The Annals of Statistics
1981, Vol. 9, No. 4, 917–921

RELATION OF THE BEST INVARIANT PREDICTOR AND THE BEST UNBIASED PREDICTOR IN LOCATION AND SCALE FAMILIES

BY YOSHIKAZU TAKADA

University of Tsukuba

In this paper a necessary and sufficient condition for a predictor to be the best invariant predictor in location and scale families is given. Using this condition, it is shown that the best invariant predictor is expressed by a linear combination of the best unbiased predictor and the best unbiased estimator of the scale parameter.

1. Introduction. Statistical prediction is the use of the data from an experiment to make some statement about the outcome of a future experiment. Let $X' = (X_1, \cdots, X_r)$ be the observed random variables and Y a future random variable. We assume that X and Y have a joint distribution given by the density

$$\sigma^{-(r+1)} f\{(x_1 - \mu)/\sigma, \cdots, (x_r - \mu)/\sigma, (y - \mu)/\sigma\}$$

for some known function f, where (μ, σ), with $\sigma > 0$, is an unknown location-scale parameter. Let $\Theta = \{(\mu, \sigma): \sigma > 0\}$ be the parameter space and $\theta = (\mu, \sigma)$. Let Φ be a class of specified statistics such that if $\delta_1 \in \Phi$ and $\delta_2 \in \Phi$, then $a\delta_1 + b\delta_2 \in \Phi$ for any real valued constants a and b. For example the class Φ may consist of all linear combinations of observed random variables. In the sequel we assume that all estimators and predictors belong to Φ.

We say that a statistic $\delta(X)$ is an unbiased predictor of Y if

$$E_\theta \delta(X) = E_\theta(Y) \quad \text{for all} \quad \theta \in \Theta. \tag{1.1}$$

We say that a statistic $\delta(X)$ is an invariant predictor of Y if for any (a, b) with $a > 0$,

$$\delta(aX + be) = a\delta(X) + b \quad \text{for almost all } X \tag{1.2}$$

(the exceptional set may depend on (a, b)), where $e = (1, \cdots, 1)'$.

Let

$$R_u = \{\delta \in \Phi: \delta \quad \text{is an unbiased predictor of } Y\}$$

and

$$R_I = \{\delta \in \Phi: \delta \quad \text{is an invariant predictor of } Y\}.$$

In this paper we adopt squared error as loss function. Then $\delta^* \in R_u$ is said to be the best unbiased predictor if δ^* minimizes $E_\theta\{\delta(X) - Y\}^2$ for all $\theta \in \Theta$ among all $\delta \in R_u$. In the same way we say that $\delta_I^* \in R_I$ is the best invariant predictor if δ_I^* minimizes $E_\theta\{\delta(X) - Y\}^2$ for all $\theta \in \Theta$ among all $\delta \in R_I$.

For the estimation problem, Pitman (1939) has given explicit formulae for the best invariant estimators of location and scale parameters. Mann (1969) has obtained the relation of the best invariant estimator and the best unbiased estimator. Using this result, Kaminsky, et al. (1975) have obtained the best linear invariant predictor of order statistic.

The purpose of this paper is to obtain the relation between the best invariant predictor and the best unbiased predictor. The method is different to that of Mann and uses a characterization of the best invariant predictor.

Received February, 1980; revised July, 1980.

AMS 1970 subject classifications. Primary 62F99; secondary 62C99.

Key words and phrases. Invariance, unbiasedness, prediction, location and scale family.

2. Characterization of the best invariant predictor. Let

$$S_0 = \{\gamma \in \Phi: E_\theta\gamma(X) = 0, E_\theta\gamma^2(X) < \infty \quad \text{for all} \quad \theta \in \Theta\}.$$

Then the following lemma is well known (see Zacks (1971), Theorem 3.3.1.).

LEMMA 1. *An unbiased estimator $\delta^* \in \Phi$ is a uniformly minimum variance* (U.M.V.) *unbiased estimator of $g(\theta)$ among the class Φ if and only if for any $\theta \in \Theta$ and $\gamma \in S_0$,*

$$E_\theta\{\delta^*(X)\gamma(X)\} = 0. \tag{2.1}$$

For the prediction problem, Ishii (1978) obtained the following lemma, the proof of which is almost the same as that of Lemma 1. We give the proof for later use.

LEMMA 2. *An unbiased predictor $\delta^* \in R_u$ is a best unbiased predictor if and only if for any $\theta \in \Theta$ and any $\gamma \in S_0$*

$$E_\theta[\{Y - \delta^*(X)\}\gamma(X)] = 0. \tag{2.2}$$

PROOF. Assume that δ^* is a best unbiased predictor. For any $\gamma \in S_0$, let $\delta(X) = \delta^*(X) + \lambda\gamma(X)$. Then $\delta \in R_u$ and

$$E_\theta\{Y - \delta(X)\}^2 = E_\theta\{Y - \delta^*(X)\}^2 - 2\lambda E_\theta[\{Y - \delta^*(X)\}\gamma(X)] + \lambda^2 E_\theta\{\gamma^2(X)\},$$

which implies $E_\theta[\{Y - \delta^*(X)\}\gamma(X)] = 0$.

Conversely, if δ^* satisfies (2.2), for any $\delta \in R_u$ let $\gamma(X) = -\delta^*(X) + \delta(X)$. Then $\gamma \in S_0$ and therefore

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

Hence δ^* is a best unbiased predictor. This completes the proof of the lemma.

THEOREM 1. *If δ^* is a best unbiased predictor, then it is an invariant predictor.*

PROOF. Let $R(\theta, \delta) = E_\theta[\{Y - \delta(X)\}^2/\sigma^2]$ and $\theta_{a,b} = (a\mu + b, a\sigma)$ for $\theta = (\mu, \sigma)$. For any (a, b) with $a > 0$, let $\delta_{a,b}(X) = \{\delta^*(aX + be) - b\}/a$. Then it is easy to see that $\delta_{a,b} \in R_u$ and

$$R(\theta, \delta_{a,b}) = R(\theta_{a,b}, \delta^*). \tag{2.3}$$

Since δ^* is a best unbiased predictor, $R(\theta, \delta^*) \leq R(\theta, \delta_{a,b})$. Therefore for any (a, b) with $a > 0$, $R(\theta, \delta^*) \leq R(\theta_{a,b}, \delta^*)$. Hence for any $\theta_1 \neq \theta_2$, $R(\theta_1, \delta^*) = R(\theta_2, \delta^*)$. Then from (2.3) we have $R(\theta, \delta_{a,b}) = R(\theta, \delta^*)$ for any $\theta \in \Theta$. This shows that $\delta_{a,b}$ is a best unbiased predictor. From Lemma 2,

$$\begin{aligned} E_\theta\{\delta_{a,b}(X) - \delta^*(X)\}^2 &= E_\theta([\{\delta_{a,b}(X) - Y\} - \{\delta^*(X) - Y\}]\{\delta_{a,b}(X) - \delta^*(X)\}) \\ &= 0, \end{aligned}$$

which implies the theorem.

REMARK 1. From Theorem 1 it turns out that the best invariant predictor is better than the best unbiased predictor.

We say that $\gamma \in \Phi$ is scale invariant if for any (a, b) with $a > 0$,

$$\gamma(aX + be) = a\gamma(X) \qquad \text{for almost all } X$$

(the exceptional set may depend on (a, b)). Let $S_I = \{\gamma \in \Phi: \gamma$ is scale invariant and $E_{\theta_0}\gamma^2(X) < \infty\}$, where $\theta_0 = (0, 1)$. Then we have the following theorem, the proof of which is similar to that of Lemma 2, noticing that the mean square error of $\delta \in R_I$ is proportional to σ^2.

THEOREM 2. *An invariant predictor $\delta^* \in R_I$ is the best invariant predictor if and only if for any $\gamma \in S_I$*

$$E_{\theta_0}[\{Y - \delta^*(X)\}\gamma(X)] = 0. \tag{2.4}$$

3. Relation of the best invariant predictor and the best unbiased predictor. Now we consider the relation of the best invariant predictor and the best unbiased predictor. For this we need the following lemma.

LEMMA 3. *If $\hat{\sigma}^* \in \Phi$ is a* U.M.V. *unbiased estimator of σ among the class Φ, then $\hat{\sigma}^* \in S_I$.*

PROOF. For any (a, b) with $a > 0$, let $\hat{\sigma}_{a,b}(X) = \hat{\sigma}^*(aX + be)/a$ and $R(\theta, \hat{\sigma}) = E[\{\hat{\sigma}(X) - \sigma\}^2/\sigma^2]$. Then the proof follows by an argument similar to that of Theorem 1, using Lemma 1.

THEOREM 3. *Let δ^* be the best unbiased predictor of Y and $\hat{\sigma}^*$ the* U.M.V. *unbiased estimator of σ. Let*

$$c_1 = E_{\theta_0}[\{Y - \delta^*(X)\}\hat{\sigma}^*(X)] \qquad \text{and} \qquad c_2 = E_{\theta_0}\{\hat{\sigma}^{*2}(X)\}.$$

Put

$$\delta_I^*(X) = \delta^*(X) + (c_1/c_2)\hat{\sigma}^*(X). \tag{2.5}$$

Then δ_I^* is the best invariant predictor of Y.

PROOF. From Theorem 1 and Lemma 3, $\delta_I^* \in R_I$. Now we show that δ_I^* satisfies the condition of Theorem 2. If $\gamma \in S_I$, then $E_\theta\{\gamma(X)\} = E_{\theta_0}\{\gamma(\sigma X + \mu e)\} = \sigma E_{\theta_0}\{\gamma(X)\}$. Therefore γ must be of the form

$$\gamma(X) = k\hat{\sigma}(X) \tag{2.6}$$

for some k, where $\hat{\sigma}$ is an unbiased estimator of σ. It follows from Lemmas 1 and 2 that

$$E_{\theta_0}[\{Y - \delta^*(X)\}\{\hat{\sigma}(X) - \hat{\sigma}^*(X)\}] = 0,$$

and

$$E_{\theta_0}[\hat{\sigma}^*(X)\{\hat{\sigma}(X) - \hat{\sigma}^*(X)\}] = 0.$$

Therefore from (2.6) and the definition of c_1 and c_2, we have that

$$\begin{aligned} E_{\theta_0}[\{Y - \delta_I^*(X)\}\gamma(X)] &= E_{\theta_0}(\{Y - \delta^*(X) - (c_1/c_2)\hat{\sigma}^*(X)\} \\ &\qquad \times [k\{\hat{\sigma}(X) - \hat{\sigma}^*(X)\} + k\hat{\sigma}^*(X)]) \\ &= kE_{\theta_0}[\{Y - \delta^*(X) - (c_1/c_2)\hat{\sigma}^*(X)\}\hat{\sigma}^*(X)] \\ &= 0, \end{aligned}$$

which proves the theorem.

REMARK 2. If X and Y are independent, let $g(\theta) = E_\theta(Y)$. Then the best unbiased predictor and the best invariant predictor of Y become the U.M.V. unbiased estimator and the best invariant estimator of $g(\theta)$, respectively and Theorem 3 coincides with Theorem 1 of Mann (1969).

EXAMPLE 1. Denote by $X_1 < X_2 \cdots < X_n$ the order statistics of a random sample of size n from a population with continuous pdf, $(1/\sigma)f\{(x - \mu)/\sigma\}$, assumed known up to location and scale. We consider the problem of predicting X_m after observing only $X_1, \cdots, X_r$, where $1 \le r < m \le n$ and consider only linear predictors, that is,

$$\Phi = \{\delta: \delta(X) = \textstyle\sum_{i=1}^r a_i X_i \quad \text{for some} \quad (a_1, \cdots, a_r)\}.$$

Since $E_\theta(X_i) = \mu + \sigma\alpha_i$, $\text{Cov}_\theta(X_i, X_j) = \sigma^2\beta_{ij}$, $i, j = 1, \cdots, n$, where the α_i and β_{ij} are known to be constant, that is, free of μ and σ (see David (1970), page 102), the Gauss-Markov least-squares theorem may be applied to give linear unbiased estimators of (μ, σ). We write δ^* and $(\hat\mu^*, \hat\sigma^*)$ for the best linear unbiased predictor of X_m and the U.M.V. linear unbiased estimator of (μ, σ). From the result of Kaminsky and Nelson (1975),

$$\delta^*(X) = \hat\mu^*(X) + \hat\sigma^*(X)\alpha_m + w'V^{-1}\{X - \hat\mu^*(X)e - \hat\sigma^*(X)\alpha\},$$

where $\alpha' = (\alpha_1, \cdots, \alpha_r)$, $w' = (\beta_{1m}, \cdots, \beta_{rm})$ and $V = (\beta_{ij})(1 \le i, j \le r)$. Then it is easy to see that

$$\begin{aligned} c_1 &= E_{\theta_0}[\{X_m - \delta^*(X)\}\hat\sigma^*(X)] \\ &= -\text{Cov}_{\theta_0}\{(1 - w'V^{-1}e)\hat\mu^*(X) + (\alpha_m - w'V^{-1}\alpha)\hat\sigma^*(X), \hat\sigma^*(X)\}. \end{aligned}$$

Then from Theorem 3 we have the same result of Kaminsky, et al. (1975), page 525.

EXAMPLE 2. In example 1, assume that the parent population is exponential, $(1/\sigma)\exp\{-(x - \mu)/\sigma\}$, $x > \mu$, $\sigma > 0$, and Φ is the class of all statistics. It is well known that for $1 \le i \le n$ the set of random variables

$$Z_i = (n - i + 1)(X_i - X_{i-1}) \qquad i = 1, \cdots, n \tag{2.9}$$

(where $X_0 = \mu$) are mutually independent with pdf, $\frac{1}{\sigma} e^{-x/\sigma}$, $x > 0$. Using (2.9), we have that

$$\begin{aligned} E(X_m \mid X) &= E_\theta\{X_r + \textstyle\sum_{i=r+1}^m (X_i - X_{i-1}) \mid X\} \\ &= X_r + \sigma \textstyle\sum_{i=r+1}^m 1/(n - i + 1). \end{aligned} \tag{2.10}$$

Let $T(X) = \sum_{i=2}^r (n - i + 1)(X_i - X_{i-1})$. Then from Theorem 3 of Epstein and Sobel (1954), (X_1, T) is sufficient and complete for θ. Therefore from (2.9) and (2.10) it is easy to see that the U.M.V. unbiased estimator of σ and the best unbiased predictor of X_m are respectively

$$\hat\sigma^*(X) = T(X)/(r - 1), \ \delta^*(X) = X_r + \hat\sigma^*(X) \textstyle\sum_{i=r+1}^m 1/(n - i + 1).$$

From (2.9) we have

$$\begin{aligned} E_\theta[\{X_m - \delta^*(X)\}\hat\sigma^*(X)] &= E_\theta[\{\textstyle\sum_{i=r+1}^m (X_i - X_{i-1}) \\ &\qquad - \hat\sigma^*(X) \textstyle\sum_{i=r+1}^m 1/(n - i + 1)\}\hat\sigma^*(X)] \\ &= -\textstyle\sum_{i=r+1}^m 1/(n - i + 1) V_\theta\{\hat\sigma^*(X)\}. \end{aligned}$$

Therefore we have that $c_1 = -\{\sum_{i=r+1}^m 1/(n - i + 1)\}/(r - 1)$ and $c_2 = r/(r - 1)$. Hence it follows from Theorem 3 that the best invariant predictor is

$$\delta_I^*(X) = X_r + (1 - 1/r)\hat\sigma^*(X) \textstyle\sum_{i=r+1}^m 1/(n - i + 1).$$

REMARK 3. Hora and Buehler (1967) give the best invariant predictor in more general invariant models, but do not deal with unbiased predictors or with restricted classes Φ of predictors, such as linear predictors.

ACKNOWLEDGMENT. The author wishes to thank the referee for his valuable comments.

REFERENCES

DAVID, H. A. (1970). *Order statistics.* Wiley, New York.

EPSTEIN, B. and SOBEL, M. (1954). Some theorems relevent to life testing from an exponential distribution. *Ann. Math. Statist.* **25** 373–381.

HORA, R. B., and BUEHLER, R. J. (1967). Fiducial theory and invariant prediction. *Ann. Math. Statist.* **38** 795–801.

ISHII, G. (1978). *Tokeiteki yosoku (Statistical prediction).* (In Japanese). In Basic sugaku No. 7. Gendai-Sugakusha, Tokyo.

KAMINSKY, K. S., MANN, N. R. and NELSON, P. I. (1975). Best and simplified linear invariant prediction of order statistics in location and scale families. *Biometrika* **62** 525–527.

KAMINSKY, K. S. and NELSON, P. I. (1975). Best linear unbiased prediction of order statistics in location and scale families. *J. Amer. Statist. Assoc.* **70** 145–150.

MANN, N. R. (1969). Optimum estimators for linear functions of location and scale parameters. *Ann. Math. Statist.* **40** 2149–2155.

PITMAN, E. J. G. (1939). The estimation of the location and scale parameters of a continuous population of any given form. *Biometrika* **30** 391–421.

ZACKS, S. (1971). *The theory of statistical inference.* Wiley, New York.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF TSUKUBA
SAKURA-MURA, NIIHARI-GUN
IBARAKI 300-31, JAPAN