Kumamoto J. Math.
Vol. 13, 67-74 March (2000)

Median unbiasedness and Pitman's measure of closeness in a prediction problem

Yushan Xiao and Yoshikazu Takada

(Received December 22, 1999)

1. Introduction

When comparing the performance of estimators, the Pitman's measure of closeness (PMC) is a useful criterion, to which considerable attention has been devoted. See Rao et al. [3] and reference therein. Ghosh and Sen [2] found an interesting role of median unbiasedness (MU) in the context of PMC, that is, a MU estimator is the Pitman-closest within an appropriate class of estimators with respect to the squared error loss.

Datta [1] applied the concepts of PMC and MU to a prediction and obtained a analogous result under the squared error loss. See also Takada [5].

The purpose of this paper is to show that the result of Datta [1] hold under not only squared error loss but also LINEX loss. The LINEX loss was proposed by Varian [6] for problems in which it is appropriate to consider asymmetric loss functions. Zellner [8] showed that the sample mean is inadmissible for estimating the mean of a univariate normal distribution with respect to the LINEX loss. See also Safie and Noorballochi [4]. Xiao [7] considered the LINEX loss in a prediction problem.

In Section 2 a MU predictor is shown to be the Pitman-closest within an appropriate family of predictors not only for the squared error loss but also for the LINEX loss. In Section 3 some examples are given.

2. Pitman-closest predictor

Suppose that X is an observable random vector, Y an unknown random variable and that the joint distribution of X and Y depends on unknown parameter θ. After observing X, we want to predict the value of Y.

Let $L(d, y)$ be the loss of predicting $Y=y$ by d. For two predictors δ_1 and δ_2, δ_1 is said to be better than δ_2 under PMC with respect to L if for all θ

$$P_\theta\{L(\delta_1(X), Y) \leq L(\delta_2(X), Y)\} \geq \frac{1}{2}.$$

Let C be a family of predictors. Then $\delta \in C$ is said to be the Pitman-closest in C with respect to L if δ is better than any other $\delta' \in C$ under PMC with respect to L. A predictor δ is said to be MU of Y if for all θ

$$P_\theta\{\delta(X) \leq Y\} = P_\theta\{\delta(X) \geq Y\}.$$

In order to discuss the role of a MU predictor, we consider two loss functions. One is $L_1(d, y) = (d-y)^2$ (squared error loss) and the other is

$$L_2(d, y) = \exp[\alpha(d-y)] - \alpha(d-y) - 1$$

(LINEX loss) with $\alpha \neq 0$. Let δ_M be a MU predictor of Y and T a statistic based on X. We consider such a family of predictors that

$$C = \{\delta;\ \delta(X) = \delta_M(X) + Z(T)\},$$

where $Z = Z(T)$ is any function of T. It follows from Theorem 1 of Datta [1] that δ_M is Pitman-closest in the family C with respect to the squared error loss. The following theorem shows that δ_M is also Pitman-closest in the same family with respect to the LINEX loss. Although the first part of theorem is an univariate version of Theorem 1 of Datta [1], the proof is added for the sake of completeness.

Theorem. *Suppose that $Y - \delta_M$ is independent of T. Then δ_M is Pitman-closest in the family C with respect to the squared error and LINEX losses.*

Proof. Let $\delta = \delta_M + Z$ be any predictor in C. First we consider the squared error loss L_1. Then

$$\begin{aligned} P_\theta\{L_1(\delta_M, Y) \leq L_1(\delta, Y)\} &= P_\theta\{Z^2 + 2Z(\delta_M - Y) \geq 0\} \\ &\geq P_\theta\{Z(\delta_M - Y) \geq 0\} \\ &= P_\theta\{\delta_M - Y \geq 0, Z > 0\} \\ &\quad + P_\theta\{\delta_M - Y \leq 0, Z < 0\} \\ &\quad + P_\theta(Z = 0). \end{aligned} \tag{2.1}$$

Since Z is independent of $\delta_M - Y$ and δ_M is MU,

$$P_\theta\{\delta_M - Y \geq 0, Z > 0\} \geq \frac{1}{2} P_\theta(Z > 0) \tag{2.2}$$

and

$$P_\theta\{\delta_M - Y \leq 0, Z<0\} \geq \frac{1}{2} P_\theta(Z<0). \tag{2.3}$$

Substituting (2.2) and (2.3) into (2.1), we have

$$\begin{aligned} P_\theta\{L_1(\delta_M, Y) \leq L_1(\delta, Y)\} &\geq \frac{1}{2}\{P_\theta(Z>0)+P_\theta(Z<0)\}+P_\theta(Z=0) \\ &\geq \frac{1}{2}. \end{aligned}$$

Hence the first part of the theorem is proved.

Next we consider the LINEX loss L_2. Then

$$\begin{aligned} &P_\theta\{L_2(\delta_M, Y) \leq L_2(\delta, Y)\} \\ &\quad = P_\theta\{\exp[a(\delta_M - Y)](\exp(aZ)-1) - aZ \geq 0\} \\ &\quad = P_\theta\{a(\delta_M - Y) \geq W, aZ>0\} \\ &\qquad + P_\theta\{a(\delta_M - Y) \leq W, aZ<0\} \\ &\qquad + P_\theta(aZ=0) \end{aligned} \tag{2.4}$$

where

$$W = \log\left(\frac{aZ}{\exp(aZ)-1}\right).$$

Noting that W is less than zero if $aZ>0$, and is larger than zero if $aZ<0$, we have

$$\begin{aligned} P_\theta\{a(\delta_M - Y) \geq W, aZ>0\} &= E_\theta\{I_{(aZ>0)} P_\theta(a(\delta_M - Y) \geq W \mid Z)\} \\ &\geq \frac{1}{2} P_\theta(aZ>0) \end{aligned} \tag{2.5}$$

and

$$\begin{aligned} P_\theta\{a(\delta_M - Y) \leq W, aZ<0\} &= E_\theta\{I_{(aZ<0)} P_\theta(a(\delta_M - Y) \leq W \mid Z)\} \\ &\geq \frac{1}{2} P_\theta(aZ<0) \end{aligned} \tag{2.6}$$

where I_A denotes the indicator function of the set A. Substituting (2.5) and (2.6) into (2.4), we have

$$P_\theta\{L_2(\delta_M, Y) \le L_2(\delta, Y)\}$$
$$\ge \frac{1}{2}\{P_\theta(aZ>0)+P_\theta(aZ<0)\}+P_\theta(aZ=0)$$
$$\ge \frac{1}{2}.$$

Hence the proof is completed.

3. Examples

In this section we consider two examples to apply Theorem in Section 2.

Example 3.1. Let $(X_1, ..., X_n, Y)$ have a multivariate normal distribution such that $X_1, ..., X_n$ are *i.i.d.* according to $N(\mu, \sigma^2)$, the distribution of Y is also $N(\mu, \sigma^2)$ and the covariance between Y and X_i is $\rho\sigma^2\left(\rho^2<\frac{1}{n}\right)$. Based on $X=(X_1, ..., X_n)$, we want to predict Y.

Let $\overline{X}=\frac{1}{n}\sum_{i=1}^{n} X_i$. Then it is easy to see that the distribution of $\overline{X}-Y$ is normal with mean zero. Hence $\overline{X}$ is a MU predictor of Y. Let $T=(X_1-\overline{X}, ..., X_n-\overline{X})$ and θ denote the unknown parameters among μ, σ^2 and ρ. Since

$$cov(\overline{X}-Y, X_i-\overline{X})=E_\theta\{(\overline{X}-Y)(X_i-\overline{X})\}=0,$$

T is independent of $\overline{X}-Y$. So we can apply Theorem to the family of predictors such that

$$C=\{\delta;\ \delta(X)=\overline{X}+Z(T)\}$$

and conclude that $\overline{X}$ is the Pitman-closest in the family C with respect to the squared error and LINEX losses irrespective of which parameters of μ, σ^2, ρ are unknown.

Under the squared error loss, $\overline{X}$ has the minimum risk among the predictors in C. In fact, since $\overline{X}-Y$ is independent of Z, for any $\delta \in C$

$$E_\theta(\delta(X)-Y)^2=E_\theta(\overline{X}-Y+Z)^2$$
$$=E_\theta(\overline{X}-Y)^2+E_\theta Z^2$$
$$\ge E_\theta(\overline{X}-Y)^2.$$

However, $\overline{X}$ is inadmissible within C under the LINEX loss.

First suppose that $\theta=\mu$ is the only unknown parameter. Let

$$\delta_1(X)=\overline{X}+\frac{1}{2}\left(2\rho-\frac{n+1}{n}\right)a\sigma^2,$$

which belongs to the family C. It is easy to see that

$$E_\theta\{L_2(\delta_1(X)),\ Y\}=\frac{\alpha^2\sigma^2}{2}\left(1+\frac{1}{n}-2\rho\right)$$

and

$$E_\theta\{L_2(\overline{X},\ Y)\}=\exp\left\{\frac{\alpha^2\sigma^2}{2}\left(1+\frac{1}{n}-2\rho\right)\right\}-1. \tag{3.1}$$

Since $e^x-1>x(x\neq 0)$, the risk function of δ_1 is less than that of $\overline{X}$. See Xiao [7].

Next suppose that $\theta=(\mu,\ \sigma^2)$ is unknown but ρ is known. Let

$$\delta_2(X)=\overline{X}+\frac{1}{2}\left(2\rho-\frac{n+1}{n}\right)\alpha\hat{\sigma}^2,$$

where $\hat{\sigma}^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2$. Note that δ_2 is contained in the family C. A straightforward calculation shows that

$$E_\theta\{L_2(\delta_2(X),\ Y)\}=\left(1+\frac{2u}{\nu}\right)^{-\frac{\nu}{2}}e^u+u-1, \tag{3.2}$$

where $u=\frac{\alpha^2\sigma^2}{2}\left(1+\frac{1}{n}-2\rho\right)$ and $\nu=n-1$. Comparing (3.1) with (3.2), we can get

$$E_\theta\{L_2(\delta(X),\ Y)\}<E_\theta\{L_2(\overline{X},\ Y)\}$$

For details, see Zellner [8] (p. 448). Therefore the risk function of δ_2 is less than that of $\overline{X}$.

Example 3.2. Let $(\lambda_1,\ X_1),\ ...,\ (\lambda_n,\ X_n)$ be *i.i.d.* random pairs where λ_i is distributed according to $N(\mu,\ \tau^2)$ and the conditional distribution of X_i given λ_i is $N(\lambda_i,\ \sigma^2)$. Suppose $\rho=\sigma^2/\tau^2$ is known. Based on $X=(X_1,\ ...,\ X_n)$, we want to predict $Y=\lambda_n$.

If all parameters were known, we would use the Bayes estimator

$$\delta_B(X)=B\mu+(1-B)X_n \tag{3.3}$$

where $B=\frac{\sigma^2}{\sigma^2+\tau^2}=\frac{\rho}{1+\rho}$. Since μ is unknown but B is known, we consider the following empirical Bayes estimator instead of δ_B

$$\delta_E(X)=B\overline{X}+(1-B)X_n \tag{3.4}$$

where $\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$.

It is easy to see that the distribution of δ_E-Y is normal with mean zero. So δ_E is a MU predictor of Y. Noting that $E_\theta(Y|X)=\delta_B$, it follows from (3.3) and (3.4) that

$$\begin{aligned} cov(Y-\delta_E, X_i-\overline{X}) &= E_\theta\{(Y-\delta_E)(X_i-\overline{X})\} \\ &= E_\theta\{(B\mu+(1-B)X_n-\delta_E)(X_i-\overline{X})\} \\ &= BE_\theta\{(\mu-\overline{X})(X_i-\overline{X})\} \\ &= 0. \end{aligned}$$

Hence $T=(X_1-\overline{X}, \ldots, X_n-\overline{X})$ is independent of $Y-\delta_E$. So applying Theorem to the family of predictors

$$C=\{\delta;\ \delta(X)=\delta_E(X)+Z(T)\},$$

we conclude that δ_E is the Pitman-closest in the family C with respect to the squared error and LINEX losses. In particular, since $\overline{X}$ and X_n are included in C, δ_E turns to be better than $\overline{X}$ and X_n under PMC.

Under the squared error loss, δ_E has the minimum risk among the predictors in C. In fact, since δ_E-Y is independent of Z, for any $\delta\in C$

$$\begin{aligned} E_\theta(\delta-Y)^2 &= E_\theta(\delta_E-Y+Z)^2 \\ &= E_\theta(\delta_E-Y)^2+E_\theta Z^2 \\ &\geq E_\theta(\delta_E-Y)^2. \end{aligned}$$

However, δ_E is inadmissible within C under the LINEX loss.

First suppose that $\theta=\mu$ is the only unknown parameter. Let

$$\delta_1(X)=B\overline{X}+(1-B)X_n-\frac{a\sigma^2}{2}\left(1-\frac{n-1}{n}B\right),$$

which belongs to the family C. It is easy to see that

$$E_\theta\{L_2(\delta_1(X), Y)\}=\frac{a^2\sigma^2}{2}\left(1-\frac{n-1}{n}B\right)$$

and

$$E_\theta\{L_2(\delta_E(X), Y)\}=\exp\left\{\frac{a^2\sigma^2}{2}\left(1-\frac{n-1}{n}B\right)\right\}-1. \tag{3.5}$$

Since $e^x-1>x (x\neq 0)$, the risk function of δ_1 is less than that of δ_E.

Next suppose that $\theta=(\mu, \tau^2)$ is unknown. Let

$$\delta_2(X)=B\overline{X}+(1-B)X_n-\frac{aB}{2}\left(1-\frac{n-1}{n}B\right)\hat{\sigma}^2,$$

where $\tilde{\sigma}^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2$. Note that δ_2 is contained in the family C. A straightforward calculation shows that

$$E_\theta\{L_2(\delta_2(X), Y)\}=\left(1+\frac{2}{\nu}u\right)^{-\frac{\nu}{2}}e^u+u-1 \tag{3.6}$$

where $u=\frac{a^2\sigma^2}{2}\left(1-\frac{n-1}{n}B\right)$ and $\nu=n-1$. Comparing (3.5) with (3.6), we can get

$$E_\theta\{L_2(\delta_2(X), Y)\}<E_\theta\{L_2(\delta_E(X), Y)\}.$$

See (3.2). Therefore the risk function of δ_2 is less than that of δ_E.

References

[1] Datta, G. S., Pitman closeness for hierarchical Bayes predictors in mixed linear models, Pitman's measure of closeness, Comm. Statist.-Theory Methods, **20** (1991), 3713-3727.

[2] Ghosh, M. and P. K. Sen, Median unbiasedness and Pitman closeness, J. Amer. Statist. Assoc., **84** (1989), 1089-1091.

[3] Rao, C. R., J. P. Keating and R. L. Mason, The Pitman nearness criterion and its determination, Comm. Statist.-Theory Methods, **15** (1986), 3173-3191.

[4] Shafie, K. and Noorbaloochi, S., Asymmetric unbiased estimation in location families, Statistcs & Decisions, **13** (1995), 307-314.

[5] Takada, Y., Median unbiasedness in an invariant prediction problem, Statistics & Probability Letter, **12** (1991), 281-283.

[6] Varian, H. R., A Bayesian approach to real estate assessment, Studies in Bayesian Econometrics and Statistics in honor of Leonard J. Savage, (eds. S. E. Fienberg and A. Zellner), 195-208 (1975), North Holland, Amsterdam.

[7] Xiao, Y. S., LINEX unbiasedness in a prediction problem, to appear in Ann. Inst. statist. Math. (1999).

[8] Zellner, A., Bayesian estimation and prediction using asymmetric loss functions, J. Amer. Statist. Assoc., **81** (1986), 446-451.