Deriving Shrunken Estimators for the Variance in One-Parameter Natural Exponential Families

A. Laheetharan[1] and P. Wijekoon[2*]

[1]*Department of Mathematics & Statistics, Faculty of Science, University of Jaffna*
[2]*Department of Statistics & Computer Science, Faculty of Science, University of Peradeniya*

Introduction

Exponential family of distributions is an important class of distributions in statistics. Estimation of mean and variance of the distributions of this class is one of the problems that the researchers are involved. Several Statisticians have considered the estimation of mean in one parameter exponential families when coefficient of variation is known, and an improved method in this regard was given by Wencheko and Wijekoon (2005). Morris (1982) and Letac and Mora (1990) showed that the variance of some basic types of distributions belong to the one parameter exponential family is a polynomial function of the mean μ with degree less than or equal to 3. The objective of this study is to derive an optimal shrunken estimator for these variance functions.

One parameter exponential family and variance functions

Consider the random vector $X = (X_1, \ldots, X_n)$ whose probability density function is a function of the parameter θ, where $\theta \in \Theta$ for some interval $\Theta \subseteq \Re$. Then the family of distributions of a model $\{P_\theta : \theta \in \Theta\}$ is said to be a one parameter exponential family, if there exists real valued functions, $\eta(\theta)$, $B(\theta)$, $T(x)$ and $h(x)$ such that the probability density (or mass) function $f(x;\theta)$ of P_θ can be written as

$$f(x;\eta) = h(x)\exp[\eta(\theta)T(x) - B(\theta)] \qquad (1)$$

The parameter $\eta(\theta)$ is called the natural parameter of the distribution, and $T(x)$ is called the natural statistic. The family of distributions obtained by taking iid samples from one-parameter exponential families are themselves one-parameter exponential families. A useful reparametrization of the exponential family can be obtained by setting $\eta = \eta(\theta)$. Then the exponential family has the form

$$f\backslash(x;\eta) = h(x)\exp[\eta T(\underline{X}) - A(\eta)], \qquad (2)$$

where, $T(\underline{X}) = \sum_{i=1} T(x_i)$ is a sufficient statistic, and $A(\eta) = \log \int \ldots \int h(x)\exp[\eta T(\underline{X})]dx$ in the continuous case and the integral is replaced by a sum in the discrete case.

Therefore, $E(T(\underline{X})) = A'(\eta)$, $Var[T(\underline{X})] = A''(\eta)$, and the moment-generating function $\psi(s) = \exp[A(s+\eta) - A(\eta)]$ for s in some neighborhood of 0, if it exists.

Morris (1982) has shown that exactly six basic types of natural exponential families (NEFs) have quadratic variance function (QVF). He considered normal, Poisson, gamma, binomial, negative binomial and the NEF generated by the generalized hyperbolic secant (GHS) distributions. However, Letac and Mora (1990) have given variance functions for the above six plus another six distributions that are inverse Gaussian, Abel, Takacs, strict arcsine, large arcsine, and Ressel, which have cubic variance function (CVF). In both these cases, it is clear that the variance is a polynomial function of the mean μ with degree less than or equal to 3, which can be described by

$$V_F(\mu) = \xi_0 + \xi_1\mu + \xi_2\mu^2 + \xi_3\mu^3 \qquad (3)$$

where, $\xi_0, \xi_1, \xi_2, \xi_3 \in \Re$ are constants.

Methodology and results

Gleaser and Hearly (1976) have considered the minimization of mean square error of linear combination of two uncorrelated and unbiased estimators of θ having a known coefficient of variation. In the following theorem, their results were generalized by considering one estimator for the parametric function $g(\theta)$ when the ratio $\nu^2 = [g(\theta)]^{-2} Var[T(\underline{X})]$ is independent of θ. Note that in this case, the estimator $T(\underline{X})$ is not necessarily an unbiased estimator for $g(\theta)$.

Theorem 1

Let $\underline{X} = (X_1, \ldots, X_n)$ be a random sample from a population with distribution $f(x;\theta)$ and $g(\theta)$ be a real-valued function on Θ. Let $T(\underline{X})$

be a point estimator of $g(\theta)$ with $E[T(\underline{X})] = kg(\theta)$ where $k \in \Re$, and without loss of generality, assume that $k > 0$. If the ratio $v^2 = [g(\theta)]^{-2} Var[T(\underline{X})]$ is independent of θ, then $T^*(\underline{X}) = \alpha^* T(\underline{X})$ has uniformly minimum mean squared error (in $g(\theta)$) among all estimators that are in the class $C_T(\alpha) = \{\alpha T(\underline{X}) | 0 < \alpha < \infty\}$, where $\alpha^* = k/(k^2 + v^2)$.

Proof: Since $E[T^*(\underline{X})] = \alpha kg(\theta)$, and $Var[T^*(\underline{X})] = \alpha^2 Var[T(\underline{X})]$, the mean squared error (MSE) of $T^*(\underline{X})$ is

$$MSE[T^*(\underline{X})] = \alpha^2 Var[T(\underline{X})] + (\alpha k - 1)^2 [g(\theta)]^2.$$

Hence, it can be easily shown that the estimator $T^*(\underline{X}) = k(k^2 + v^2)^{-1} T(\underline{X})$ minimizes MSE over $C_T(\alpha)$, and the minimum MSE is given by

$$MSE[T^*(\underline{X})] = v^2 (k^2 + v^2)^{-1} [g(\theta)]^2$$

Then it can be shown that

$$MSE[T^*(\underline{X}) < MSE[T(\underline{X})].$$

Therefore, it is clear that the estimator $T^*(\underline{X})$ is uniformly better than the estimator $T(\underline{X})$, and that $T(\underline{X})$ is inadmissible, and more generally $T^*(\underline{X})$ is uniformly better than any other estimators in the class $C_T(\alpha)$.

The above theorem can be applied for estimation of certain variance functions $V_F(\mu)$, for which the estimator is of the form $T(\underline{X})$ with $v^2 = c(n) k^2 \tau^2$, where $c(n)$, is a known constant depends on n, and τ is the coefficient of variation of the probability distribution which is assumed to be known.

Applications

The following applications illustrate the use of the above results.

Binomial distribution: Let $X_i \sim B(r, p), i = 1 \cdots n, 0 < p < 1, and\ r$ be fixed. In this case for an iid sample of B(r,p) one can identify

$$\eta = \log\left(\frac{p}{1-p}\right), \quad p = \frac{\exp(\eta)}{1+\exp(\eta)}, \quad B(p) = -nr\log(1-p), \quad A(\eta) = nr\log[1+\exp(\eta)].$$

Then the mean and variance for the natural sufficient statistic $T(\underline{X}) = \sum_{i=1}^{n} X_i$ is $A'(\eta) = nrp$ and $A''(\eta) = nrp(1-p)$ respectively. Then $\alpha = 1/r(n + \tau^2)$, where τ is the coefficient of variation of the distribution, which is assumed to be known, and $v^2 = nr^2\tau^2$. Thus, according to Theorem 1, the optimal shrunken estimator of the parameter p is $T^*(\underline{X}) = \sum_{i=1}^{n} X_i / r(n + \tau^2)$.

Hence, $\sum_{i=1}^{n} X_i / (n + \tau^2)$ is the optimal shrunken estimator of the mean $\mu = rp$, which is precisely the estimator proposed by Wencheko and Wijekoon (2005).

Using the derivatives of,

$$\Psi_T(t) = \exp[A(\eta + t) - A(\eta)] = ([1 + \exp(\eta + t)]/[1 + \exp(\eta)])^{nr}$$

which is the moment generating function of $T(\underline{X})$ and Theorem 1, it can be shown that the optimal shrunken estimator of the function $p(1-p)$ is

$$T^*(\underline{X}) = r\sum_{i=1}^{n} X_i \left(nr - \sum_{i=1}^{n} X_i\right) / (nr-1)(r\tau + \tau^{-1})^2 + r(nr-2)(nr-3)$$

Note that in this case $k = nr(nr-1)$ and $g(p) = p(1-p)$. Hence,

$$r\sum_{i=1}^{n} X_i \left(nr - \sum_{i=1}^{n} X_i\right) / (nr-1)(r\tau + \tau^{-1})^2 + r(nr-2)(nr-3)$$ is

the optimal shrunken estimator of the variance $V_F(\mu) = rp(1-p)$. Similarly, the optimal shrunken estimator of the variance functions of some basic types of NEF"s are given in the following Table 1.

Conclusions

When the coefficient of variation τ of the distribution is known the above admissible shrunken estimators of the variance functions have minimum mean squared errors in the class $C_T(\alpha) = \{\alpha T(\underline{X}) | 0 < \alpha < \infty\}$.

References

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

Letac, G. and Mora, M. (1990) Natural exponential families with cubic variance functions, *The Annals of Statistics,* 18, 1-37.

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

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

Table 1. Optimal shrunken estimators of the variance functions of several distributions, which belong to NEFs

	Name of the type	$V_F(\mu)$
1	Normal $N(\mu, \sigma^2)$	$\sum_{i=1}^{n}(X_i - \overline{X}_n)^2 / (n+1)$
2	Poisson Poiss (λ)	$\sum_{i=1}^{n} X_i / (n + \tau^2)$
3	Negative binomial $NB(r, p)$	$r^2 \sum_{i=1}^{n} X_i \left(nr + \sum_{i=1}^{n} X_i \right) / (nr+1)(r\tau - \tau^{-1})^2 + r(nr+2)(nr+3)$
4	Gamma $Gam(r, \lambda)$	$r\left(\sum_{i=1}^{n} X_i\right)^2 / (nr+2)(nr+3)$