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 $\theta$. 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 $\delta_1$ and $\delta_2$, $\delta_1$ is said to be better than $\delta_2$ under PMC with respect to $L$ if for all $\theta$
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 $\delta$ is better than any other $\delta' \in C$ under PMC with respect to $L$. A predictor $\delta$ is said to be MU of $Y$ if for all $\theta$

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

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 $\delta_m$ 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 $\delta_m$ is Pitman-closest in the family $C$ with respect to the squared error loss. The following theorem shows that $\delta_m$ 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 $\delta_m$ 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

$$P_{\theta}(L_1(\delta_m, Y) \leq L_1(\delta, Y)) = P_{\theta}(Z^2 + 2Z(\delta_m - Y) \geq 0)$$

$$\geq P_{\theta}(Z(\delta_m - Y) \geq 0)$$

$$= P_{\theta}(\delta_m - Y \geq 0, Z > 0)$$

$$+ P_{\theta}(\delta_m - Y \leq 0, Z < 0)$$

$$+ P_{\theta}(Z = 0).$$

(2.1)

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

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

(2.2)
Median unbiasedness and Pitman's measure of closeness in a prediction problem

and

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

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

\[ P_\theta(L_1(\delta_{\theta}, Y) \leq L_1(\delta, Y)) \geq \frac{1}{2} (P_\theta(Z > 0) + P_\theta(Z < 0)) + P_\theta(Z = 0) \]

\[ \geq \frac{1}{2}. \]

Hence the first part of the theorem is proved.

Next we consider the LINEX loss \( L_2 \). Then

\[
P_\theta\{L_2(\delta_{\theta}, Y) \leq L_2(\delta, Y)\}
= P_\theta(\exp[\alpha(\delta_{\theta} - Y)](\exp(\alpha Z) - 1) - aZ \geq 0)
= P_\theta(\alpha(\delta_{\theta} - Y) \geq W, aZ > 0)
+ P_\theta(\alpha(\delta_{\theta} - Y) \leq W, aZ < 0)
+ P_\theta(\alpha Z = 0) \tag{2.4}
\]

where

\[ W = \log \left( \frac{aZ}{\exp(\alpha Z) - 1} \right). \]

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

\[ P_\theta(\alpha(\delta_{\theta} - Y) \geq W, aZ > 0) = E_\theta(I_{aZ > 0} P_\theta(\alpha(\delta_{\theta} - Y) \geq W | Z)) \]

\[ \geq \frac{1}{2} P_\theta(aZ > 0) \tag{2.5} \]

and

\[ P_\theta(\alpha(\delta_{\theta} - Y) \leq W, aZ < 0) = E_\theta(I_{aZ < 0} P_\theta(\alpha(\delta_{\theta} - Y) \leq W | Z)) \]

\[ \geq \frac{1}{2} P_\theta(aZ < 0) \tag{2.6} \]

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

\[ P(\ell_\delta, Y) \leq \ell_\delta(\delta, Y) \]
\[ \geq \frac{1}{2} (P(\ell(aZ > 0) + P(\ell(aZ < 0)) + P(\ell(aZ = 0)) \]
\[ \geq \frac{1}{2}. \]

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 \( \rho \sigma^2 \left\{ \frac{1}{\sqrt{n}} \right\} \). Based on \( X = (X_1, ..., X_n) \), we want to predict \( Y \).

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

\[ \text{cov}(\bar{X} - Y, X_i - \bar{X}) = E(\bar{X} - Y)(X_i - \bar{X}) = 0, \]

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

\[ C = \{ \delta; \delta(X) = \bar{X} + Z(T) \} \]

and conclude that \( \bar{X} \) is the Pitman-closest in the family \( C \) with respect to the squared error and LINEX losses irrespective of which parameters of \( \mu, \sigma^2, \rho \) are unknown.

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

\[ E(\delta(X) - Y)^2 = E(\bar{X} - Y + Z)^2 \]
\[ = E(\bar{X} - Y)^2 + E(Z^2) \]
\[ \geq E(\bar{X} - Y)^2. \]

However, \( \bar{X} \) is inadmissible within \( C \) under the LINEX loss.

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

\[ \delta(X) = \bar{X} + \frac{1}{2} \left( 2 \rho - \frac{n+1}{n} \right) \sigma^2, \]

which belongs to the family \( C \). It is easy to see that
Median unbiasedness and Pitman's measure of closeness in a prediction problem

\[ E_s(L_\delta(\delta(X)), Y) = \frac{\sigma^2}{2} \left( 1 + \frac{1}{n} - 2\rho \right) \]

and

\[ E_s(L_\delta(\overline{X}, Y)) = \exp \left( \frac{\sigma^2}{2} \left( 1 + \frac{1}{n} - 2\rho \right) \right) - 1. \] \hspace{1cm} (3.1)

Since \( e^x - 1 > x(x > 0) \), the risk function of \( \delta_1 \) is less than that of \( \overline{X} \). See Xiao [7].

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

\[ \delta_2(X) = \overline{X} + \frac{1}{2} \left( 2\rho - \frac{n+1}{n} \right) \alpha \delta^2, \]

where \( \delta^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2 \). Note that \( \delta_2 \) is contained in the family \( C \). A straightforward calculation shows that

\[ E_s(L_\delta(\delta_2(X), Y)) = \left( 1 + \frac{2\mu}{\nu} \right)^{-\frac{1}{2}} e^u + u - 1, \]

(3.2)

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

\[ E_s(L_\delta(\delta(X), Y)) < E_s(L_\delta(\overline{X}, Y)) \]

For details, see Zellner [8] (p. 448). Therefore the risk function of \( \delta_2 \) is less than that of \( \overline{X} \).

Example 3.2. Let \( (\lambda_i, X_i), ... , (\lambda_n, X_n) \) be i.i.d. random pairs where \( \lambda_i \) is distributed according to \( N(\mu, \tau^2) \) and the conditional distribution of \( X_i \) given \( \lambda_i \) 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 \( Y = \lambda_1 \).

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

\[ \delta_b(X) = B\mu + (1 - B)X_n \]

(3.3)

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

\[ \delta_e(X) = B\overline{X} + (1 - B)X_n \]

(3.4)

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

It is easy to see that the distribution of \( \delta_e - Y \) is normal with mean zero. So \( \delta_e \) is a MU predictor of \( Y \). Noting that \( E_s(Y | X) = \delta_b \), it follows from (3.3) and (3.4) that
\[ \text{Yusun Xiao and Yoshikazu Takada} \]
\[
cov(Y - \delta_\varepsilon, X_i - \overline{X}) = E_\varepsilon((Y - \delta_\varepsilon)(X_i - \overline{X}))
\]
\[
= E_\varepsilon((B\mu + (1-B)X_n - \delta_\varepsilon)(X_i - \overline{X}))
\]
\[
= B\varepsilon_\varepsilon((\mu - \overline{X})(X_i - \overline{X}))
\]
\[
= 0.
\]

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

\[ C = \{ \delta; \delta(X) = \delta_\varepsilon(X) + Z(T) \}, \]

we conclude that \( \delta_\varepsilon \) 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 \( C \), \( \delta_\varepsilon \) turns to be better than \( \overline{X} \) and \( X_n \) under PMC.

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

\[ E_\delta(\delta - Y)^2 = E_\delta(\delta_\varepsilon - Y + Z)^2 \]
\[ = E_\delta(\delta_\varepsilon - Y)^2 + E_\delta Z^2 \]
\[ \geq E_\delta(\delta_\varepsilon - Y)^2. \]

However, \( \delta_\varepsilon \) is inadmissible within \( C \) under the LINEX loss.

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

\[ \delta_\varepsilon(X) = B\overline{X} + (1-B)X_n - \frac{a\sigma^2}{2} \left(1 - \frac{n-1}{n} B \right), \]

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

\[ E_\delta(L_\varepsilon(\delta_\varepsilon(X), Y)) = \frac{a^2 \sigma^2}{2} \left(1 - \frac{n-1}{n} B \right) \]

and

\[ E_\delta(L_\varepsilon(\delta_\varepsilon(X), Y)) = \exp \left( \frac{a^2 \sigma^2}{2} \left(1 - \frac{n-1}{n} B \right) \right) - 1. \tag{3.5} \]

Since \( e^{\varepsilon - 1} > x(x \neq 0) \), the risk function of \( \delta_\varepsilon \) is less than that of \( \delta_\varepsilon \).

Next suppose that \( \theta = (\mu, \sigma) \) is unknown. Let

\[ \delta(X) = B\overline{X} + (1-B)X_n - \frac{aB}{2} \left(1 - \frac{n-1}{n} B \right) \sigma^2, \]

\[ \text{Yusun Xiao and Yoshikazu Takada} \]
\[
cov(Y - \delta_\varepsilon, X_i - \overline{X}) = E_\varepsilon((Y - \delta_\varepsilon)(X_i - \overline{X}))
\]
\[
= E_\varepsilon((B\mu + (1-B)X_n - \delta_\varepsilon)(X_i - \overline{X}))
\]
\[
= B\varepsilon_\varepsilon((\mu - \overline{X})(X_i - \overline{X}))
\]
\[
= 0.
\]

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

\[ C = \{ \delta; \delta(X) = \delta_\varepsilon(X) + Z(T) \}, \]

we conclude that \( \delta_\varepsilon \) 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 \( C \), \( \delta_\varepsilon \) turns to be better than \( \overline{X} \) and \( X_n \) under PMC.

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

\[ E_\delta(\delta - Y)^2 = E_\delta(\delta_\varepsilon - Y + Z)^2 \]
\[ = E_\delta(\delta_\varepsilon - Y)^2 + E_\delta Z^2 \]
\[ \geq E_\delta(\delta_\varepsilon - Y)^2. \]

However, \( \delta_\varepsilon \) is inadmissible within \( C \) under the LINEX loss.

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

\[ \delta_\varepsilon(X) = B\overline{X} + (1-B)X_n - \frac{a\sigma^2}{2} \left(1 - \frac{n-1}{n} B \right), \]

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

\[ E_\delta(L_\varepsilon(\delta_\varepsilon(X), Y)) = \frac{a^2 \sigma^2}{2} \left(1 - \frac{n-1}{n} B \right) \]

and

\[ E_\delta(L_\varepsilon(\delta_\varepsilon(X), Y)) = \exp \left( \frac{a^2 \sigma^2}{2} \left(1 - \frac{n-1}{n} B \right) \right) - 1. \tag{3.5} \]

Since \( e^{\varepsilon - 1} > x(x \neq 0) \), the risk function of \( \delta_\varepsilon \) is less than that of \( \delta_\varepsilon \).

Next suppose that \( \theta = (\mu, \sigma) \) is unknown. Let

\[ \delta(X) = B\overline{X} + (1-B)X_n - \frac{aB}{2} \left(1 - \frac{n-1}{n} B \right) \sigma^2, \]
where $\sigma^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$. Note that $\delta_2$ is contained in the family $C$. A straightforward calculation shows that

$$
E_\delta (L_2(\delta_2(X), Y)) = \left(1 + \frac{2}{\nu} \mu\right)^{-\frac{1}{2}} e^{\mu} + \mu - 1
$$

(3.6)

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

$$
E_\delta (L_2(\delta_2(X), Y)) < E_\delta (L_2(\delta_\delta(X), Y)).
$$

See (3.2). Therefore the risk function of $\delta_2$ is less than that of $\delta_\delta$.

References


