Statistical Papers **46**, 101-115 (2005)

Statistical Papers

Improved estimation of the mean in one-parameter exponential families with known coefficient of variation

E. Wencheko[1] and P. Wijekoon[2]

[1] Department of Statistics, Addis Ababa University, Ethiopia
[2] Department of Statistics and Computer Sciene, University of Peradeniya, Sri Lanka

Received: April 26, 2002, revised version: September 6, 2003

Abstract: The value for which the mean square error of a biased estimator aT for the mean μ is less than the variance of an unbiased estimator T is derived by minimizing $MSE(aT)$. The resulting optimal value is $1/[1+c(n)v^2]$, where $v = \sigma/\mu$, is the coefficient of variation. When T is the UMVUE $\bar{X}$, then $c(n) = 1/n$, and the optimal value becomes $1/(n+v^2)$ (Searls, 1964). Whenever prior information about the size of v is available the shrinkage procedure is useful. In fact for some members of the one-parameter exponential families it is known that the variance is at most a quadratic function of the mean. If we identify the pertinent coefficients in the quadratic function, it becomes easy to determine v.

Introduction

It is known that one important way of improving estimation of distribution parameters is by way of employing biased estimation procedures. The mean square is the most popular and widely used criterion to show when a biased estimator is better than an unbiased estimator. In other words the region of improvement of estimation is obtained from comparing the scalar or (matrix) mean square error (MSE) of the biased estimator with the variance (covariance matrix) of the unbiased estimator. In this connection Perlman (1972) provides necessary and sufficient condition for the existence of a real $a \in (0,1)$ such that a shrinkage of an unbiased vector-valued estimator would result in an improvement in MSE. Bibby (1972) and Bibby and Toutenburg (1977, 1978) discuss biased estimation procedures and estimators that result from unbiased procedures and estimators. Kleffe (1985) uses a general result due to Perlman to derive uniformly better estimators. Further results which relate to the improvement of the estimator of the population mean are given in Searls (1964), Khan (1968), Gleser and Healy (1976), Arnholt and Hebert (1995). In the current paper we concentrate on certain members of the one-parameter exponential families. Under the assumption that the coefficient of variation is known *a priori* we derive estimators for the mean with minimum MSE.

A Review of some Results

Suppose that a univariate variable $\mathbf{X} \sim \mathcal{D}(\mu, \sigma^2)$, $\mu \neq 0$, where $\mathcal{D}$ is any distribution, and that the coefficient of variation $v = \sigma/\mu$ or the noise-to-signal, $\sigma/|\mu|$, $\mu < 0$, is known. As a consequence the variance of $\mathbf{X}$ can be written as $\sigma^2 = v^2\mu^2$. It is not uncommon that experience and results obtained from applications justify the assumption of known v. The prior information about v can be incorporated in distribution models. For example, in the case of the exponential distribution $v = 1$. The statistical literature provides some results that relate to this topic.

Suppose that an iid sample $\mathbf{X} = (X_1, \cdots, X_n)'$ is available. Based on the assumption that v is known Searls (1964) shows that an improvement of es-

timation of the mean is possib

$\mathcal{C}(a)$

According to Khan (1968), in

$\mathcal{C}(a, 1-a) = \{T(\mathbf{X}) \,|\, T($

(see Examples below) the wei

of $T_1(\mathbf{X})$ or $T_2(\mathbf{X})$ in MSE w

distribution is known. Glese

to a wider class

$\mathcal{C}(a_1, a_2) = \{T(\mathbf{X}) \,|\, T($

(see Examples below) that c

$T_1(\mathbf{X})$ and $T_2(\mathbf{X})$ of the mea

that has minimum MSE. Th

ther enrichment to the abo

estimators. This paper stud

discusses intervals on which

one-parameter exponential f

Example 1 [Searls (1964)

Suppose $\mathbf{X} = (X_1, \cdots, X_n)$

mean μ, $\mu \neq 0$, and varia

$a\sum_{i=1}^n X_i$, $a \in (0,1)$, as an e

the purpose here is a real n

The optimal a is

timation of the mean is possible by considering a class of shrunken estimators

$$\mathcal{C}(a) = \{aT(\mathbf{X}) | a \in (0,1)\} .$$

According to Khan (1968), in the class

$$\mathcal{C}(a, 1-a) = \{T(\mathbf{X}) \,|\, T(\mathbf{X}) = aT_1(\mathbf{X}) + (1-a)T_2(\mathbf{X}) \ , \ a \in [0,1]\}$$

(see Examples below) the weighted unbiased estimator $\mathrm{T}(\mathbf{X})$ dominates each of $T_1(\mathbf{X})$ or $T_2(\mathbf{X})$ in MSE when the coefficient of variation for the normal distribution is known. Gleser and Healy (1976) extend the results of Khan to a wider class

$$\mathcal{C}(a_1, a_2) = \{T(\mathbf{X}) \,|\, T(\mathbf{X}) = a_1 T_1(\mathbf{X}) + a_2 T_2(\mathbf{X}) \ , \ a_1, a_2 \in (0,1)\}$$

(see Examples below) that consists of both biased and unbiased estimators $T_1(\mathbf{X})$ and $T_2(\mathbf{X})$ of the mean. They derive a biased estimator of the mean that has minimum MSE. The work by Arnholt and Hebert (1995) is a further enrichment to the above estimation procedure for a general class of estimators. This paper studies improvement of estimation of the mean and discusses intervals on which improvement in estimation is attained for the one-parameter exponential families of distributions.

Example 1 [Searls (1964)]

Suppose $\mathbf{X} = (X_1, \cdots, X_n)'$ is a random sample from a distribution with mean μ, $\mu \neq 0$, and variance σ^2. We can consider the statistic $T(\mathbf{X}) = a\sum_{i=1}^n X_i$, $a \in (0,1)$, as an estimator for μ. The best choice of a constant for the purpose here is a real number $a \in (0,1)$ which minimizes $MSE[aT(\mathbf{X})]$. The optimal a is

$$a^* = \frac{1}{n + v^2}.$$

Hence, the estimator of the mean μ with minimum MSE is $T^*(\mathbf{X}) = a^*T(\mathbf{X})$.

Example 2 [Khan (1968)]

Suppose $\mathbf{X} = (X_1, \cdots, X_n)'$ is an iid random sample from $\mathcal{N}(\mu, v^2\mu^2), \mu \neq 0$, and the coefficient of variation v is known. Khan considers two unbiased estimators for μ, namely the sample mean $T_1(\mathbf{X}) = \bar{X}$ and a linear function of the square root of the total sum of squares S_n

$$T_2(\mathbf{X}) = c_n[\sum_{i=1}^{n} (Xi - \bar{X})^2]^{1/2} = c_n S_n \ ,$$

where the constant

$$c_n = n^{1/2}(2v)^{-1/2}\frac{\Gamma((n-1)/2)}{\Gamma(n/2)} \ .$$

The variances of the two estimators are

$$VarT_1(\mathbf{X}) = n^{-1}v^2\mu^2$$

and

$$VarT_2(\mathbf{X}) = \mu^2[-1 + \frac{n-1}{2} \times \frac{\Gamma^2((n-1)/2)}{\Gamma^2(n/2)}] \ .$$

The weighted sum of the unbiased estimators $T_1(\mathbf{X})$ and $T_2(\mathbf{X})$ belongs to the class $\mathcal{C}(a, 1-a)$ given above with $VarT(\mathbf{X}) = a^2VarT_2(\mathbf{X}) + (1-a)^2VarT_1(\mathbf{X})$

This last result is given purely as the sum of two variances because the two estimators $T_1(\mathbf{X})$ and $T_2(\mathbf{X})$ are uncorrelated. In fact, by normality, they are independent.

That value of a which minimizes $VarT(\mathbf{X})$ is

$$a^* = \frac{VarT_1(\mathbf{X})}{VarT_1(\mathbf{X}) + VarT_2(\mathbf{X})}$$

$$= v\{(v + n[-1 + \frac{n-1}{2} \times \frac{\Gamma^2((n-1)/2)}{\Gamma^2(n/2)}])\}^{-1} \ .$$

The resulting unbiased estin
each one of $T_1(\mathbf{X})$ and $T_2(\mathbf{X}$
any other unbiased estimato

Example 3 [Gleser and H

An extension of the result
they consider a wider class
necessarily a weighted linea

$\mathcal{C}(a_1, a_2) = \{T(\mathbf{X}) \,|\, T($

where $T_1(\mathbf{X})$ and $T_2(\mathbf{X})$ are
They show that a shrinkage

$s(v, n) = (d_r$

where $d_n = -1 + v(n-1)c_r$

with uniformly minimum M

Example 4 [Arnholt and

Suppose $\mathbf{X} = (X_1, \cdots, X_n)'$
and that the coefficient of
median. Note that in this c
where κ_n is the variance of t
optimal biased estimator o
M_n given by

T_r

The resulting unbiased estimator $T_3(\mathbf{X}) = a^* T(\mathbf{X})$ is uniformly better than each one of $T_1(\mathbf{X})$ and $T_2(\mathbf{X})$, and more generally it is uniformly better than any other unbiased estimator in the class $\mathcal{C}(a)$.

Example 3 [Gleser and Healy (1976)]

An extension of the result by Khan is given by Gleser and Healy where they consider a wider class (a linear combination of $T_1(\mathbf{X})$ and $T_2(\mathbf{X})$, not necessarily a weighted linear combination)

$$\mathcal{C}(a_1, a_2) = \{T(\mathbf{X}) \mid T(\mathbf{X}) = a_1 T_1(\mathbf{X}) + a_2 T_2(\mathbf{X}) \ , \ a_1, \ a_2 \in [0,1]\} \ ,$$

where $T_1(\mathbf{X})$ and $T_2(\mathbf{X})$ are as given in Khan.

They show that a shrinkage of the unbiased estimator $T_3(\mathbf{X})$ multiplied by

$$s(v,n) = (d_n + vn^{-1})/(d_n + vn^{-1} + vd_n n^{-1}) \ ,$$

where $d_n = -1 + v(n-1)c_n{}^2 n^{-1}$, gives rise to a biased estimator

$$T_4(\mathbf{X}) = s(a,n) T_3(\mathbf{X})$$

with uniformly minimum MSE showing that $T_4(\mathbf{X})$ dominates $T_3(\mathbf{X})$.

Example 4 [Arnholt and Hebert (1995)]

Suppose $\mathbf{X} = (X_1, \cdots, X_n)'$ is an iid random sample from $\mathcal{N}(\mu, v^2\mu^2), \mu \neq 0$, and that the coefficient of variation v is known. Let M_n be the sample median. Note that in this case we have $E(M_n) = \mu$ and $Var(M_n) = \kappa_n v^2 \mu^2$, where κ_n is the variance of the median of n iid standard normal variates. The optimal biased estimator of the mean μ is a shrinkage of the sample median M_n given by

$$T_{\kappa_n}(\mathbf{X}) = (\kappa_n v^2 + 1)^{-1} M_n$$

with

$$MSE(T_{\kappa_n}(\mathbf{X})) = (\kappa_n v^2\mu^2 + 1)(\kappa_n v^2 + 1)^{-1}$$

Comparison of the $MSE(M_n)$ and $MSE(T_{\kappa_n}(\mathbf{X}))$ shows that $T_{\kappa_n}(\mathbf{X})$ is the more efficient of the two estimators.

Improved Estimation and Intervals

A Generalization of the Result by Searls

Suppose $\mathbf{X} = (X_1, \cdots, X_n)'$ is an iid sample from $\mathcal{D}(\mu, \sigma^2)$, $\mu \neq 0$, and that on the basis of the sample there exists an unbiased estimator, say $T(\mathbf{X})$, for μ. Let the variance of $T(\mathbf{X})$ be $Var[T(\mathbf{X})] = c(n)\sigma^2$, where $c(n)$ is a function of the sample size n. If we consider biased estimators of μ from the class $C(a) = \{aT(\mathbf{X}) | a \in (0,1)\}$, we have $E[aT(\mathbf{X})] = a\mu$, $Var[aT(\mathbf{X})] = c(n)a^2\sigma^2$, and hence $MSE[aT(\mathbf{X})] = c(n)a^2\sigma^2 + (a-1)^2\mu^2$. Since $(d/da)[MSE(a\bar{X})] = 2c(n)a\sigma^2 + 2(a-1)\mu^2$ and $(d^2/da^2)[MSE(aT(\mathbf{X}))] = 2c(n)\sigma^2 + 2\mu^2 > 0$, it follows that the biased estimator with minimum MSE in the class $C(a)$ is

$$T^{**}(\mathbf{X}) = a^{**}T(\mathbf{X}) \ , \quad a^{**} = \frac{1}{c(n)v^2 + 1} \ .$$

Hence $MSE[T^{**}(\mathbf{X})] = c(n)\sigma^2(c(n)v^2+1)^{-1} < c(n)\sigma^2$, from which it is evident that the relative efficiency of $T^{**}(\mathbf{X})$ in relation to $T(\mathbf{X})$, defined as $Var[T(\mathbf{X})]/MSE[T^{**}(\mathbf{X})]$, is $c(n)v^2 + 1$. Since $c(n)v^2 + 1 > 1$ the estimator $T^{**}(\mathbf{X})$ dominates the unbiased $T(\mathbf{X})$.

Note that the result by Searls is a special case of the above with $c(n) = 1/n$.

The Improvement Interv

An issue closely associated
of regions of improvement;
domination in terms of the
while considering this poin
improvement interval lead t

As a consequence of the ge
which improvement is gaine

$I(c(n), v) = \{a \in$

We observe that when v is
to diminish, and therefore t
will not be of great interest
then all values of v betwee
mean.

In the following section we
pertinent matters about on

One-Parameter Exp

The family of distributio
parameter exponential fam
on Θ, real-valued functions
functions $p(x, \theta)$ of P_θ may

$p(x, \theta) = h(x)$

$_{\kappa_n} v^2 + 1)^{-1}$

)) shows that $T_{\kappa_n}(\mathbf{X})$ is the

$\mathcal{D}(\mu, \sigma^2)$, $\mu \neq 0$, and that
estimator, say $T(\mathbf{X})$, for μ.
where $c(n)$ is a function of
s of μ from the class $C(a) =$
$ar[aT(\mathbf{X})] = c(n)a^2\sigma^2$, and
ince $(d/da)[MSE(a\bar{X})] =$
$))] = 2c(n)\sigma^2 + 2\mu^2 > 0$, it
MSE in the class $C(a)$ is

$\frac{1}{)v^2 + 1}$.

$c(n)\sigma^2$, from which it is
lation to $T(\mathbf{X})$, defined as
$n)v^2 + 1 > 1$ the estimator

he above with $c(n) = 1/n$.

The Improvement Interval

An issue closely associated with improved estimation is the determination of regions of improvement; in the univariate case these are intervals. Since domination in terms of the scalar MSE is limited to an interval it is worthwhile considering this point as well because all shrinkage factors from the improvement interval lead to admissibility of the more efficient estimator.

As a consequence of the generalized result above we can give the interval on which improvement is gained by

$$I(c(n), v) = \{a \in \Re_+ \mid \frac{1 - c(n)v^2}{1 + c(n)v^2} < a < 1, v \in (0,1)\} .$$

We observe that when v is very close to zero the improvement interval seems to diminish, and therefore the possibility of improved estimation by shrinkage will not be of great interest. On the other hand if v is not very close to zero, then all values of v between zero and one lead to improved estimation of the mean.

In the following section we introduce definitions and terminology and discuss pertinent matters about one-parameter exponential families of distributions.

One-Parameter Exponential Families of Distributions

The family of distributions of a model $\{P_\theta | \theta \in \Theta\}$, is said to be a one-parameter exponential family, if there exist real-valued functions $\eta(\theta)$, $B(\theta)$ on Θ, real-valued functions T and h on $\Re^p$, such that the density (frequency) functions $p(x, \theta)$ of P_θ may be written

$$p(x, \theta) = h(x)\, exp\, [\eta(\theta)T(x) - B(\theta)], x \in X \subset \Re^p .$$

Note that the functions η, B and T are not unique (Bickel and Doksum, 2000, p. 49).

The families of distributions obtained from one-parameter exponential families are also themselves one-parameter exponential families. Suppose $\mathbf{X} = (X_1, \cdots, X_n)'$ is an iid sample with common distribution P_θ where P_θ form a one-parameter exponential family. If $\{P_\theta{}^{(n)}\}$ is the family of distributions and $p(x, \theta)$ are the corresponding density (frequency) functions, we have

$$p(x,\theta) = \prod_{i=1}^{n} h(x_i)\, exp\, [\eta(\theta)T(x_i) - B(\theta)] = [\prod_{i=1}^{n} h(x_i)]\, exp\, [\eta(\theta) \sum_{i=1}^{n} T(x_i) - nB(\theta)]$$

where $x = (x_1, \cdots, x_n)'$. Therefore, the $\{P_\theta{}^{(n)}\}$ form a one-parameter exponential family.

We obtain a re-parameterization of the exponential family by letting the model to be indexed by η rather than θ (Bickel and Doksum, 2000, p. 52). The exponential family then has the form

$$q(x,\eta) = h(x)\, exp\, [\eta T(x) - A(\eta)], x \in X \subset \Re^p$$

where $A(\eta) = log \int \cdots \int h(x)\, exp[\eta T(x)]dx$ in the continuous case and the integral is replaced by a sum in the discrete case. If $\theta \in \Theta$, then $A(\eta)$ must be finite. The model $q(x, \eta)$ with η ranging over the space $\mathcal{E}$ is called the canonical one-parameter exponential family generated by T and h. The space $\mathcal{E}$ is called the natural parameter space and T is called the natural sufficient statistic. As in the case of $p(x, \theta)$ for an iid sample $X_1, \cdots, X_n$ we have

$$q(x,\eta) = [\prod_{i=1}^{n} h(x_i)]\, exp\, [\eta \sum_{i=1}^{n} T(x_i) - nA(\eta)]\ .$$

For example, for the Poisson distribution we have

$$p(x,\theta) = \frac{1}{x!}\, exp\, [log(\theta)T(x) - \theta]\ , \quad x = 0, 1, \cdots; \theta > 0\ ,$$

with $h(x) = 1/x!$, $\eta(\theta) = lo$

$q(x, \eta)$

where $h(x) = 1/x!$, $\eta = log$

The re-parameterized cano

with $\theta = (\mu, \sigma^2)', \mu \neq 0,$

$\theta = (\alpha, \beta)'$, and the negati

of the cases the statistic $\sum$

the mean by $a \sum X_i$, $a \in$ (

the normal), and n, α, and

Morris (1982) gives a chara

mial and negative binomia

tions enjoy wide applicatio

because they are natural

functions. In these cases, t

mean μ. This latter relatio

σ^2

where the constants $\zeta_0, \zeta_1,$

It can be shown that the

variance function), $\zeta_1 =$

$\zeta_1 = 1$ (linear), $\zeta_2 = 0$),

$0, \zeta_2 = 1/\alpha^2$), the binomia

$0, \zeta_1 = 1, \zeta_2 = -1/n$)(with

$\mu = rq/p$, $g(\mu) = rq/p^2 =$

of the natural exponentia

ique (Bickel and Doksum,

arameter exponential fam-
ial families. Suppose $\mathbf{X}$ =
ribution P_θ where P_θ form
the family of distributions
ncy) functions, we have

$$x_i)] \, exp\,[\eta(\theta) \sum_{i=1}^{n} T(x_i) - nB(\theta)]$$

orm a one-parameter expo-

ntial family by letting the
and Doksum, 2000, p. 52).

$\in X \subset \Re^p$

e continuous case and the
If $\theta \in \Theta$, then $A(\eta)$ must
the space $\mathcal{E}$ is called the
ated by T and h. The space
alled the natural sufficient
le $X_1, \cdots, X_n$ we have

$- nA(\eta)]$.

$0, 1, \cdots; \theta > 0$,

with $h(x) = 1/x!$, $\eta(\theta) = log(\theta)$, $B(\theta) = \theta$, and $T(x) = x$, and

$$q(x, \eta) = h(x)\, exp\,[\eta T(x) - exp(\eta)] \ ,$$

where $h(x) = 1/x!$, $\eta = log(\theta)$, $A(\theta) = exp(\eta)$, and $T(x) = x$.

The re-parameterized canonical forms of the Poisson with $\theta = \lambda$, the normal with $\theta = (\mu, \sigma^2)', \mu \neq 0$, the binomial with $\theta = (n,p)'$, the gamma with $\theta = (\alpha, \beta)'$, and the negative binomial with $\theta = (r, \lambda)'$ show that in each one of the cases the statistic $\sum X_i$ is sufficient for the mean μ. While estimating the mean by $a \sum X_i$, $a \in (0,1)$, we assume that the variance $\sigma^2 = {\sigma_0}^2$ (for the normal), and n, α, and r for the other distributions are fixed.

Morris (1982) gives a characterization of the normal, Poisson, gamma, binomial and negative binomial distributions. He explains that these distributions enjoy wide applications and have many useful mathematical properties because they are natural exponential families and have quadratic variance functions. In these cases, the variance is at most a quadratic function of the mean μ. This latter relationship is given as

$$\sigma^2 = g(\mu) = \zeta_0 + \zeta_1\mu + \zeta_2\mu^2 \ ,$$

where the constants $\zeta_0, \zeta_1, \zeta_2 \in \Re$.

It can be shown that the normal $\mathcal{N}(\mu, \sigma^2)$ with $g(\mu) = \sigma^2$ ($\zeta_0 = \sigma^2$ constant variance function), $\zeta_1 = \zeta_2 = 0$), the Poisson, $\mathcal{P}(\lambda)$, $g(\mu) = \lambda$, ($\zeta_0 = 0$, $\zeta_1 = 1$ (linear), $\zeta_2 = 0$), the gamma, $\Gamma(\alpha, \beta)$, $\mu = \alpha\beta$, $g(\mu) = \alpha\beta^2$ ($\zeta_1 = 0, \zeta_2 = 1/\alpha^2$), the binomial, $\mathcal{B}(n,p)$, $\mu = np$, $g(\mu) = npq = \mu - \mu^2/n$, ($\zeta_0 = 0, \zeta_1 = 1, \zeta_2 = -1/n$)(with $q = 1 - p$) , and the negative binomial $\mathcal{NB}(r,p)$, $\mu = rq/p$, $g(\mu) = rq/p^2 = \mu(\mu + r)/r$, ($\zeta_0 = 0, \zeta_1 = 1, \zeta_2 = 1/r$) are members of the natural exponential families with quadratic variance functions. Note that the assumption that σ^2, λ, n, α, and r will remain constant while estimating the mean still holds. Based on the foregoing we now give the coefficient of variation for the respective distributions in the order they are presented above: $v = \sigma/\mu$, $\lambda^{-1/2}$, $\alpha^{-1/2}$, $q/(np)^{1/2}$ and $(rq)^{-1/2}$, respectively.

Hence, with the pertinent choice of v, the resulting improvement interval has the form

$$I(v) = \{a \in \Re_+ \mid \frac{1-v^2}{1+v^2} < a < 1, v \in (0,1)\}.$$

This interval is a special case of $I(c(n), v)$ where $c(n) = 1$. In fact, it is the interval of improvement given by Bibby and Toutenburg (1977).

Finally, having given the interval above we can conclude that those values of a from the open interval $(0,1)$ satisfying the condition in $I(v)$ can be taken as shrinkage factors; the value $v = 1$ gives a lower bound zero of the interval, and this bound is exactly zero. One particular case for which $v = 1$ is the exponential distribution. Thus, for the exponential distribution, all shrunken estimators with shrinkage factors $a \in (0,1)$ are improvements against the unbiased estimator of the mean, namely $\bar{X} = n^{-1}\sum_{i=1}^{n} X_i$. The best among these is $a^* = 1/(n+1)$. The resulting improved biased estimator $(n+1)^{-1}\sum_{i=1}^{n} X_i = (n\bar{X})/(n+1)$ has a smaller MSE than the variance of $\bar{X}$.

Applications

In this section we provide optimal shrinkage factors for the sufficient statistic $\sum_{i=1}^{n} X_i$ as an improved estimation against the unbiased estimator of the mean. Assume that $\mathbf{X} \sim \mathcal{F}(\mu, \sigma^2)$, $\mu > 0$, where $\mathcal{F}$ is any of the five distributions from the one-parameter exponential families we discussed above. We

use the general relationship
Then for any real valued a
result in each of the cases t
a is the solution of the equa

1. The Poisson distribu

Suppose $\mathbf{X} \sim \mathcal{P}(\lambda)$, $\lambda > 0$,
and $Var(a\bar{X}) = a^2\lambda/n$, MS
$2(a-1)\lambda^2 + 2a\lambda/n$ and (
$(d/da)[MSE[a\bar{X}]] = 0$, we
showing that the optimal s

2. The Exponential dist

Suppose $\mathbf{X} \sim \mathcal{E}(\lambda)$, $\lambda > 0$,
follow $E(a\bar{X}) = a/\lambda$ and
$a^2/\lambda^2 n$, $(d/da)[MSE(a\bar{X})]$
$2/\lambda^2 + 2/\lambda^2 n > 0$. The sc
$a = n/(1+n)$. Thus, the o

3. The Binomial distrib

Suppose $\mathbf{X} \sim \mathcal{B}(n,p), 0$ <
np, $Var(\mathbf{X}) = npq$; $E(a\bar{X}$
$(a-1)^2 n^2 p^2 + a^2 npq$, $(d/d$

will remain constant while
foregoing we now give the
tions in the order they are
^{2} and $(rq)^{-1/2}$, respectively.

g improvement interval has

$v \in (0,1)\}$.

$c(n) = 1$. In fact, it is the
enburg (1977).

conclude that those values
condition in $I(v)$ can be
es a lower bound zero of
particular case for which
e exponential distribution,
$\in (0,1)$ are improvements
ely $\bar{X} = n^{-1}\sum_{i=1}^{n} X_i$. The
improved biased estimator
MSE than the variance of

rs for the sufficient statis-
unbiased estimator of the
$\mathcal{F}$ is any of the five distri-
es we discussed above. We

use the general relationship, $E(\bar{X}) = \mu$, $Var(\bar{X}) = \sigma^2/n$, to derive results. Then for any real valued a it is $E(a\bar{X}) = a\mu$ and $Var(a\bar{X}) = a^2\sigma^2/n$. As a result in each of the cases the optimal shrinkage parameter $a^* = a/n$, where a is the solution of the equation $d/da[MSE(a\bar{X})] = 0$.

1. The Poisson distribution

Suppose $\mathbf{X} \sim \mathcal{P}(\lambda)$, $\lambda > 0$, that is $E(\mathbf{X}) = \lambda = Var(\mathbf{X})$. Then $E(a\bar{X}) = a\lambda$ and $Var(a\bar{X}) = a^2\lambda/n$, $MSE(a\bar{X}) = (a-1)^2\lambda^2 + a^2\lambda/n$, $(d/da)[MSE(a\bar{X})] = 2(a-1)\lambda^2 + 2a\lambda/n$ and $(d^2/da^2)[MSE(a\bar{X})] = 2\lambda^2 + 2\lambda/n > 0$. Setting $(d/da)[MSE[a\bar{X}]] = 0$, we get a solution $a = n\lambda/(1+n\lambda)$ for the equation, showing that the optimal shrunken estimator of the mean λ is

$$T^*(\mathbf{X}) = \frac{\lambda}{1+n\lambda}\sum_{i=1}^{n} X_i \ .$$

2. The Exponential distribution

Suppose $\mathbf{X} \sim \mathcal{E}(\lambda)$, $\lambda > 0$, that is $E(\mathbf{X}) = 1/\lambda$, $Var(\mathbf{X}) = 1/\lambda^2$ from which follow $E(a\bar{X}) = a/\lambda$ and $Var(a\bar{X}) = a^2/\lambda^2 n$, $MSE(a\bar{X}) = (a-1)^2/\lambda^2 + a^2/\lambda^2 n$, $(d/da)[MSE(a\bar{X})] = 2(a-1)/\lambda^2 + 2a/\lambda^2 n$ and $(d^2/da^2)[MSE(a\bar{X})] = 2/\lambda^2 + 2/\lambda^2 n > 0$. The solution of the equation $(d/da)[MSE[a\bar{X}]] = 0$ is $a = n/(1+n)$. Thus, the optimal shrunken estimator of the mean λ is

$$T^*(\mathbf{X}) = \frac{1}{1+n}\sum_{i\ 1}^{n} X_i \ .$$

3. The Binomial distribution

Suppose $\mathbf{X} \sim \mathcal{B}(n,p), 0 < p < 1$, and n is fixed. In this case $E(\mathbf{X}) = np$, $Var(\mathbf{X}) = npq$; $E(a\bar{X}) = anp$ and $Var(a\bar{X}) = a^2npq$, $MSE(a\bar{X}) = (a-1)^2 n^2p^2 + a^2npq$, $(d/da)[MSE(a\bar{X})] = 2(a-1)n^2p^2 + 2anpq$ and

$(d^2/da^2)[MSE(a\bar{X})] = 2n^2p^2+2npq > 0$. The equation $(d/da)[MSE(aX)] = 0$ has the solution $a = n^2p^2/(n^2p^2 + npq)$. Thus, the optimal shrunken estimator of the mean np is

$$T^*(\mathbf{X}) = \frac{p^2}{np^2 + pq}\sum_{i=1}^{n} X_i \ .$$

4. The Normal distribution

Suppose $\mathbf{X} \sim \mathcal{N}(\mu, \sigma^2)$, $\mu \neq 0$ and σ^2 is fixed. $E(a\bar{X}) = a\mu$ and $Var(a\bar{X}) = a^2\sigma^2/n$, $MSE(a\bar{X}) = (a-1)^2\mu^2 + a^2\sigma^2/n$; $(d/da)[MSE(a\bar{X})] = 2(a-1)\mu^2 + 2a\sigma^2/n$ and $(d^2/da^2)[MSE(a\bar{X})] = 2\mu^2 + 2\sigma^2/n > 0$. Solving $(d/da)[MSE(a\bar{X})] = 0$ for a gives the solution $a = n\mu^2/(n\mu^2 + \sigma^2)$ giving the optimal shrunken estimator

$$T^*(\mathbf{X}) = \frac{\mu^2}{n\mu^2 + \sigma^2}\sum_{i=1}^{n} X_i$$

for μ.

5. The Gamma distribution

Suppose $\mathbf{X} \sim \mathcal{G}(\alpha, \beta)$, $\alpha, \beta > 0$. We assume that the scale parameter α is held constant. From $E(\mathbf{X}) = \alpha\beta$, $Var(\mathbf{X}) = \alpha\beta^2$ follows $E(a\bar{X}) = a\alpha\beta$ and $Var(a\bar{X}) = a^2\alpha\beta^2/n$, $MSE(a\bar{X}) = (a-1)^2\alpha^2\beta^2 + a^2\alpha\beta^2/n$. Hence $(d/da)[MSE(a\bar{X}] = 2(a-1)\alpha^2\beta^2 + 2a\alpha\beta^2/n$ and $(d^2/da^2)[MSE([a\bar{X})] = 2\alpha^2\beta^2 + 2\alpha\beta^2/n > 0$. We get the solution $a = n\alpha/(1+n\alpha)$ for the equation $(d/da)[MSE(a\bar{X})] = 0$ thereby giving the shrunken estimator

$$T^*(\mathbf{X}) = \frac{\alpha}{1 + n\alpha}\sum_{i=1}^{n} X_i$$

as the optimal estimator of the mean $\alpha\beta$ in the sense of MSE. Observe that if $a = 1$ we have the exponential distribution in which case the optimal shrinkage is attained at $1/(1+n)$as shown above.

6. The Negative Binomia

Suppose $\mathbf{X} \sim \mathcal{NB}(r,p)$, r =
that the parameter r (numb
distribution $E(\mathbf{X}) = rq/p$,
a^2rq/np^2, $MSE(a\bar{X}) = (a -$
$2(a-1)r^2q^2/p^2 + 2arq/np^2$, a
0. The first derivative set ec
The minimum MSE estimat

T^*

It is known that the geom
tive binomial with $r = 1$ fo
estimator of the mean q/p i

T

We note that the shrinkage

In conclusion we have so f
optimal shrinkage factors a
Poisson ($v^2 = 1/\lambda$), expon
($v^2 = \sigma^2/\mu^2$), gamma ($v^2 =$
course of the sample size n
to zero in which case it bec
large a_{opt} is useful provided
with the exception of the e

6. The Negative Binomial distribution:

Suppose $\mathbf{X} \sim \mathcal{NB}(r,p)$, $r = 1,2,\cdots;\ p \in (0,1)$. Here the assumption is that the parameter r (number of degrees of freedom) is held fixed. For this distribution $E(\mathbf{X}) = rq/p$, $Var(\mathbf{X}) = rq/p^2$, $E(a\bar{X}) = arq/p$, $Var(a\bar{X}) = a^2rq/np^2$, $MSE(a\bar{X}) = (a-1)^2r^2q^2/p^2 + a^2rq/np^2$. Thus $(d/da)[MSE(a\bar{X})] = 2(a-1)r^2q^2/p^2 + 2arq/np^2$, and $(d^2/da^2)[MSE(a\bar{X})] = 2r^2q^2/p^2 + 2rq/np^2 > 0$. The first derivative set equal to zero has the solution $a = nrq/(nrq+1)$. The minimum MSE estimator of the mean $r(1-p)/p$ is given by

$$T^*(\mathbf{X}) = \frac{rq}{nrq+1}\sum_{i=1}^{n} X_i \ .$$

It is known that the geometric distribution is a special case of the negative binomial with $r = 1$ for which we obtain the minimum MSE shrunken estimator of the mean q/p is

$$T^*(\mathbf{X}) = \frac{q}{nq+1}\sum_{i=1}^{n} X_i \ .$$

We note that the shrinkage factors we have provided can be written as

$$a_{opt} = \frac{1}{n+v^2} \ .$$

In conclusion we have so far noted the following. First, the values of the optimal shrinkage factors a_{opt} are functions of the distribution parameters - Poisson ($v^2 = 1/\lambda$), exponential ($v^2 = 1$), binomial ($v^2 = q/np$), normal ($v^2 = \sigma^2/\mu^2$), gamma ($v^2 = 1/\alpha$), and negative binomial ($v^2 = 1/rq$), and of course of the sample size n. For large n the optimal shrinkage factor tends to zero in which case it becomes unattractive. For sample sizes that are not large a_{opt} is useful provided at the same time v^2 is not very large. Secondly, with the exception of the exponential distribution, all results depend on the unknown parameters because the MSE is itself a function of the distribution parameters. And as such they are relevant from the theoretical point of view. However, whenever prior information about the size of v^2 is available the shrunken estimator with minimum MSE is determined precisely. In that case the shrunken minimum MSE estimator becomes the operational substitute for the unbiased estimator $\bar{X}$, especially for small sample sizes, although $\bar{X}$ is UMVUE within the class of linear unbiased estimators of the mean.

Acknowledgement: The authors are thankful for the support they received from the German Academic Exchange Service (DAAD) during their research visit in Germany.

References

Arnholt AT, Hebert JL (1995) Estimating the mean with known coefficient of variation. *The American Statistician* **49**, 367-369.

Bibby J (1972) Minimum mean square error estimation, ridge regression and some unanswered questions. *Progress in Statistics* **1**, 107-121.

Bibby J, Toutenburg H (1977) *Prediction and Improved Estimation in Linear Models.* Wiley, Chichester.

Bibby J, Toutenburg H (1978) Improved estimation and prediction. *Zeitschrift für angewandte Mathematik und Mechanik* **58**, 45-49.

Bickel JP, Doksum KA (2000) *Mathematical Statistics, Basic Ideas and Selected Topics*, Vol.1 (2nd Ed.). Prentice-Hall, New Jersey.

Gleser LJ, Healy JD (19
tion with known coeffi
tistical Association **71**

Khan RA (1968) A note
with known coefficient
Association **63**, 1039-1

Kleffe J (1985) Some rem
tiplication with a cons
T. and Klonecki, W.,

Morris CN (1982) Natur
functions. *The Annals*

Searls DT (1964) The uti
estimation procedure.
59, 1225-1226.

function of the distribution
he theoretical point of view.
size of v^2 is available the
nined precisely. In that case
the operational substitute
ll sample sizes, although $\bar{X}$
timators of the mean.

or the support they received
AAD) during their research

the mean with known coef-
ian **49**, 367-369.

estimation, ridge regression
in Statistics **1**, 107-121.

nd Improved Estimation in

imation and prediction. *Zeitschrift*
58, 45-49.

l Statistics, Basic Ideas and
Hall, New Jersey.

Gleser LJ, Healy JD (1976) Estimating the mean of a normal distribution with known coefficient of variation. *Journal of the American Statistical Association* **71**, 977-981.

Khan RA (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.

Kleffe J (1985) Some remarks on improving unbiased estimators by multiplication with a constant. *Linear Statistical Inference*. Ed.: Calinski, T. and Klonecki, W., 150-161. Cambridge, Massachussets.

Morris CN (1982) Natural exponential families with quadratic variance functions. *The Annals of Statistics* **10**, 65-80.

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