment turns out to be unsatisfactory.

The proposed two-stage estimation scheme is as follows:

First, it is assumed that there is a prior knowledge of the mean μ in the form of an initial guessed value μ_o. The degree of belief on μ_o can be expressed by a constant k, $o < k \leqslant 1$. For example, k near zero implies a strong belief on μ_o while k near 1 implies a poor knowledge of μ_o. It is also assumed that there are two samples at hand; or there is only a first sample available, but the second sample can be produced either from the already collected data or by performing a new experiment.

With these initial assumptions satisfied, the experimenter carries out the actual two-stage procedure as follows:

1. Select two positive integers n_1 and n_2.
2. Choose a region S (depending on μ_o) in the real axis.
3. Obtain independent random observations $\tilde{x}_{11}, \tilde{x}_{12}, \ldots, \tilde{x}_{1n}$, and compute the sample mean $\bar{x}_1$.
4. If $\bar{x}_1 \in S$ take $k(\bar{x}_1 - \mu_o) + \mu_o$, the shrinkage estimate, as the estimate of μ.
5. If $\bar{x}_1 \notin S$, take a second set of observations $\tilde{x}_{21}, \tilde{x}_{22}, \ldots$. $\tilde{x}_{2n_2}$ and take the weighted mean

$$\frac{n_1 \bar{x}_1 + n_2 \bar{x}_2}{n_1 + n_2}$$

as the estimate of μ.

The case $k = 1$ was first studied by Katti [7] and then extended to the multivariate case by Waikar and Katti [14]. Their estimator is given by

$$d_{12}(x_{11}, x_{12}, \ldots, x_{2n_2}) = \begin{cases} \bar{x}_1 & \text{if } \bar{x}_1 \in S \\ \dfrac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2} & \text{if } \bar{x}_1 \notin S \end{cases} \quad (2.1.3)$$

$$\text{where } S = (\mu_o - c\,\sigma, \mu_o + c\,\sigma) \quad (2.1.4)$$

$$c = \sqrt{\frac{1}{n_1} \quad \frac{1}{2+u}} \quad (2.1.5)$$

$$u = n_2/n_1 \quad (2.1.6)$$

The case for arbitrary k was first proposed by Arnold and Al-Bayyatti and is given by

$$d_{22}(x_{11}, x_{12} \ldots, x_{2n_2}) = \begin{cases} k(\bar{x}_1 - \mu_o) + \mu_o & \text{if } \bar{x}_1 \in S \\ \dfrac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2} & \text{if } x_1 \notin S \end{cases} \quad (2.1.7)$$

$$\text{where } S = (\mu_o - c\,\sigma, \mu_o + c\,\sigma)$$

$$c = \sqrt{\frac{1}{n_1} \quad \frac{u}{(1+u)^2\; k^2 - 1}} \quad (2.1.8)$$

u and k are as defined above. It is suggested that for the shrinkage factor k, the one used by Thompson can be substituted, thus the first stage estimate may be taken as the Thompson estimator in Equation 2.1.1.

2.2 *Bayesian Estimates* – Suppose such aspects as experience and knowledge can be evaluated numerically from recorded data and can be quantified into a probability distribution. Or suppose that the prior information can be represented by a probability distribution, the form and parameters of which are entirely based on the personal judgment of the experimenter. Then the parameter of interest may now be considered as a random variable with a probability distribution whose mean is μ_o.

Bayes' theorem for density functions can be used to combine this prior information with the information contained in the sample to produce a posterior probability density function which expresses the experimenter's degree of belief after the performance of the experiment.

When there is substantial information about the parameters μ, it is convenient to take as prior distribution the natural conjugate prior given by

$$\pi(\mu) = \frac{n_o}{\sqrt{2\pi\sigma}} \exp.\left\{-\frac{no}{2\sigma^2}(\mu - \mu_o)^2\right\} \tag{2.2.1}$$

where $n_o = \dfrac{\sigma^2}{\sigma_o^2}$

which is a normal density with mean μ_o and variance σ^2/no. If n_o is large for fixed σ^2, the variance of the prior distribution approaches zero implying that this prior information about the mean is very precise and approaches the point estimate μ_o. If n_o is near zero for fixed σ^2, the prior variance becomes infinite which implies a very imprecise knowledge about μ_o.

On the other hand, if there is no prior knowledge about μ, it is convenient to use a prior distribution which has constant density for all μ, that is, any value of μ is as likely as the other. In this case, it is suggested by Lindley [9] to use the vague prior with density written as

$$\pi(\mu)\, d\mu \;\alpha\; d\mu$$

The point estimation problem will now be modified to include the assumption that $\widetilde{\mu}$ is a random variable with prior density $\pi(\mu)$. Similarly, the density function of the sample observation will now be conditional on a given value of μ, that is, we write as

$$f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left\{\frac{-1}{2\sigma^2} (x-\mu)^2\right\} \tag{2.2.2}$$

The *MSE* at a given value μ of $\tilde{\mu}$ is now obtained by taking expectation relative to the conditional distribution of $\tilde{x}_1, \tilde{x}_2 \ldots, \tilde{x}_n$ given μ. Since $\tilde{\mu}$ is now a random variable, we define the risk at a given prior density of $\tilde{\mu}$, i.e.

$$R(\pi, d) = E\{E[(d(x_1, \ldots x_n) - \mu)^2]\} \tag{2.2.3}$$

where the outer expectation is with respect to the prior density. This expression is called the Bayes' risk. It can be shown (see Lavalle [8]) that the Bayes' estimator which minimizes $R(\pi, d)$ is the mean of the posterior distribution. The following results for the normal distribution can be verified:

(a) Let the single observation $\tilde{x}$ be normal with mean μ and variance σ^2, where μ is unknown and σ^2 is known. Let the natural conjugate prior of $\tilde{\mu}$ be also normal with mean μ_o and variance σ^2/n_o then the posterior distribution of $\tilde{\mu}$ given x is also normal with mean

$$\frac{n_o\mu_o + x}{1 + n_o}$$

and variance

$$\frac{\sigma^2}{1 + n_o}$$

The Bayes' estimate is thus

$$d'_{11}(x) = \frac{n_o\mu_o + x}{1 + n_o} \tag{2.2.4}$$

and the Bayes' risk is

$$R(\pi, d_1) = \frac{\sigma^2}{1 + n_o} \tag{2.2.5}$$

(b) If instead of a single observation x, we have a sample x_1, $x_2, \ldots, x_n$, then the Bayes' estimate is given by

$$d'_{21}(x_1, \ldots, x_n) = \frac{n_o \mu_o + n\bar{x}}{n_o + n} \tag{2.2.6}$$

with Bayes' risk

$$R(\pi_1, d_{21}) = \frac{\sigma^2}{n_o + n} \tag{2.2.7}$$

(c) If we have a random sample from the normal population with mean μ and variance σ^2 and the prior density of μ is flat, then the Bayes' estimate is

$$d'_{31}(x_1, \ldots x_n) = \bar{x} \tag{2.2.8}$$

with Bayes' risk

$$R(\pi_2, d_{31}) = \frac{\sigma^2}{n} \tag{2.2.9}$$

Note that this is exactly the maximum likelihood estimate.

(d) Let the prior distribution be normal with mean μ_o and variance σ^2/n_o. A random sample of size n_1 is taken from the normal population with unknown mean μ and known variance σ^2. Then the posterior distribution is normal with mean

$$\frac{n_o \mu_o + n_1 \bar{x}}{n_o + n_1}$$

and variance

$$\frac{\sigma^2}{n_o + n}$$

If another experiment is conducted using a second random sample of size n_2 and if the prior density for this stage is taken as the posterior of the first stage, then the Bayes' estimate of the two-stage scheme is given by

$$d'_{12}\ (x_{11}, x_{12} \ldots, x_{2n_2}) = \frac{n_o\,\mu_o + n_1\,\overline{x}_1 + n_2\,\overline{x}_2}{n_o + n_1 + n_2} \tag{2.2.10}$$

with Bayes' risk

$$R\ (\pi_1, d'_{22}) = \frac{\sigma^2}{n_o + n_1 + n_2} \tag{2.2.11}$$

2.3 *A Compromise Estimator* – With respect to a given prior density, Bayes' estimators always give the minimum Bayes' risk among all other estimators. There is, however, a bad property of Bayes' estimators mentioned by Efron and Morris [2]. This is the fact that the Bayes' estimator designed to do well versus the population as a whole may do very poorly against sub-populations which the statistician recognizes and cares about but for which he has no distributional information.

To illustrate, suppose the researcher has available information on the average *IQ* of the general population and he uses this as prior density in estimating the mean of a subpopulation (say, children of Nobel Prize winners). If the true mean of the subpopulation is higher than the prior mean of the general population, then the resulting Bayes' estimator will have a high risk.

To overcome this bad property, Efron and Morris proposed a compromise estimator. The proposed estimation rule is to fix some allowable deviation from the maximum likelihood estimate x, say M, and follow the Bayes' rule as closely as possible. For the case $\mu_o = 0$, $\sigma^2 = 1$, $n = 1$, $A = 1/n_o$, the estimate is given by

$$d''_{11}\ (x) = \begin{cases} x + M & \text{if } x < -c \\ \dfrac{A}{A+1}\ x & \text{if } x \in [-c, c] \\ x - M & \text{if } x > c \end{cases} \tag{2.3.1}$$

where $c = M\ (A + 1)$

The risk is given by

$$R(\pi, d''_{11}) = \frac{D^2 + A + 1}{A + 1} - \frac{D^2 + 1}{A + 1} \int_{-D}^{D} \phi(z)\, dz + \frac{2D\phi(D)}{A + 1} \qquad (2.3.2)$$

where $D = \frac{c}{\sqrt{A+1}}$ and ϕ is the normal dènsity function.

If the observable is a random sample, then the estimate is given by,

$$d''_{12}(x_1, \ldots x_n) = \begin{cases} \bar{x} + M \text{ if } \bar{x} < -c \\ \frac{n_o \mu_o + n\bar{x}}{n_o + n} \text{ if } x \in [-c, c] \\ \bar{x} - M \text{ if } \bar{x} > c \end{cases} \qquad (2.3.3)$$

The constant M is usually taken as a multiple of the standard deviation σ.

To evaluate how well the estimator d''_{11}, compares with the Bayes'estimator d'_{11}, Efron and Morris proposed a measure known as relative savings loss, denoted by 1-s, where

$$1 - s = \frac{R(\pi, d''_{11}) - R(\pi, d'_{11})}{R(\pi, d_{11}) - R(\pi, d'_{11})} \qquad (2.3.4)$$

where

$R(\pi, d''_{11})$ = risk of the compromise estimator at the prior π

$R(\pi, d'_{11})$ = risk of the Bayes' estimator at π which is equal to $A/(A + 1)$.

$R(\pi, d_{11})$ = risk of the maximum likelihood estimator which is equal to 1 (since $\sigma^2 = 1$ and $n = 1$ by assumption), regardless of any π.

The relative savings loss can be derived to be equal to

$$1 - s = 2 \quad [(D^2 + 1)(1 - \phi(D)) - D\phi(D)] \qquad (2.3.5)$$

which is only a function of D. Hence, the relative savings loss (that is, how much he will sacrifice the savings in risk by using d''_{11} instead of d'_{11}) is derived in advance and find D where

$$D = M \sqrt{A+1} \tag{2.3.6}$$

For different values of 1-s, Efron and Morris have tabulated the corresponding values of D.

2.4 *Some Comparisons*

An attempt is made to compare the different estimators so far discussed by making use of the *MSE* or $R(\mu, d)$ for the single-stage estimators and the Bayes' risk based on an equivalent sample size for the two-stage estimators.

The *MSE* for the single-stage shrinkage estimator is given by

$$R(\mu, d_{21}) = \frac{k^2 \sigma^2}{n} + (k-1)^2(\mu - \mu_o)^2 \tag{2.4.1}$$

The Bayesian estimate and the shrinkage estimate are equivalent if we let

$$k = \frac{n_1}{n_o + n_1} \tag{2.4.2}$$

since

$$\frac{n_1}{n_o + n_1}(\bar{x}_1 - \mu_o) + \mu_o = \frac{n_1\bar{x}_1 + n_o\mu_o}{n_o + n_1}$$

Hence

$$R(\mu, d'_{21}) = \left(\frac{n_1}{n_o + n_1}\right)^2 \frac{\sigma^2}{n_1} + \left(\frac{n_o}{n_o + n_1}\right)^2 (\mu - \mu_o)^2 \tag{2.4.3}$$

Without loss of generality, assume that $n_1 = 1$, $\sigma^2 = 1$, $\mu_o = 0$ and

$$A = \frac{1}{n_o} \; ; \text{ then}$$

$$R\ (\mu, d_{21}) = k^2 + (k-1)^2\ \mu^2 \tag{2.4.4}$$

$$R\ (\mu, d'_{21}) = \left(\frac{1}{A+1}\right)^2 \ (A^2 + \mu^2) \tag{2.4.5}$$

The compromise estimator can be shown to have *MSE* equal to

$$R\ (\mu, d'_{11}) = 1 + \frac{C^2}{(A+1)^2} + Q\ (\mu) \tag{2.4.6}$$

where

$$Q\ (\mu) = \mu^2 - C^2 - 1 - 2A\ [\phi\ (c-\mu) - \phi\ (-c-\mu)]$$

$$+ \frac{(2A+1)\ (C+2\mu) - 2A}{(A+1)^2}\ \phi\ (c+\mu) \tag{2.4.7}$$

$$+ \frac{(2A+1)\ (c-\mu) + 2\,\mu\, A}{(A+1)^2}\ \phi\ (c-\mu)$$

The MSEs of these three estimators are not directly comparable because they depend on different parameters. The shrinkage estimator is a function of k, the Bayesian estimator is a function of A and the compromise estimator is a function of both A and C.

The two-stage shrinkage estimators can be compared with the Bayesian and maximum likelihood estimates by numerical evaluation of their Bayes' risks and relative savings loss for different values of k and $\mu = n_2/n_1$.

It is noted that the two-stage estimators perform better than the Bayesian estimator and the maximum likelihood estimator only when u and k are small. That is, the prior variance is small or the belief that μ is close to μ_o is strong and the second sample size is more or less equal to the first sample size.

3. Applications of Bayesian Estimation

3.1 *The Educational Placement Test* – It is not uncommon in social science research where the researcher has certain prior information which he wants to use in estimating a certain parameter. The methods so far discussed will be applied specifically to the Educational

Placement Test [16] which is a project of the National Education Center.

In line with the government's efforts to develop and utilize human resources at all age levels, the group of school leavers are encouraged to continue their education, formal or informal or to train for jobs they might have some aptitudes for or interest in.

The educational placement test, therefore has been developed as an instrument to assess knowledge and work experience in various areas of endeavor which will be given credit for academic equivalence. This equivalency may be used for grade/year placement in the formal school system.

The purely academic part of the Educational Placement Test will be administered to the school leavers who like to go back to school and finish a certain level of formal education. This part covers the subject areas of English, Pilipino and Mathematics for the secondary and elementary levels. Based on the rationale that learning is a continuum, the test for each subject area was prepared in such a way that the test items for each subject area cover, in a sequential manner, basic skills required for the grade/year level.

A student who takes the tests can go as far as his acquired skills will allow him to reach the appropriate grade level he can manage. It might happen that a leaver would be placed in different "grades" or "levels" depending upon his own abilities and skills. The leavers will be placed in the highest grade level/year he has attained as shown by the test results. Whatever deficiencies he has with respect to the other subject areas will be taken cared of in the proposed Learning Center.

3.2 *Development of Norms* – Initially, it is desired to determine scores that will classify a particular school leaver to a certain year or grade level. Since these school leavers will eventually be absorbed into the educational system, they will be mixed with those in-school students. Hence, it was decided to administer the same test to students presently in school to determine the norms for each subject area and grade level. The scores were standardized for each subject area for each year level such that the scores have been converted to a normally distributed variable with mean 50 and standard deviation 10.

The norms are basically obtained from the mean and standard deviation and it is therefore a problem to determine which mean to use – whether the in-school or out-of-school person for a given subject or year level. It may be argued that the school leaver because of his maturity, experience and training will eventually have a higher score than his in-school counterpart. It is also possible that many of these out-of-school youths have not tried to keep up intellectually and would really be far behind from those with formal education.

Putting it in the context of statistical estimation, we have two subpopulations namely the in-school and out-of-school persons who will eventually be mixed together in one population. Using either one of the means of these subpopulations as estimate of the general population might produce unwanted results because of the arguments cited above. Combining the two means in such a manner that a reasonable mean is produced would be the best alternative.

One way is to consider the information on the in-school as the prior distribution and the out-of-school as the sample. The resulting posterior mean can serve as a reasonable estimate. Since all the single-stage estimators discussed in the previous section can be given a Bayesian interpretation, it is possible to consider the estimation procedure using these methods.

Another way is to consider the information on the in-school as the prior distribution and the out-of-school who left school at the same year or grade level as the first stage. The second stage would be the scores of all other out-of-school regardless of the year or grade level when they left school.

3.3 *Applications of Bayesian and Shrinkage (single-stage) estimators*
We apply the above methods on the sample of out-of-school examinees coming from Region IV shown in Table 1 in the two subject areas, Mathematics and English.

If the Bayesian estimator using a normal prior with mean 50 and standard deviation 10 (so that $n_o = 1$) is used for the data in Table 1 and 2, it follows that the Bayesian estimates would be the same as the maximum likelihood estimate which is the sample mean $\bar{x}$. This is because the sample sizes are large and the sample mean has a greater weight than the prior mean. The information corresponding to the in-school students would therefore not be reflected here.

The shrinkage (single-stage) estimator can be applied to Table 2 since it is known that the in-school students in Fourth Year have mean 50 and standard deviation 10. Furthermore, as the grade or year level decreases, the true mean would be at a greater distance

TABLE 1

STANDARD SCORES FOR MATHEMATICS EXAM (SAME EXAM FOR ALL GRADE/YEAR LEVELS) TAKEN BY OUT-OF-SCHOOL EXAMINEES

Year Level of Dropping Out	*n*	*Mean*	*s.d.*
Grade V	47	41.23	9.46
VI	235	44.33	9.17
Year I	199	47.71	9.54
II	178	55.69	13.58
III	193	55.94	12.99
IV	96	58.31	13.63

TABLE 2

STANDARD SCORES FOR FOURTH YEAR ENGLISH EXAM (TAKEN BY OUT-OF-SCHOOL EXAMINEES)

Year or Grade Level of Dropping Out	*n*	*Mean*	*s.d.*	*Shrunken Mean*
Grade V	12	35.33	9.31	41.20
VI	155	42.21	9.57	46.10
Year I	188	45.41	10.56	48.16
II	166	51.73	11.90	50.52
III	189	52.29	11.75	50.46
IV	97	53.54	11.87	50.35

than the 4th year mean which is 50. Thus using the estimate $k(\overline{x}_1 - \mu_o) + \mu_o$ with $k = 0.1, 0.2, \ldots 0.6$ for *4th* year, *3rd* year . . . Grade V respectively, we obtain the shrunken mean shown in Table 2.

For the results of the Mathematics exam shown in Table 1, the maximum likelihood estimate would suffice since all students both in-school and out-of-school take the same exam regardless of year level unlike the English exam where there is a separate exam for each grade or year level.

In fact, the above results show that a single exam can be designed such that it will be taken by all students regardless of year levels.

3.4 *Applications of Bayesian and two-stage estimators*

The Bayesian and two-stage estimators are generally applicable to situations where a small sample is available. In fact, it would be ideal in cases where the taking of the sample is costly since the estimation procedure is sequential in nature. Furthermore, a Bayesian viewpoint is that the experimenter changes his belief everytime he acquires new information in the value of the parameter.

We can apply the Bayesian and two-stage estimation procedure to the data on school leavers of age 40 and above, shown in Tables 3 and 4.

In Table 3, the sample mean would be considered as the estimate of the true mean for the same reason as in the previous section.

TABLE 3

STANDARD SCORES OF SCHOOL LEAVERS OF AGE 40 AND ABOVE IN MATH EXAM

Year of Drop-out	*n*	*Mean*	*s.d.*
Year I	7	47.86	11.70
II	12	61.25	14.58
III	9	54.56	8.63
IV	4	50.75	20.76

In Table 4, the Bayesian estimates were computed sequentially, that is, the prior mean of 50 and $n_o = 1$ was used for the fourth year

TABLE 4

STANDARD SCORES OF SCHOOL LEAVERS OF AGE 40 AND ABOVE IN ENGLISH MEAN

Year of Drop-out	*n*	*Mean*	*s.d.*	*Bayesian Estimate*	*Shrinkage Estimate*
I	6	58.33	9.69	57.14	53.33
II	11	59.73	12.84	58.72	52.92
III	8	54.5	10.66	57.42	50.90
IV	4	53.5	11.09	56.90	50.35

to find the posterior. The posterior in fourth year becomes the prior in third year, etc.

The shrinkage estimators were computed by using $k = 0.1, 0.2, 0.3, 0.4$ for the *4th*, *3rd*, *2nd* and *1st* year levels, respectively.

It can be observed that the maximum likelihood estimate, the Bayesian and shrinkage estimators are low for higher year levels and high for lower year levels of drop-outs. It appears that the year level when a particular school leaver of age 40 and above dropped out has no relation with his score in English. Presumably, English proficiency can be acquired even without the benefit of a formal education.

The two-stage estimation scheme proposed by Katti will be illustrated by taking as example the first two rows of Table 4. The region S is found to be equal to (47.915, 52.085) with $n_1 = 6$, $n_2 = 11$. Hence, a second sample is taken and the estimate is taken to be the weighted mean which is equal to 59.24. Here, the two samples can be considered independent since as previously stated, the year level of drop-out does not affect the scores in the English exam, for those who are 40 and over.

Bibliography

Arnold and Al-Bayyatti, "On Double Stage Estimation of the Mean Using Prior Knowledge," Biomatius, 1970, pp. 787-800.

Efron and Morris, "Limiting the Risk of Bayes and Empirical Bayes' Estimators, Part I, The Bayes' Case," Journal of the American Statistical Association, Vol. 66, 1971, pp. 807-815.

Ferguson, T., *Mathematical Statistics* – a Decision Theoritic Approach, Academic Press, New York, 1967.

Girschick and Savage, Bayes and Minimax Estimates for Quadratic Loss Functions, Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, Berkeley and Los Angeles, University of California Press, 1951, pp. 53-73.

Hodges and Lehman. Some Applications of Cramer – Rao Inequality, Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, Berkeley and Los Angeles, University of California Press, 1951, pp. 13-22.

Karlin, S. "Admissibility for Estimation with Quadratic Loss," Annals of Mathematical Statistics, Vol. 29, 1958, p. 406.

Katti, S. "Use of Some a Priori Knowledge in the Estimation of Means from Double Samples," Biometius, Vol. 18, 1962, pp. 139-147.

Lavalle, I. *An Introduction to Probability Decision and Inference,* New York, Holt, Rinehart and Winston, 1972.

Lindley, *Introduction to Probability and Statistics from a Bayesian Viewpoint,* Cambridge University Press, 1965.

Mehta and Shrinioisan. "Estimation of the Mean by Shrinkage to a Point," Journal of the American Statistical Association, Vol. 66, 1971, pp. 86-90.

Raiffa and Schlaifer, "Applied Statistical Decision Theory, Boston, Division of Research, Graduate School of Business Administration," Harvard University, 1964.

Strawdermann and Cohen, "Admissibility of Estimators of the Mean Vector of a Multivariate Normal Distribution with Quadratic Loss," Annals of Mathematical Statistics, Vol. 42, p. 27, 1971.

Thompson, J. "Some Shrinkage Techniques for Estimating the Mean," Journal of the American Statistical Association, Vol. 63, 1968 pp. 113-122.

Waikar and Katti, "On a Two-Stage Estimate of the Mean," Journal of the American Statistical Association, Vol. 66, 1971, pp. 75-81.

Garcia, M. "Estimation of the Mean of the Normal Distribution Using Prior Information." Unpublished Master's Thesis, Statistical Center, U.P.

National Educational Testing Center, Development of Assessment Instruments for Accreditation and Equivalency Purposes.