JIRSS (2005)
Vol. 4, No. 2, pp 87-95

Estimation Based on an Appropriate Choice of Loss Function

M. Ganjali, K. Shafie

Department of Statistics, Faculty of Mathematical Sciences, Shahid Beheshti University, Tehran, Iran. (m-ganjali@sbu.ac.ir, k-shafie@sbu.ac.ir)

Abstract. Some examples of absurd uniformly minimum variance unbiased estimators are discussed. Two reasons, argued in the literature, for having such estimators are lack of enough information in the available data and property of unbiasedness. In this paper, accepting both of these views, we show that an appropriate choice of loss function using a general concept of unbiasedness leads to risk unbiased, admissible and reasonable estimators. For this we extend the Rao-Blackwell theorem using a new way of defining unbiased estimator.

1 Introduction

The possibility of an absurd mean-unbiased estimator which may be the uniformly minimum variance unbiased estimator (UMVUE) is illustrated by Kendall and Stuart (1979, Ex. 17.26), Lehmann (1983) and Romano and Siegel (1986, Ch.9) in the following example.

Received: May 2005
Key words and phrases: Admissibility, loss function, risk Function, unbiased estimator, uniformly minimum variance unbiased estimator.

Example 1.1. Let X have the Poisson distribution with θ as its mean. For the estimand $g(\theta) = \exp(-b\theta)$, the estimator $\delta(x) = (1-b)^x$ is the UMVUE, but this estimator is absurd for $b > 1$, since it is negative for odd x and does not preserve the range of estimand.

Looking at the large maximum bias of the maximum likelihood estimator (MLE), $\delta_1(x) = \exp(-bx)$, for the estimand $g(\theta) = \exp(-b\theta)$ in example 1.1, Lehmann (1983) concludes that the available information is inadequate for the existence of reasonable unbiased estimator or MLE. Meeden (1987) argues that the absurd estimator in example 1.1 arises because unbiasedness requires the estimator to be correct on the average (property of unbiasedness.) In other words, the concept of unbiasedness needs to be generalized to incorporate the loss function in use. This generalization, called loss-unbiasedness or risk unbiasedness, is presented by Lehmann (1951), Klebanov (1974) and Noorbaloochi and Meeden (1983). Klebanov (1974) considers possibility of generalizing Rao-Blackwell theorem for the risk unbiased estimators. Parsian and Farsipour (1999) use LINEX loss to compare several estimators for the mean of the selected population. They use the general concept of unbiasedness to obtain some of their estimators. Parsian and Kirmani (2002) also discuss the use of the LINEX loss function as a criterion of estimation.

The purpose of this paper, using the general definition of unbiasedness of Lehmann (1951), is to show that the absurd estimator in example 1.1 is obtained because of inappropriate choice of the loss function. We shall choose a suitable loss function and show that the MLE has some optimal properties that make it reasonable to be used. In fact, the appropriate choice of loss function gives some intuitive information about the parametric function of interest. It also leads us to think about the property of unbiasedness in its general form (Lehmann, 1951).

In Section 2, we review the general definition of unbiasedness (Lehmann, 1951, Klebanov, 1974, Noorbaloochi and Meeden, 1983) and other necessary concepts of decision theory. In Section 3, we shall give some useful theorems which show the optimal properties of our estimators. Some more examples are illustrated in Section 4 and we conclude in Section 5.

2 Basic definitions

Lehmann (1951 and 1959) gives the following definition for unbiasedness.

Definition 2.1. The estimator $\delta(x)$ is said to be L-unbiased for the estimand $g(\theta)$ if

$$E_\theta\{L[g(\theta),\delta(X)]\} = \min_{\theta'} E_\theta\{L[g(\theta'),\delta(X)]\} \quad \text{for all } \theta \in \Theta, \tag{1}$$

where $L[g(\theta),\delta(x)]$ is the loss of estimating $g(\theta)$ by $\delta(x)$.

In this definition of unbiasedness, one needs to have the loss of estimating the estimand $g(\theta)$ by the estimator $\delta(x)$. We shall use the loss function

$$L[g(\theta),\delta(x)] = [m(\delta(x)) - m(g(\theta))]^2, \tag{2}$$

where $m(.)$ is a strictly monotone and continuous function and we shall call the loss in (2)m-loss. For instance, by ln-loss we mean

$$L[g(\theta),\delta(x)] = [\ln(\delta(x)) - \ln(g(\theta))]^2. \tag{3}$$

For a given $m(.)$ in (2), we shall call the estimator obtained using the following theorem an m-unbiased estimator.

Theorem 2.1. *The estimator $\delta(x)$ is m-unbiased for the estimand $g(\theta)$ if*

$$E_\theta m(\delta(X)) = m(g(\theta)) \quad \text{for all } \theta \in \Theta, \tag{4}$$

where m is a strictly monotone and continuous function.

The proof of this theorem is similar to the main theorem of Lehmann (1951). Note also that for $m(x) = x$ this theorem gives the mean-unbiased estimator.

The risk for the estimator $\delta(x)$ corresponding to the m-loss is defined to be

$$R(\theta,\delta) = E_\theta[m(\delta(X)) - m(g(\theta))]^2. \tag{5}$$

An estimator δ is said to be better than estimator η if, for all θ

$$R(\theta,\delta) \leq R(\theta,\eta), \tag{6}$$

with strict inequality for some $\theta_0 \in \Theta$ and δ is said to be admissible in the class of estimators Δ if it is not worse than any other estimator in Δ (Lehmann and Casella, 1998). As an example, consider example 1.1 with $m(x) = x$. The estimator $\delta^*(x) = \max((1-b)^x, 0)$; using squared error loss function, has uniformly smaller mean squared error than UMVUE $(1-b)^x$ for $b > 1$. Hence, δ is inadmissible with respect to squared error loss function (Romano and Siegel, 1986, Ch. 9)

We shall show that log-unbiased estimator for $g(\theta)$ in example 1.1 is the uniformly minimum risk unbiased estimator (MRUE) and admissible. The log function is an appropriate choice for estimating $g(\theta)$ in this example because it forces the estimator $\delta(x)$ to have a measurement scale like that of $g(\theta)$ (exponential).

3 Main Results

The following theorem extends Rao-Blackwell theorem to m-unbiased estimators.

Theorem 3.1. *Suppose $\{F_\theta : \theta \in \Theta\}$ is a family of probability distributions and δ is an m-unbiased estimator for $g(\theta)$ where $E_\theta[m(\delta(x))]^2 < \infty$. Suppose T is a sufficient statistic for $\{F_\theta : \theta \in \Theta\}$. Then, $\eta(t) = m^{-1}E[m(\delta(x)) \mid t]$ is an m-unbiased estimator for $g(\theta)$ and*

$$R(\theta, \eta) \quad \leq \quad R(\theta, \delta).$$

The proof of this theorem is similar to that of Rao-Blackwell theorem (see Lehmann and Casella, 1998) and is omitted.

If T is complete for its family of distributions, then $\eta(t)$ given in theorem 3.1 is uniformly minimum risk m-unbiased estimator (UMRUE) for $g(\theta)$. Note that for $m(x) = x$, $\eta(t)$ is uniformly minimum variance unbiased estimator (UMVUE).

Theorem 3.2. *The estimator $\delta(x)$ is UMVUE for $g(\theta)$ iff $m^{-1}(\delta(x))$ is UMRUE for $m^{-1}(g(\theta))$ with respect to the m-loss.*

Proof. Suppose $\delta(x)$ is UMVUE for $g(\theta)$ then $m^{-1}(\delta(x))$ is m-unbiased for $m^{-1}(g(\theta))$ with respect to m-loss. Let $\delta^*(x)$ be another m-unbiased estimator for $m^{-1}(g(\theta))$. Then, $m(\delta^*(x))$ is unbiased for $g(\theta)$. Let denote the risk of δ for estimating g when the loss function is the $m-$loss by $R(\theta, \delta)$. As $\delta(x)$ is UMVUE for $g(\theta)$ for all θ, we have

$$\begin{aligned} R(\theta, \delta^*) &= E_\theta[m(\delta^*(X)) - g(\theta)]^2 \\ &\geq E_\theta[\delta(X) - g(\theta)]^2 \\ &= E_\theta[m(m^{-1}(\delta(X))) - m(m^{-1}(g(\theta)))]^2 \\ &= R(\theta, m^{-1}\delta) \end{aligned}$$

So, $m^{-1}\delta(x)$ is UMRUE for $m^{-1}(g(\theta))$. The other direction of the theorem can be proved in a similar manner. ■

Example 3.1. (1, continued) As $E_\theta(X) = \theta$, thus X is UMVUE. So, if $m(x) = ln(x)$, $\delta(x) = \exp(-bX)$ is ln-unbiased estimator and in fact, the UMRUE for $\exp(-b\theta)$ with respect to the ln-loss.

Lehmann (1983) does not accept $\delta(x) = \exp(-bX)$ as a reasonable estimator because of its large bias with respect to squared error loss. However, it should be noted that this estimator is unbiased (i.e. with 0 absolute bias) with respect to ln-loss and as mentioned before this loss seems more appropriate for estimating $\exp(-b\theta)$.

Theorem 3.3. *The estimator $\delta(x)$ is admissible for $g(\theta)$ with respect to squared error loss function iff $m^{-1}(\delta(x))$ is admissible for $m^{-1}(g(\theta))$ with respect to m-loss.*

Proof. Suppose $\delta(x)$ is admissible for $g(\theta)$ with respect to squared error loss, but $m^{-1}(\delta(x))$ is not admissible for $m^{-1}(g(\theta))$ with respect to m-loss. Then, there exist δ^* such that

$$\begin{aligned} &E_\theta[m(\delta^*(X)) - m(m^{-1}(g(\theta)))]^2 \\ &\quad \leq E_\theta[m(m^{-1}(\delta(X))) - m(m^{-1}(g(\theta)))]^2 \text{ for all } \theta, \end{aligned} \tag{7}$$

and the inequality in (7) is strict for at least one θ. This gives

$$E_\theta[m(\delta^*(X)) - g(\theta)]^2 \leq E_\theta[\delta(X) - g(\theta)]^2 \text{ for all } \theta, \tag{8}$$

and strict inequality for at least one θ. So, $m(\delta^*(X))$ is better than $\delta(X)$ with respect to squared error loss. This is a contradiction. The other side of the theorem can be proved in a similar manner. ■

Remark. A similar theorem can be established for minimaxity of an estimator.

Example 3.2. (1, continued) As X is admissible for θ with respect to squared error loss (see Lehmann and Casella, 1998), $\exp(-bX)$ is admissible for $\exp(-b\theta)$ with respect to ln-loss.

The following theorem emphasizes that if m is strictly convex, then m-unbiased and mean-unbiased estimators for a parametric function would not be the same.

Theorem 3.4. *Suppose m is a monotone and strictly convex function. Neither is the m-unbiased estimator $\delta(x)$ for $g(\theta)$ a mean-unbiased estimator nor the mean-unbiased estimator for $g(\theta)$ is the m-unbiased estimator.*

Proof. If $\delta(x)$ is m-unbiased for $g(\theta)$, then

$$E_\theta m(\delta(X)) = m(g(\theta)) \text{ for all } \theta. \tag{9}$$

Since m is strictly convex

$$m(E_\theta \delta(X)) < E_\theta m(\delta(X)) \ , \tag{10}$$

which implies $E_\theta \delta(X) \neq g(\theta)$.

This means $\delta(x)$ is not unbiased for $g(\theta)$. On the other hand. If $E_\theta \delta(x) = g(\theta)$, we have

$$m(g(\theta)) < E_\theta m(\delta(X)), \tag{11}$$

which means $\delta(x)$ is not m-unbiased for $g(\theta)$. ■

Theorem 3.4 can be extended for two strictly monotone convex functions (m_1 and m_2) which says $\delta(x)$ can not be both m_1-unbiased and m_2-unbiased.

4 Examples

Example 4.1. Suppose the distribution of X is $Bin(n, \theta)$. A mean-unbiased estimator for $g(\theta) = \frac{1}{\theta+1}$ dose not exist. The squared error loss function does not, however, incorporate the change of measurement scale. A more appropriate loss function, where the action space is $(0, 1)$, would be a $\frac{1}{x}$-loss. As $\frac{X}{n} + 1$ is UMVUE for $\theta + 1$, $\frac{n}{X+n}$ is a reasonable and UMRUE for $\frac{1}{\theta+1}$ with $\frac{1}{X}$-loss. As $\frac{X}{n} + 1$ is admissible for $\theta + 1$ with respect to squared error loss, then $\frac{n}{X+n}$ is admissible for $\frac{1}{\theta+1}$ with respect to $\frac{1}{X}$-loss.

Example 4.2. Suppose $X+n$ is the number of Bernouilli trials required to obtain n successes for $n>1$. So X has the following negative binomial distribution

$$Pr(X=x)=\binom{n+x-1}{n-1}p^n q^x \quad x=0,1,..., \tag{12}$$

the UMVU estimator $\delta(x)$ for p is given by

$$\delta(x)=\frac{\binom{n+x-2}{x}}{\binom{n+x-1}{n-1}}=\frac{n-1}{n+x-1}. \tag{13}$$

Then the $\sqrt[M]{x}$-unbiased estimator for p^M is

$$[\delta(x)]^M=(\frac{n-1}{n+x-1})^M, \tag{14}$$

which is UMRU estimator with respect to $\sqrt[M]{x}$-loss. It should be noted that this estimator is between 0 and 1 and hence preserves the range of estimand. On the other hand, UMVUE for $p^M(M>n)$ is absurd (see Lehmann, 1983).

Example 4.3. Let X have exponential density

$$P_\theta(x)=\beta(\theta)\exp(\theta T(x)) \quad \theta,\, T \text{ real-valued} \tag{15}$$

where θ is the canonical parameter. Suppose that the natural parameter space of θ is real line. As it can be found in Lehmann and Casella (1998, corollary 2.18, page 336), T is admissible for $E_\theta T$ with squared error loss. So, $g(T)$ is admissible for $g(E_\theta T)$ with $g^{-1}-$loss. We can call the function $g(.)$ a link function as in generalized linear models (McCullagh and Nelder, 1989). For instance for a random sample of Binomial family with success probability p Logit of T is admissible for logit of p with $g^{-1}-$loss where $g^{-1}(x)=\frac{\exp(x)}{1+\exp(x)}$ and $T(\mathbf{x})=\frac{\sum X_i}{n}$.

5 Conclusion

Although the lack of information to estimate some of the estimands in this paper, as argued by Lehmann (1983), remains valid, the change of loss function helped us to find better estimators with some optimal properties using available information. We believe the m-loss discussed in this note can increase our subjective information about the statistical problem. The m-loss also leads us to use the property of unbiasedness for estimands and estimators which are in the same scale of measurement. If one is not sure about the choice of m, one can use a class of m-loss functions, for different m, to see the robustness of the results to the choice of different elements of the class (see, Dey, Lou and Bose, 1998.).

Acknowledgments

We are grateful to Professor S. Noorbaloochi for his comments on the first draft of the paper. We are also grateful to the Editor and two referees for their comments that improved the paper.

References

Dey, D. K., Lou, K., and Bose, S. A. (1998), Bayesian approach to loss robustness. Statistics & Decisions, **16**, 65-87.

Lehmann, E. L. (1951), A general concept of unbiasedness. Ann. Math. Statis, **22**, 587-592.

Lehmann, E. L. (1959), Testing statistical hypotheses. New York: John Wiley.

Lehmann, E. L. (1983), Estimation with inadequate information. Journal of the American Statistical Association, **78**, 624-627.

Lehmann, E. L. and Casella, G. (1998), Theory of point estimation. Springer-Verlag.

Klebanov, L. B. (1974), Unbiased estimators and sufficient statistics. Theory of Prob. and its Applications, **19**(2), 379-383.

Kendall, M. and Stuart, A. (1979), The advanced theory of statistics. Vol. 2, 4th ed., New York: McMillan.

McCullagh, P. and Nelder, J. A. (1989), Generalized linear models. Chapman & Hall.

Meeden, G. (1987), Estimation when using a statistics that is not sufficient. American Statistician, **41**(2), 135-136.

Noorbaloochi, S. and Meeden, G. (1983), Unbiasedness as the Dual of Being Bayes. Journal of the American Statistical Association, **78**, 619-623.

Parsian, A. and Farsipour, N. S. (1999), Estimation of the mean of the selected population under asymmetric loss function. Metrika, **50**, 89-107.

Parsian, A. and Kirmani, S. N. U. A. (2002), Estimation under LINEX loss function. Handbook of applied econometrics and statistical inference, 53-76, Statist. Textbooks Monoger., 165 Dekker, New York.

Romano, J. P. and Siegel A. F. (1986), Counterexamples in probability and statistics. The Wadsworth and Brooks.