Sri Lankan Journal of Applied Statistics, Vol.10, 2009, p.137-152

Optimal Combined Estimators for Population Parameters with Known Prior Information

A. LAHEETHARAN AND P. WIJEKOON*

Department of Mathematics and Statistics, University of Jaffna, Sri Lanka

ABSTRACT

Estimation of population parameters by combining the available estimators when some prior information is available, is considered by several statisticians. However these combined estimators for an unknown parameter θ are optimal under some rigid conditions, and hence they are applicable only if these conditions are satisfied. In this paper a more general method for deriving optimal combined estimators for $g(\theta)$, a function of θ, is given. Since the population variance is a function of the mean, these results are useful to obtain optimal combined estimators for both mean and the variance. As an illustration, optimal combined estimators are obtained for the normal distribution, and they are compared with the single estimators existing in the combination using scalar mean square error loss.

Keywords: Coefficient of variation, Mean square error, Minimal sufficient statistic, Optimal combined estimator

INTRODUCTION

The use of prior information in inference is well established in the Bayesian arena of statistical methodology. Such information is usually incorporated into a model by choosing an appropriate prior distribution. Also in a variety of agricultural and biological situations several statisticians have focused their research to find optimal estimators when prior information is available. Assuming the coefficient of variation is known, Marshall (1936) studied the frequency distribution of insect population with special reference to American bollworm. Sen (1978) has considered bird surveys for six provinces in Canada during 1967-1970. Preliminary analysis of the data showed that the coefficient of variation of a particular bird species was steady over the period for each of the provinces. He has further estimated the efficiency of the combined estimate of the population mean μ by assuming that the coefficient of variation is known for each province. Utilization of prior information is also common in quantitative population genetics problems when estimating the intraclass and interclass correlations.

* Department of Statistics and Computer Science, University of Peradeniya, Sri Lanka.

Some basic results concerning optimal estimators are now reviewed. Searls (1964) and Arnholt and Hebert (1995) utilized the known coefficient of variation on estimating the population mean. Wencheko and Wijekoon (2005) improved their results further and derived the optimal shrunken estimators for the mean in one parameter exponential families. A general method to obtain optimal estimators for both mean and variance of distributions when additional information (coefficient of variation, kurtosis etc.) is given in Laheetharan and Wijekoon (2008).

According to their method the optimal estimator is derived based on the complete sufficient statistic for the unknown parameter. If such an estimator does not exit then a uniformly minimum mean squared error estimator can be obtained by using a minimal sufficient statistic defined over a certain class as in the following theorem. In this situation, more than one uniformly minimum mean squared error estimators defined over different classes may exist.

Theorem 1 (Laheetharan and Wijekoon, 2008)

Let $X = (X_1, ..., X_n)'$ be a random sample from a population with distribution $f(x;\theta)$ and $g(\theta)$ be a real-valued function on the parameter space Θ. Let $T_1(X)$ and $T_2(X)$ be minimal sufficient statistics for θ. Let $\phi_1(T_1(X))$ and $\phi_2(T_2(X))$ be estimators of $g(\theta)$ with $E[\phi_i(T_i(X))] = k_i g(\theta)$ where $k_i \in \Re$, and without loss of generality, assume that $k_1, k_2 > 0$. If the ratios $\tau_i^2 = [g(\theta)]^{-2} Var[\phi_i(T_i(X))]$ are independent of θ, $i = 1,2$, then the estimator $\phi_i^*(T_i(X)) = \alpha_i^* \phi_i(T_i(X))$ has uniformly minimum mean squared error (in $g(\theta)$) among all estimators that are in the class $C_{T_i}(\alpha_i) = \{\alpha_i[\phi_i(T_i(X))] \mid 0 < \alpha_i < \infty\}$ respectively, where $\alpha_i^* = k_i/(k_i^2 + \tau_i^2)$. Further, if $k_2 < k_1(\tau_2/\tau_1)$ then $\phi_1^*(T_1(X))$ has smaller mean square error and if $k_2 > k_1(\tau_2/\tau_1)$ then $\phi_2^*(T_2(X))$ has smaller mean square error.

However, it is well known that if more than one estimator is given for a particular parameter, the most suitable one is chosen based on the desirable properties of estimators.

In this regard, the statisticians have also considered the option of combining available estimators. According to this option, Khan (1968) and Gleser and Healy

(1976) have selected an optimal estimator from a set of all linear combinations of two uncorrelated, unbiased estimators of mean θ having a known coefficient of variation. Ester Samuel-Cahn (1994) have further improved these results by taking linear combination of two correlated, unbiased estimators of parameter θ with known correlation coefficient. In 2001 Arnholt and Hebert have obtained an optimal estimator by considering a non convex combination of two correlated and biased estimators for an unknown parameter θ as follows:

If $T_1(X)$ and $T_2(X)$ are two correlated biased estimators for θ, and if $v_i = \theta^{-2} Var_\theta T_i(\mathbf{X})$, $\lambda^2 = k_2^2 v_1 / k_1^2 v_2$, the correlation coefficient ρ of $T_1(X)$ and $T_2(X)$ are independent of θ, and known then the optimal estimator is $T^*(X) = \alpha_1^* T_1(X) + \alpha_2^* T_2(X)$, with $\alpha_1^* = \frac{1-\rho\lambda}{k_1(1-2\rho\lambda+\lambda^2+(1+\rho^2)(v_1/k_1^2))}$, and $\alpha_2^* = \frac{\lambda(\lambda-\rho)}{k_2(1-2\rho\lambda+\lambda^2+(1+\rho^2)(v_1/k_1^2))}$, where k_1 and k_2 are constants. In the derivations of α_1^* and α_2^*, Arnholt and Hebert (2001) mistakenly provide the term $(1+\rho^2)$ in the denominator, and it should be corrected as $(1-\rho^2)$ in both α_1^* and α_2^*.

In all the methods described above, the selection of initial estimators $T_1(X)$ and $T_2(X)$ is arbitrary. In this paper a general method to derive an optimal combined estimator for population parameters, when additional information is available, is considered. This method is primarily based on the theorem 1. As the initial estimators, the minimal sufficient statistics are used, and the method can be applied to any function $g(\theta)$ of the unknown parameter θ.

A MORE GENERALIZED APPROACH FOR COMBINING ESTIMATORS

It is well known that one important way of improving estimation of distribution parameters is by employing biased estimation procedures. In general, since mean squared error (MSE) is a function of the parameter, there will not be one "best" estimator. Often, the MSE's of two estimators will cross each other, showing that each estimator is better than the other with respect to the MSE, in only a portion of e parameter space. This partial information is also sometimes useful since it ovides guidelines to choose a good estimator for a particular situation.

If a complete sufficient statistic exits for an unknown parameter then there exits a uniformly minimum mean square error estimator (Laheetharan and Wijekoon (2008)). However in certain situations a complete sufficient statistic does not exit and hence theorem 1 can be used to obtain a uniformly minimum mean squared error estimator in a certain class. Now we consider the linear combination of two such estimators, and obtain the optimal combined estimator. It can be shown that this optimal combined estimator is better than the two uniformly minimum mean squared error estimators defined over the corresponding classes.

Theorem 2

Suppose $X = (X_1,\ldots,X_n)'$ is a random sample from a population with distribution $f(x;\theta)$ and let $g(\theta)$ be a real-valued function on the parameter space Θ . Suppose $T_1(X)$ and $T_2(X)$ are minimal sufficient statistics for θ. Let $\phi_1(T_1(X))$ and $\phi_2(T_2(X))$ be uncorrelated estimators of $g(\theta)$ with $E[\phi_i(T_i(X))] = k_i g(\theta)$ where $k_i \in \Re$, and without loss of generality, assume that $k_1, k_2 > 0$. If the ratios $\tau_i^2 = [g(\theta)]^{-2} Var[\phi_i(T_i(X))]$ are independent of θ, $i = 1,2$. Then the estimator $\phi^*(T_1(X),T_2(X)) = \alpha_1^*\phi_1(T_1(X)) + \alpha_2^*\phi_2(T_2(X))$ has uniformly minimum mean squared error (in $g(\theta)$) among all estimators that are in the class

$$C_\phi(\alpha_1,\alpha_2) = \{\phi(T_1(X),T_2(X)) = \alpha_1\phi_1(T_1(X)) + \alpha_2\phi_2(T_2(X)) \mid 0 < \alpha_i < \infty, i = 1,2\},$$

where $\alpha_i^* = k_i\tau_j^2 / (k_1^2\tau_2^2 + k_2^2\tau_1^2 + \tau_1^2\tau_2^2)$ for $i, j = 1,2$ and $i \neq j$ are constants.

Proof

Let $\phi^*(T_1(X),T_2(X)) = \alpha_1\phi_1(T_1(X)) + \alpha_2\phi_2(T_2(X))$; $0 < \alpha_1, \alpha_2 < \infty$.

Then $E[\phi^*(T_1(X),T_2(X))] = (\alpha_1 k_1 + \alpha_2 k_2)g(\theta)$ and

$Var[\phi^*(T_1(X),T_2(X))] = \alpha_1^2 Var[\phi_1(T_1(X))] + \alpha_2^2 Var[\phi_2(T_2(X))]$.

The mean square error (MSE) of $\phi^*(T_1(X),T_2(X)) \in C_\phi(\alpha_1,\alpha_2)$ can be obtained as

$$\begin{aligned} MSE[\phi^*(T_1(X),T_2(X))] &= Var[\phi(T_1(X),T_2(X))] + \{E[\phi(T_1(X),T_2(X))] - g(\theta)\}^2 \\ &= \alpha_1^2 Var[\phi_1(T_1(X))] + \alpha_2^2 Var[\phi_2(T_2(X))] + (\alpha_1 k_1 + \alpha_2 k_2 - 1)^2 [g(\theta)]^2. \qquad (1) \end{aligned}$$

Using the standard techniques of partial differentiation

$\partial MSE\left[\phi^*(T_1(X),T_2(X))\right]/\partial\alpha_1 = 2\alpha_1 Var[\phi_1(T_1(X))] + 2k_1(\alpha_1 k_1 + \alpha_2 k_2 - 1)[g(\theta)]^2$ and

$\partial MSE\left[\phi^*(T_1(X),T_2(X))\right]/\partial\alpha_2 = 2\alpha_2 Var[\phi_2(T_2(X))] + 2k_2(\alpha_1 k_1 + \alpha_2 k_2 - 1)[g(\theta)]^2$.

Equating these equations to zero, we get

$\alpha_1\left\{Var[\phi_1(T_1(X))] + k_1^2[g(\theta)]^2\right\} + \alpha_2 k_1 k_2[g(\theta)]^2 - k_1[g(\theta)]^2 = 0$

$\alpha_1 k_1 k_2[g(\theta)]^2 + \alpha_2\left\{Var[\phi_2(T_2(X))] + k_2^2[g(\theta)]^2\right\} - k_2[g(\theta)]^2 = 0$.

Now by solving these equations the results follow,

$\alpha_1^* = k_1\tau_2^2/(k_1^2\tau_2^2 + k_2^2\tau_1^2 + \tau_1^2\tau_2^2)$ and $\alpha_2^* = k_2\tau_1^2/(k_1^2\tau_2^2 + k_2^2\tau_1^2 + \tau_1^2\tau_2^2)$.

By substituting α_1^* and α_2^* to $\phi^*(T_1(X),T_2(X))$ it can be easily shown that

$$\phi^*(T_1(X),T_2(X)) = \frac{k_1\tau_2^2\phi(T_1(X)) + k_2\tau_1^2\phi(T_2(X))}{k_1^2\tau_2^2 + k_2^2\tau_1^2 + \tau_1^2\tau_2^2} \qquad (2)$$

minimizes MSE over $C_\phi(\alpha_1,\alpha_2)$, and the minimum MSE is given by

$$MSE\left[\phi^*(T_1(X),T_2(X))\right] = \tau_1^2\tau_2^2[g(\theta)]^2/(k_1^2\tau_2^2 + k_2^2\tau_1^2 + \tau_1^2\tau_2^2). \qquad (3)$$

Then $MSE\left[\phi^*(T_1(X),T_2(X))\right] = \tau_1^2\tau_2^2[g(\theta)]^2/(k_1^2\tau_2^2 + k_2^2\tau_1^2 + \tau_1^2\tau_2^2)$

$\le \tau_1^2\tau_2^2[g(\theta)]^2/(2k_1k_2\tau_1\tau_2 + \tau_1^2\tau_2^2)$

$= \tau_1\tau_2[g(\theta)]^2/(2k_1k_2 + \tau_1\tau_2) \qquad < \tau_1\tau_2[g(\theta)]^2$;

$\forall\, k_1,k_2 > 0$.

$\le (\tau_1^2 + \tau_2^2)[g(\theta)]^2/2$

$\le (\tau_1^2 + \tau_2^2 + (k_1 + k_2 - 1)^2)[g(\theta)]^2$

$= MSE[\phi_1(T_1(X)) + \phi_2(T_2(X))]$; $\forall g(\theta), \theta \in \Theta$.

Also, note that

$$MSE\left[\phi^*(T_1(X),T_2(X))\right] = \tau_1^2\tau_2^2[g(\theta)]^2/(k_1^2\tau_2^2 + k_2^2\tau_1^2 + \tau_1^2\tau_2^2)$$
$$< \tau_1^2\tau_2^2[g(\theta)]^2/(k_1^2\tau_2^2 + \tau_1^2\tau_2^2)$$
$$= \tau_1^2[g(\theta)]^2/(k_1^2 + \tau_1^2) = MSE\left[\phi_1^*(T_1(X))\right]$$

Thus, $MSE\left[\phi^*(T_1(X),T_2(X))\right] < MSE\left[\phi_i^*(T_i(X))\right]$; for $i = 1,2$, $\forall g(\theta), \theta \in \Theta$. (4)

Therefore it is clear that the optimal estimator $\phi^*(T_1(X),T_2(X))$ dominates the optimal estimators $\phi_i^*(T_i(X))$ in *MSE*, and $\phi_i^*(T_i(X))$'s *inadmissible* for $i = 1,2$, and more generally $\phi^*(T_1(X),T_2(X))$ is uniformly better than any other estimators in the class $C_\phi(\alpha_1,\alpha_2)$.

The above results hold for any probability distribution function which satisfies the given conditions, and applicable to any estimator $\phi(T_1(X),T_2(X))$ if the quantities $\tau_i^2 = [g(\theta)]^{-2} Var[\phi_i(T_i(X))]$ are independent of the unknown parameter θ for $i = 1,2$.

Note

If $g(\theta) = \theta$, $\phi_i(T_i(X)) = T_i(X)$ and $E[T_i(X)] = k_i\theta$ for $i = 1,2$. Then $T_i(X)$ are uncorrelated and biased estimators of θ, and the optimal estimator for θ is $T^*(X) = (k_1\tau_2^2T_1(X) + k_2\tau_1^2T_2(X))/(k_1^2\tau_1^2 + k_2^2\tau_2^2 + \tau_1^2\tau_2^2)$.

Arnholt and Hebert (2001) derived the same estimator when $\rho = 0$ and $\tau_i^2 = \nu_i$ for $i = 1,2$. Also, if $\tau_i^2 = \nu_i$ and $T_i(X)$ are unbiased, then with $k_i = 1$, for $i = 1,2$, the optimal combined estimator is $T^*(X) = (\nu_2T_1(X) + \nu_1T_2(X))/(\nu_1 + \nu_2 + \nu_1\nu_2)$, which was first shown by Gleser and Healy (1976).

APPLICATIONS

In this section the results obtained in section 2 are used to estimate the population mean and variance of normal distribution with known coefficient of variation.

Numerical Illustration - Normal Distribution

Suppose $X_i \sim N(\mu, \sigma^2)$, $\mu \neq 0$, $x \in \Re$ and Mean $E(X_i) = \mu$ and $Var(X_i) = \sigma^2$.

The joint probability density function is

$$P(x_1, \ldots, x_n \mid \mu, \sigma^2) = \exp\left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{n} x_i^2 \quad + \mu \frac{\sum_{i=1}^{n} x_i}{\sigma^2} \quad - \frac{n\mu^2}{2\sigma^2} - \frac{n}{2} \log 2\pi\sigma^2 \right\}. \qquad (5)$$

Since the distribution properties of k parameter exponential families are used for the derivations, now we recall some of those important properties.

For an iid sample $X_1, X_2, \ldots, X_n$ if vector of estimators $T(X) = (T_1(X), \ldots, T_k(X))$ for the parameter vector $\theta \in \Theta$ exists, and if the joint probability function can be written as

$q(x, \eta) = \prod_{i=1}^{n} h(x_i) \exp\left[\sum_{j=1}^{k} \eta_j \sum_{i=1}^{n} T_j(x_i) - nA(\eta)\right]$, where $\eta_1(\theta), \ldots, \eta_k(\theta), B(\theta)$ are real-valued functions on Θ, $T_1, \ldots, T_k, h$ are real valued functions on $\Re^p$ such $T_j(X) = \sum_{i=1}^{n} T_j(x_i)$ and $A(\eta) = \log \int \ldots\ldots \int h(x) \exp\left[\sum_{j=1}^{k} \eta_j T_j(x)\right] dx$, then the distribution is said to belong to a *k-parameter exponential family* for which the dimension of the parameter vector θ is equal to k (or $\dim(\theta) = k$). A k-parameter exponential family has a k-dimensional sufficient statistic regardless of the sample size. The vector $T(x) = (T_1(x), \ldots, T_k(x))$ is called the natural sufficient statistic, which is also *complete*. If the dimension of the vector θ , i.e. $\dim(\theta) < k$ then the k parameter exponential family transforms to a curved exponential family which also enjoy many of the properties of the full families. (For more details refer Bickel and Doksum, 2000). Note that curved exponential families have sufficient statistic, but they are not complete, since the parameter space does not contain a k-dimensional open set. (Casella and Berger, 2002). In such two-dimensional cases, a sufficient statistic is a vector, say $T(X) = (T_1(X), T_2(X))$. This situation often occurs when the parameter θ is a one-dimensional parameter, and then the parameter space is a curve in two dimensions which is a subset of the full two-dimensional space (Lehmann and Casella, 1998). This follows from the fact that according to the theorem 2, we can construct better estimator for θ by combining the two estimators in the joint sufficient statistic $(T_1(X), T_2(X))$.

Note that according to the above description probability distribution (3.1) belongs to the two-parameter exponential family with,

$$\eta = (\eta_1, \eta_2) = \left(\frac{\mu}{\sigma^2}, \frac{-1}{2\sigma^2}\right) \text{ and } A(\eta) = \frac{1}{2}\left(\frac{\mu^2}{\sigma^2} + \log 2\pi\sigma^2\right) = \frac{1}{2}\left(\frac{-\eta_1^2}{2\eta_2} + \log\left(-\frac{\pi}{\eta_2}\right)\right)$$

If the coefficient of variation of the distribution $\nu = \sigma/\mu$ is assumed to be known, then $X_i \sim N(\mu, \nu^2\mu^2)$. Now the joint probability distribution function (3.1) can be written as

$$p(x_1,\ldots,x_n) = \left(\frac{1}{\sqrt{2\pi}\nu\mu}\right)^n \exp\left\{-\frac{n}{2\nu^2}\left(\frac{\bar{x}}{\mu} - 1\right)^2 - \frac{1}{2\nu^2\mu^2}\sum_{i=1}^{n}(x_i - \bar{x})^2\right\}. \qquad (6)$$

Since $\dim(\theta) < k$ (6) represents a curved exponential family with,

$$T(X) = (T_1(X), T_2(X)) = \left(\sum_{i=1}^n x_i, \sum_{i=1}^n (x_i - \bar{x})^2\right),$$

which is minimal sufficient for μ, but not a complete sufficient statistic. Since the mean gradient vector, the covariance matrix (or Hessian matrix) and the moment generating function of $T(X)$ are $E[T(X)] = nA'(\eta)$, $Var[T(X)] = nA''(\eta)$, and $\Psi_{T(X)}(t) = \{\exp[A(\eta + t) - A(\eta)]\}^n$ respectively, it can be easily shown that

$E[T_1(X)] = E\left[\sum x_i\right] = n\mu$, $\quad Var[T_1(X)] = Var\left[\sum x_i\right] = n\sigma^2 = n\nu^2\mu^2$,

$E[T_2(X)] = E\left[\sum_{i=1}^n x_i^2\right] = n(\mu^2 + \sigma^2)$, and $Var[T_2(X)] =$ $Var\left[\sum_{i=1}^n x_i^2\right] = 2n\sigma^2(2\mu^2 + \sigma^2)$.

Since $\sigma^2 = \nu^2\mu^2$ and $W = (n-1)S^2/\sigma^2 \sim \chi^2_{(n-1)}$,

$E[T_2(X)] = (n-1)\sigma^2 = (n-1)\nu^2\mu^2$ and $Var[T_2(X)] = 2(n-1)\sigma^4 = 2(n-1)\nu^4\mu^4$,

where $S^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2 \big/ (n-1)$. For the transformation $U = \sqrt{W}$ it is clear that $F_U(u) = F_W(u^2)$. Hence the probability density function of U is $f_U(u) = u^{n-2}e^{\frac{-1}{2}u^2} \big/ 2^{(n-3)/2}\Gamma((n-1)/2);\quad u > 0$.

Then the m^{th} moment of U is given by

$E(U^m) = \int_0^\infty u^m f_U(u)\,du = \left[2^{m/2}\Gamma((n+m-1)/2)\right]/\Gamma((n-1)/2)$ where

$\Gamma(r) = \int_0^\infty \lambda^r y^{r-1}e^{-\lambda y}dy$.

Thus the m^{th} moment of S is

$$E(S^m) = \frac{E(U^m)}{(n-1)^{m/2}}\sigma^m = \frac{2^{m/2}\Gamma((n+m-1)/2)}{(n-1)^{m/2}\Gamma((n-1)/2)}\sigma^m = \frac{2^{m/2}\Gamma((n+m-1)/2)}{(n-1)^{m/2}\Gamma((n-1)/2)}\nu^m\mu^m. \quad (7)$$

By applying $m = 1$, we can obtain an unbiased estimator of σ as $\frac{(n-1)^{1/2}\Gamma((n-1)/2)}{2^{1/2}\Gamma(n/2)}S$,

and an unbiased estimator of $\mu(>0)$ as $\frac{(n-1)^{1/2}\Gamma((n-1)/2)}{\nu 2^{1/2}\Gamma(n/2)}S$.

Also note that

$$Var(S^m) = E(S^{2m}) - (E(S^m))^2$$

$$Var(S^m) = \frac{2^m\nu^{2m}}{(n-1)^m}\left\{\frac{\Gamma((n+2m-1)/2)}{\Gamma((n-1)/2)} - \frac{[\Gamma((n+m-1)/2)]^2}{[\Gamma((n-1)/2)]^2}\right\}\mu^{2m} \quad (8)$$

Since $E(S^m)=k\mu^m$ where $k=\nu^m 2^{m/2}\Gamma((n+m-1)/2)/(n-1)^{m/2}\Gamma((n-1)/2)$, we can use S^m as an estimator for μ^m and

$$\tau^2=\left(\mu^m\right)^{-2}Var\left(S^m\right)=\frac{2^m\nu^{2m}}{(n-1)^m}\left\{\frac{\Gamma((n+2m-1)/2)}{\Gamma((n-1)/2)}-\frac{[\Gamma((n+m-1)/2)]^2}{[\Gamma((n-1)/2)]^2}\right\}.$$

Note that $\phi_1(T_1(X))=\sum_{i=1}^n x_i$, $\phi_2(T_2(X))=S$,are uncorrelated. Then substituting $\phi_1(T_1(X))$ and $\phi_2(T_2(X))$, $k_1=n$, $k_2=\nu 2^{1/2}\Gamma(n/2)/(n-1)^{1/2}\Gamma((n-1)/2)$, and $g(\mu)=\mu$ to theorem 2.1 we can easily show that $\tau_1^2=[\mu]^{-2}Var[\phi_1(T_1(X))]=n\nu^2$,

$$\tau_2^2=[\mu]^{-2}Var[\phi_2(T_2(X))]$$

$$=\frac{2\nu^2}{(n-1)}\left\{\frac{\Gamma((n+1)/2)}{\Gamma((n-1)/2)}-\frac{[\Gamma(n/2)]^2}{[\Gamma((n-1)/2)]^2}\right\}=\nu^2\left\{1-\frac{2[\Gamma(n/2)]^2}{(n-1)[\Gamma((n-1)/2)]^2}\right\},$$

and $k_1^2\tau_2^2+k_2^2\tau_1^2+\tau_1^2\tau_2^2=n\nu^2\left\{n+\nu^2-\frac{2n[\Gamma(n/2)]^2}{(n-1)[\Gamma((n-1)/2)]^2}\right\}$.

Thus the optimal combined estimator of the mean μ is

$$\phi^*(T_1(X),T_2(X))=\frac{\left[1-\frac{2[\Gamma(n/2)]^2}{(n-1)[\Gamma((n-1)/2)]^2}\right]\left(\sum_{i=1}^n X_i\right)+\frac{\nu 2^{1/2}\Gamma(n/2)}{(n-1)^{1/2}\Gamma((n-1)/2)}S}{\left[n+\nu^2-2n(\Gamma(n/2))^2/(n-1)(\Gamma((n-1)/2))^2\right]},\qquad (9)$$

and the optimal combined estimator of standard deviation $\sigma(=\nu\mu)$ is given by

$$\frac{\nu\left[1-\frac{2[\Gamma(n/2)]^2}{(n-1)[\Gamma((n-1)/2)]^2}\right]\left(\sum_{i=1}^n X_i\right)+\frac{\nu^2 2^{1/2}\Gamma(n/2)}{(n-1)^{1/2}\Gamma((n-1)/2)}S}{\left[n+\nu^2-2n(\Gamma(n/2))^2/(n-1)(\Gamma((n-1)/2))^2\right]}.$$

Note that Gleser and Healy (1976), have taken the two estimators as

$T_1 = \overline{X}$ and $T_2 = ((n-1)/n)^{1/2} c_n S$ with $c_n = n^{1/2}\Gamma[(n-1)/2]/(2a)^{1/2}\Gamma(n/2)$, and $d_n = \lfloor n^{-1}(n-1)\, ac_n^2 - 1 \rfloor$, $a = v^2$, and shown that the optimal combined estimator of μ as $\phi^*(T_1(\boldsymbol{X}), T_2(\boldsymbol{X})) = (d_n T_1 + n^{-1}aT_2)/(d_n + n^{-1}a + n^{-1}ad_n)$. Instead of $\phi_1(T_1(\boldsymbol{X})) = \sum_{i=1}^{n} X_i$ and $\phi_2(T_2(\boldsymbol{X})) = S$, if we substitute the estimators $T_1 = \overline{X}$ and $T_2 = ((n-1)/n)^{1/2} c_n S$ to theorem 2.1 the same optimal combined estimator can be obtained.

Next, to derive the optimal combined estimator of σ^2, consider, $\overline{X} \sim N(\mu, \sigma^2)$.

Then $E(\overline{X}) = \mu$, $Var(\overline{X}) = \sigma^2/n$, and

$E(\overline{X}^2) = Var(\overline{X}) + [E(\overline{X})]^2 = \sigma^2/n + \mu^2 = (1 + v^2/n)\mu^2$.

Hence $\overline{X}^2$ is an estimator of μ^2. Similarly it can be shown that

$$E(\overline{X}^4) = \mu^4 + 6\mu^2\sigma^2/n + 3\sigma^4/n^2 . \tag{10}$$

Then $$Var(\overline{X}^2) = E(\overline{X}^4) - [E(\overline{X}^2)]^2 = \mu^4 + 6\mu^2\sigma^2/n + 3\sigma^4/n^2 - (\sigma^2/n + \mu^2)^2$$
$$= 2\sigma^2(2n\mu^2 + \sigma^2)/n^2 = 2v^2(2n + v^2)\mu^4/n^2 . \tag{11}$$

Now to derive an optimal combined estimator for σ^2 apply

$\phi_1(T_1(\boldsymbol{X})) = \overline{X}^2$, $\phi_2(T_2(\boldsymbol{X})) = S^2$, $k_1 = (1 + v^2/n)$,

$k_2 = v^2 2\Gamma((n+1)/2)/(n-1)\Gamma((n-1)/2) = v^2$, and $g(\mu) = \mu^2$ to theorem 2,

since $\phi_1(T_1(\boldsymbol{X}))$ and $\phi_2(T_2(\boldsymbol{X}))$ are uncorrelated. Then we can show that

$\tau_1^2 = [\mu^2]^{-2} Var[\phi_1(T_1(\boldsymbol{X}))] = [\mu^2]^{-2} Var[\overline{X}^2] = 2v^2(2n + v^2)/n^2$,

$\tau_2^2 = [\mu^2]^{-2} Var[\phi_2(T_2(\boldsymbol{X}))] = [\mu^2]^{-2} Var[S^2] = 2v^4/(n-1)$, and

$$k_1^2\tau_2^2 + k_2^2\tau_1^2 + \tau_1^2\tau_2^2 = \frac{2v^4}{n^2}\left\{\frac{(n+v^2)^2}{(n-1)} + \frac{(n+1)v^2(2n+v^2)}{(n-1)}\right\}.$$

Thus, the optimal combined estimator of μ^2 is

$$\frac{\left(n\left(n+\nu^2\right)/(n-1)\right)\overline{X}^2+\left(2n+\nu^2\right)S^2}{\left(\left(n+\nu^2\right)^2/(n-1)\right)+\left((n+1)\nu^2\left(2n+\nu^2\right)/(n-1)\right)}$$, and hence,

$$\phi^*(T_1(X),T_2(X))=\frac{\left(n\left(n+\nu^2\right)/(n-1)\right)\overline{X}^2+\left(2n+\nu^2\right)S^2}{\left(\left(n\nu^{-1}+\nu\right)^2/(n-1)\right)+\left((n+1)\left(2n+\nu^2\right)/(n-1)\right)} \tag{12}$$

is the optimal combined estimator of the population variance $\nu^2\mu^2=\sigma^2$.

To assess the efficiency of estimators Scaler Mean Square Error Loss (SMSEL) (Kanfuji and Iwase, 1998) is used. The SMSEL of any estimator $T(X)$ for $g(\theta)$ is $SMSEL[T(X)]=MSE[T(X)]/(g(\theta))^2=\tau^2\left(k^2+\tau^2\right)^{-1}$, where $MSE[T(X)]$ is the mean squared error of the estimator $T(X)$. Thus the SMSEL of $\left|\phi^*(T_1(X),T_2(X))\right|$ and $\left|\phi_i^*(T_i(X))\right|$ are $SMSEL\left|\phi^*(T_1(X),T_2(X))\right|=\tau_1^2\tau_2^2/(k_1{}^2\tau_2^2+k_2^2\tau_1^2+\tau_1^2\tau_2^2)$, and $SMSEL\left|\phi_i^*(T_i(X))\right|=\tau_i^2\left(k_i^2+\tau_i^2\right)^{-1}$ respectively, and the $SMSEL=0$ when the estimated value is equal to the actual value so that the distance between $\phi_i^*(T_i(X))$ and $g(\theta)$ is around zero. The notations OSE1, OSE2 and OSE3 are used to represent the optimal shrunken estimators for $\phi_1^*(T_1(X))$, $\phi_2^*(T_2(X))$ and the combined estimator $\phi^*(T_1(X),T_2(X))$ respectively.

SMSEL of mean estimators

To estimate mean μ the selected minimal sufficient estimators are $\phi_1(T_1(X))=\sum_{i=1}^{n}X_i$, and $\phi_2(T_2(X))=S$, and the optimal combined estimator is found to be

$$\phi^*(T_1(X),T_2(X))=\frac{\left[1-\dfrac{2[\Gamma(n/2)]^2}{(n-1)[\Gamma((n-1)/2)]^2}\right]\left(\sum_{i=1}^{n}X_i\right)+\dfrac{\nu 2^{1/2}\Gamma(n/2)}{(n-1)^{1/2}\Gamma((n-1)/2)}S}{\left[n+\nu^2-2n(\Gamma(n/2))^2/(n-1)(\Gamma((n-1)/2))^2\right]}.$$

The SMSEL's of the estimators are

$$SMSEL\left[\phi_1^*(T_1(X))\right] = \nu^2/(n+\nu^2),\quad SMSEL\left[\phi_2^*(T_2(X))\right] = 1 - \frac{2(\Gamma(n/2))^2}{(n-1)(\Gamma((n-1)/2))^2} \text{ and}$$

$$SMSEL\left[\phi^*(T_1(X),T_2(X))\right] = \frac{\nu^2\left\{1 - \frac{2(\Gamma(n/2))^2}{(n-1)(\Gamma((n-1)/2))^2}\right\}}{\nu^2 + n\left\{1 - \frac{2(\Gamma(n/2))^2}{(n-1)(\Gamma((n-1)/2))^2}\right\}}.$$

The SMSEL's of these estimators are compared for three different values of coefficient of variations CV = 0.3, CV = 0.5, and CV = 0.7 with increasing sample sizes.

According to the Figure 1 (see Appendix) it is clear that OSE3, which is the optimal combined estimator, has a minimum mean square error with respect to the other two estimators. Also the estimated value is very close to the actual value when the sample size is increased.

SMSEL of variance estimators

The minimal sufficient statistics $\phi_1(T_1(X)) = \overline{X}^2$ and $\phi_2(T_2(X)) = S^2$ are taken as the possible estimators, and the corresponding optimal combined estimator is

$$\phi^*(T_1(X),T_2(X)) = \frac{\left(n(n+\nu^2)/(n-1)\right)\overline{X}^2 + (2n+\nu^2)S^2}{\left((n\nu^{-1}+\nu)^2/(n-1)\right) + \left((n+1)(2n+\nu^2)/(n-1)\right)}.$$

Their SMSEL's are

$$SMSEL\left[\phi_1^*(T_1(X))\right] = \frac{2\nu^2(2n+\nu^2)}{n^2+6n\nu^2+3\nu^4},\quad SMSEL\left[\phi_2^*(T_2(X))\right] = \frac{2}{n+1}, \text{ and}$$

$$SMSEL\left[\phi^*(T_1(X),T_2(X))\right] = \frac{2\nu^2(2n+\nu^2)}{n^2+2n(n+2)\nu^2+(n+2)\nu^4} \text{ respectively.}$$

In this case also it is clear that the optimal combined estimator for the variance, OSE3, (see Figure 2 in Appendix) has a minimum mean square error with respect to the other two estimators. As in the mean estimator, the estimated variance is very close to the actual value when the sample size is increased.

CONCLUSION

It has been seen that the suggested optimal combined estimators are more efficient with known prior information. Although in this paper normal distribution with known coefficient of variation is considered, theorem 2.1 can be applied to other probability distributions with any other available prior information (kurtosis, skewness etc.).

REFERENCES

Arnholt, A.T. and J.L. Hebert (1995). Estimating the mean with known coefficient of variation. *The American Statistician,* 49: 367-369.

Arnholt, A.T. and J.L. Hebert (2001). Optimal combinations of pairs of estimators. *Interstat,*<http://interstat.statjournals.net/YEAR/2001/articles/0103002.pdf>.

Bickel, P.J. and K.A. Doksum (2000) *Mathematical Statistics, Basic Ideas and Selected Topics,* Vol.1 (2nd Ed). Prentice Hall, New Jersey.

Casella, G. and R.L. Berger (2002). *Statistical Inference,* (2nd Ed). Pacific Grove, Calif. USA.

Ester, Samuel-Cahn (1994). Combining unbiased estimators. *The American Statistician*, 48: 34-36.

Gleser, L.J. and J.D. Healy (1976). Estimating the mean of a normal distribution with known coefficient of variation. *Journal of the American Statistical Association,* 71: 977-981.

Kanfuji, K. and K. Iwase (1998). Estimation for a scale parameter with known coefficient of variation, *Statistical Papers*, 39: 377-388.

Khan, R.A. (1968). A note on estimating the mean of a normal distribution with known coefficient of variation. *Journal of the American Statistical Association,* 63: 1039-1041.

Laheetharan, A. and P. Wijekoon (2008). Improved estimation of the population parameters when some additional information is available. *Statistical Papers*, DOI:10:1007/s00362-008-0185-5.

Lehmann, E.L. and G. Casella (1998) *Theory of Point Estimation*, (2nd Ed). Springer-Verlag Berlin Heidelberg, New York.

Marshall, J. (1936). The distribution in sampling of insect populations in the field with special reference to the American bollworm. *Annals of Applied Biology*, 2B: 150.

Searls, D.T. (1964). The utilization of a known coefficient of variation in the estimation procedure. *Journal of the American Statistical Association*, 59: 1225-1226.

Sen, A.R. (1978). Estimation of population mean when the coefficient of variation is known. *Communication in Statistics-Theory and Methods*, A8(7): 657-672.

Wencheko, E. and P. Wijekoon (2005). Improved estimation of the mean in one-parameter exponential families with known coefficient of variation. *Statistical Papers*, 46: 101-115.

APPENDIX

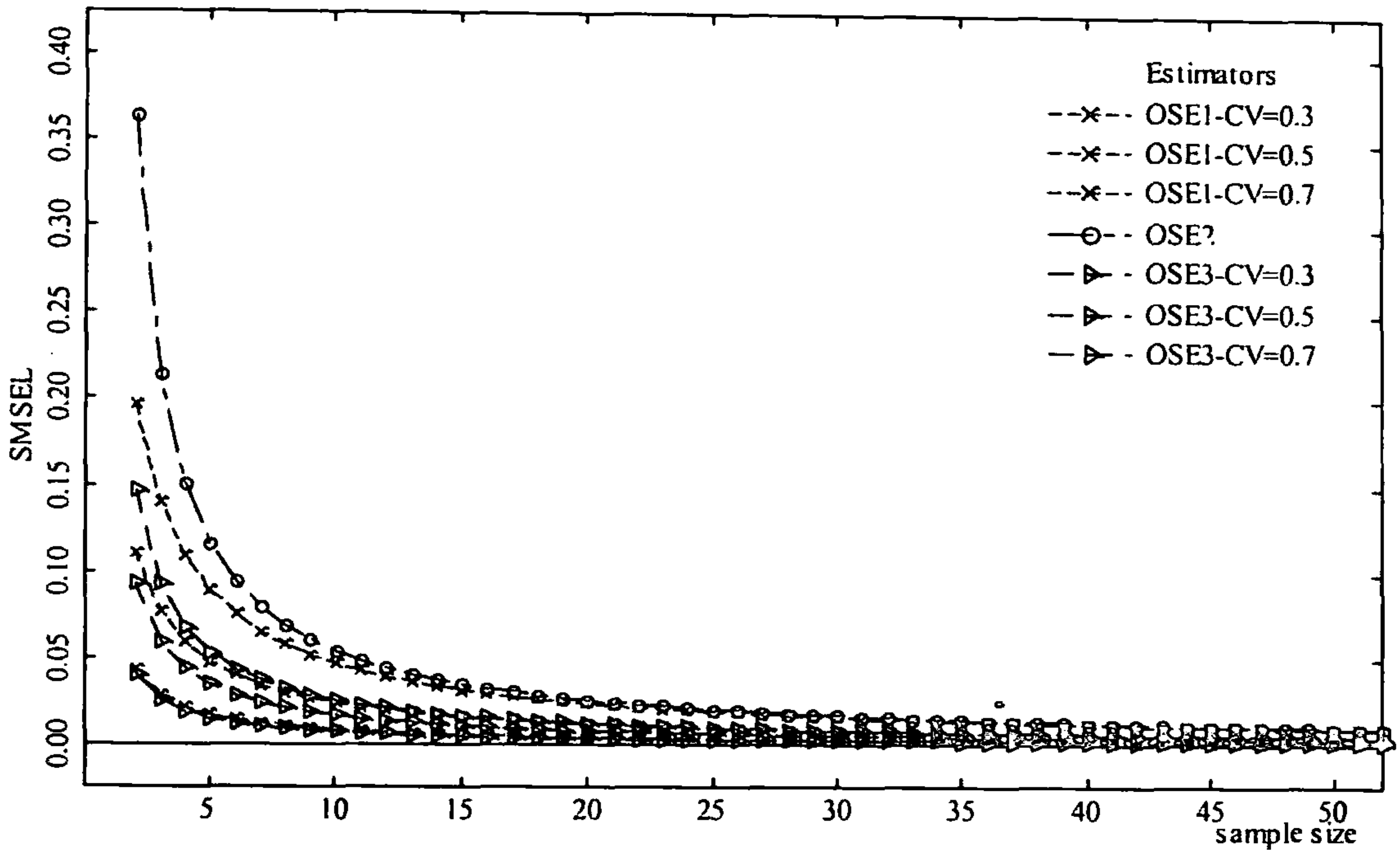

Fig 1:SMSEL vs sample size for estimation of mean (Normal distribution)

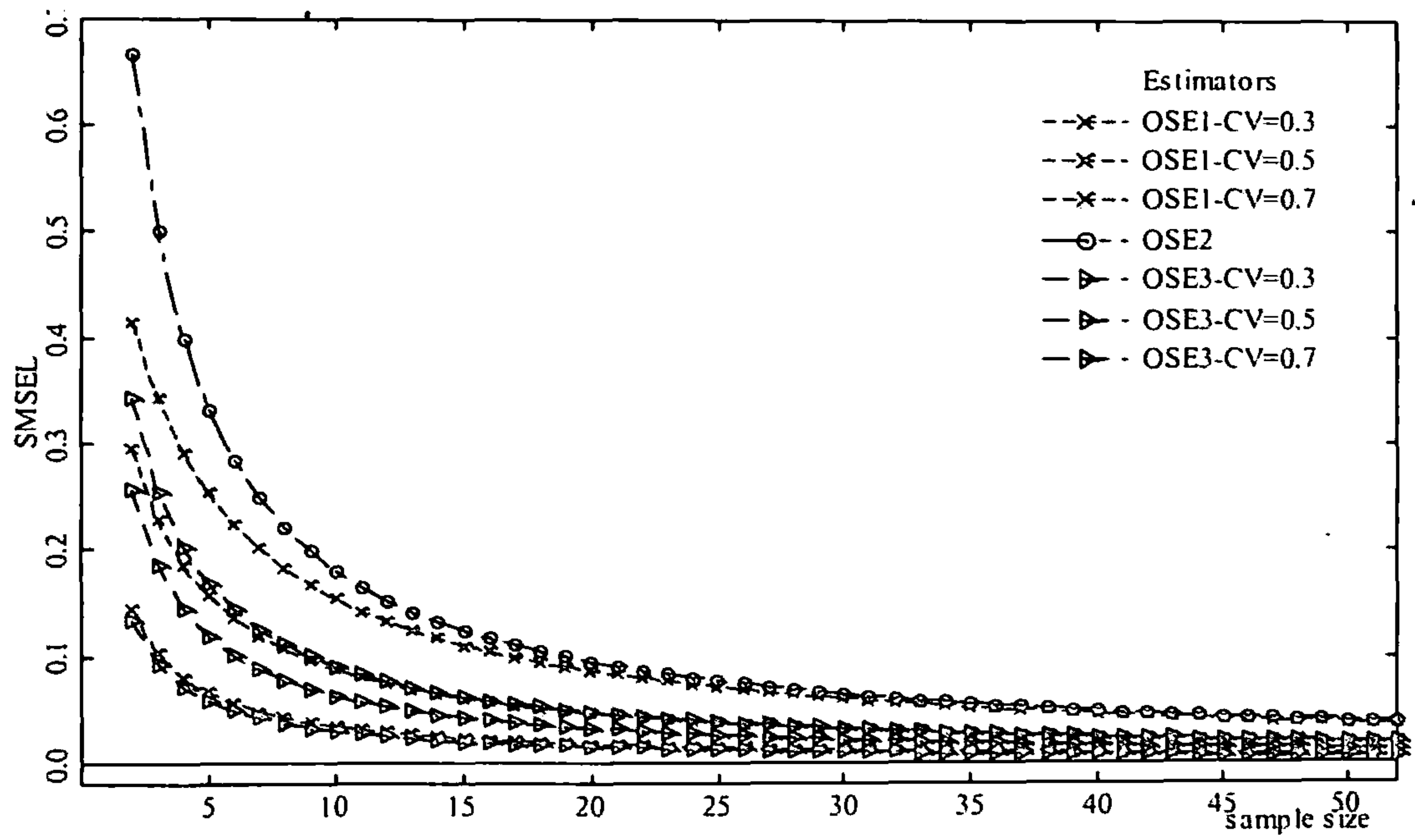

Fig 2: SMSEL vs sample size for estimation of variance (Normal distribution)