Kumamoto J. Sci. (Math.)
Vol. 17, 75～81 March (1987)

BAYES INVARIANT PREDICTOR

Yoshikazu TAKADA

(Received November 17, 1986)

1. INTRODUCTION

The prediction problem which is invariant under a certain group of transformations was treated by many authors (e.g. Hora and Buehler [2] and Takada [4]). Under the condition that the parameter space is isomorphic to the group, it was shown that the best invariant predictor, which has the minimum risk among the class of invariant predictors, can be represented by using the Haar measure on the group. The purpose of this paper is to treat the case in which the condition is not fulfilled.

In a statistical decision problem, such a case was studied by Zidek [5], among other things. For simplicity we consider the location and scale family with shape parameter as the probability model, though the general treatment is possible.

Let $X=(X_1,\dots,X_n)$ be observable random variables and Y a future random variable we want to predict on the basis of X. We suppose that the joint probability density function of X and Y with respect to Lebesgue measure is given by

$$p(x,y\mid\theta,\rho)=\sigma^{-(n+1)}f\{(x_1-\mu)/\sigma,\dots,(x_n-\mu)/\sigma,(y-\mu)/\sigma\mid\rho\}, \tag{1.1}$$

where the location-scale parameter $\theta=(\mu,\sigma)$ and shape parameter ρ are unknown but have specified ranges $-\infty<\mu<\infty$, $\sigma>0$, $\rho\varepsilon\psi$ and $f(x,y\mid\rho)$ is a known function for each $\rho\varepsilon\psi$. Hence if ρ is known, the family of probability distributions becomes the location and scale family.

Consider the following transformation on the sample space;

$$(x_1,\dots,x_n,y)\rightarrow(ax_1+b,\dots,ax_n+b,ay+b) \tag{1.2}$$

with $a>0$. Then the family of probability distributions (1.1) is invariant under such a transformation group. But the parameter space is not isomorphic to the group unless ρ is known.

After observing X, we want to predict the value of Y. As the loss of erroneously predicting $Y=y$ by the value d, we adopt squared error $(y-d)^2$. We say that a predictor δ is invariant under (1.2) if for any (a,b) with $a>0$

$$\delta(ax_1+b,\ldots,ax_n+b)=a\delta(x_1,\ldots,x_n)+b.$$

It is easy to ee that for any invariant predictor δ, $\sigma^{-2}E(Y-\delta(X))^2$ depends only on ρ. Hence we denote it by $R(\rho,\delta)$. Since ρ is unknown, generally there does not exists the best invariant predictor. Therefore we seek the Bayes solution among the class of invariant predictors.

Let $G(\rho)$ be a probability distribution on ψ. Then we seek the invariant predictor which minimizes

$$\int R(\rho,\delta)G(d\rho) \tag{1.3}$$

among the class of invariant predictors. We say that such a predictor is the Bayes invariant predictor with respect to G.

In Section 2 we give a representation of the Bayes invariant predictor by using the Haar measure on the group. It is also shown that the Bayes invariant predictor is admissible among the class of invariant predictors. In Section 3 we consider the application to the first order autoregressive process.

2. BAYES INVARIANT PREDICTOR

To obtain the representation of the Bayes invariant predictor, we need the following lemma, the proof of which is obtained from Theorem 2 of Takada [4].

Lemma 1. *Suppose that the joint probability density function of X and Y is written as*

$$q(x,y\mid\theta)=\sigma^{-(n+1)}g\{(x_1-\mu)/\sigma,\ldots,(x_n-\mu)/\sigma,(y-\mu)/\sigma\} \tag{2.1}$$

for some known function g, where $\theta=(\mu,\sigma)$ with $\sigma>0$ is unknown. Then the predictor which minimizes $\sigma^{-2}E(Y-\delta(X))^2$ among the class of invariant predictors is given by

$$\delta^*(x) = \int\int\int y\sigma^{-2}q(x,y\,|\theta)\nu(d\theta)dy \,/ \int\int\int \sigma^{-2}q(x,y\,|\theta)\nu(d\theta)dy$$

where $\nu(d\theta) = \sigma^{-1}d\mu d\sigma$.

Now we give a representation of the Bayes invariant predictor with respect to G by using the Haar measure ν.

Theorem 1. *The Bayes invariant predictor with respect to G is given by*

$$\delta_G(x) = \int\int\int\int y\sigma^{-2}p(x,y\,|\theta,\rho)\nu(d\theta)G(d\rho)dy \,/ \int\int\int\int \sigma^{-2}p(x,y\,|\theta,\rho)\nu(d\theta)G(d\rho)dy$$

where p is (1.1) *and* $\nu(d\theta) = \sigma^{-1}d\mu d\sigma$.

Proof. Let

(2.2) $$f_G(x,y) = \int f(x,y\,|\rho)G(d\rho).$$

Then it follows from (1.3) that

(2.3) $$\int R(\rho,\delta)G(d\rho) = \sigma^{-2}E(Y-\delta(X))^2$$

where the expectation is taken to (2.1) with $g=f_G$. Hence we have the result from Lemma 1.

From the Bayesian point of view, it turns out that the Bayes invariant predictor with respect to G is the conditional mean of Y given X when the prior distribution on (θ,ρ) is $\sigma^{-2}\nu(d\theta)G(d\rho)$ (cf. Theorem 3.1 of Zidek [5]). If ρ is known, the best invariant predictor δ_ρ is easily obtained from Lemma 1. The next result shows that the Bayes invariant predictor can be expressed as the expectation of δ_ρ with respect to the posterior distribution of ρ given X.

Theorem 2. *The Bayes invariant predictor* δ_G *is written as*

(2.4) $$\delta_G(x) = \int \delta_\rho(x)G(d\rho\,|x)$$

where

(2.5) $$\delta_\rho(x) = \int\int\int yp(x,y\,|\theta,\rho)\sigma^{-2}\nu(d\theta)dy \,/ \int\int\int p(x,y\,|\theta,\rho)\sigma^{-2}\nu(d\theta)dy$$

and

(2.6) $G(d\rho \mid x)=G(d\rho)\int\int\int p(x,y\mid\theta,\rho)\sigma^{-2}\nu(d\theta)dy/\int\int\int\int p(x,y\mid\theta,\rho)\sigma^{-2}\nu(d\theta)G(d\rho)dy.$

Proof. From Theorem 1 δ_G can be written as

$$\delta_G(x)=\int\{\int yp(y\mid x,\rho)dy\}G(d\rho\mid x)$$

where

$$p(y\mid x,\rho)=\int\int p(x,y\mid\theta,\rho)\sigma^{-2}\nu(d\theta)/\int\int\int p(x,y\mid\theta,\rho)\sigma^{-2}\nu(d\theta)dy$$

and $G(d\rho\mid x)$ is (2.6). Lemma 1 shows that $\delta_\rho(x)=\int yp(y\mid x,\rho)dy$ is the best invariant predictor for the case of known ρ, which completes the proof.

If $G(d\rho\mid x)$ is concentrated about some modal value, by (2.4) we can write approximately

$$\delta_G(x)=\delta_{\hat\rho}(x)$$

where $\hat\rho$ is the mode of $G(d\rho\mid x)$.

Next we shown the admissibility of the Bayes invariant predictor among the class of invariant predictors.

Theorem 3. *Suppose that for each* $\rho\varepsilon\psi$, $\{x;\int f(x,y\mid\rho)dy>0\}$ does not de-*pend on* ρ. *Then* δ_G *is admissible among the class of invariant predictors.*

Proof. Without loss of generality, we can assume $\theta=(0,1)$. If there exists an invariant predictor δ such that for each $\rho\varepsilon\psi$

$$E(Y-\delta(X))^2\leq E(Y-\delta_G(X))^2$$

then

$$\int R(\rho,\delta)G(d\rho)\leq\int R(\rho,\delta_G)G(d\rho),$$

which implies that δ is also the Bayes invariant predictor. It follows from (2.3) that the Bayes invariant predictor is the best invariant predictor for the distribution (2.2). Hence from Theorem 2 of Takada [3] we have that

$$\int \{E(\delta(X)-\delta_G(X))^2\}G(d\rho)=E\{[(Y-\delta_G(X))-(Y-\delta(X))](\delta(X)-\delta_G(X))\}$$
$$=0$$

where the expectation of the left hand side is taken to (2.2). Then there exists a $\rho\varepsilon\psi$ such that

$$E(\delta(X)-\delta_G(X))^2=0.$$

By the assumption of the theorem, we have that for each $\rho\varepsilon\psi$

$$\delta(X)=\delta_G(X) \quad \text{a.e.,}$$

which implies the admissibility of δ_G.

3. FIRST ORDER AUTOREGRESSIVE PROCESS

Let X_i's be random variables generated by the first order autoregressive process;

$$X_i-\mu=\rho(X_{i-1}-\mu)+\varepsilon_i, \quad i=\ldots,-1,0,1,\ldots,$$

where ε_i's are independent and identically distributed normal random variables with mean 0 and variance $(1-\rho^2)\sigma^2$ ($|\rho|<1$). Here we assume that (μ,σ,ρ) is unknown. We observe $X=(X_1,\ldots,X_n)$ and wish to use it to predict $Y=X_{n+1}$.

It is easy to see that the joint probability density function of X and Y is given by

$$p(x,y\,|\theta,\rho)=(2\pi)^{-(n+1)/2}(1-\rho^2)^{-n/2}\sigma^{-(n+1)}exp\{-[(x_1-\mu)^2$$
$$+(1-\rho^2)^{-1}\sum_{i=1}^{n}(x_{i+1}-\mu-\rho(x_i-\mu))^2]/2\sigma^2\}$$

where $\theta=(\mu,\sigma)$.

The straightforward calculation shows that the predictor δ_ρ given by (2.5) becomes

$$(3.1) \qquad \delta_\rho(x)=\rho x_n+\left(\frac{1-\rho}{n-(n-2)\rho}\right)[(1+\rho)x_1+(1-\rho)\sum_{i=1}^{n-1}(x_{i+1}-\rho x_i)]$$

and for a distribution G(ρ) of ρ, (2.6) is equal to

$$(3.2)\quad G(d\rho\,|x)=\frac{(1-\rho^2)^{-(n-1)/2}[(1+\rho)/(n-(n-2)\rho)]^{1/2}Q(\rho;x)^{-(n+1)/2}G(d\rho)}{\int_{-1}^{1}(1-\rho^2)^{-(n-1)/2}[(1+\rho)/(n-(n-2)\rho)]^{1/2}Q(\rho;x)^{-(n+1)/2}G(d\rho)}$$

where

$$Q(\rho;x)=x_1^2+(1-\rho^2)^{-1}\sum_{i=1}^{n-1}(x_{i+1}-\rho x_i)^2$$

$$-(1+\rho)^{-1}(n-(n-2)\rho)^{-1}[(1+\rho)x_1+\sum_{i=i}^{n-1}(x_{i+1}-\rho x_i)]^2.$$

After the tedious calculation we have that

$$Q(\rho;x)=s^2(x)[Q-(1-\rho)\rho^2NP^2]/(1-\rho^2)$$

where

$$N=n-(n-2)\rho,\qquad \bar{x}=\sum_{i=1}^{n}x_i/n,\qquad s^2(x)=\sum_{i=1}^{n}(x_i-\bar{x})^2,$$

$$a_1=(x_1-\bar{x})/s(x),\quad a_n=(x_n-\bar{x})/s(x),\quad P=(a_1+a_n)/N,$$

$$(3.3)\qquad r=\sum_{i=1}^{n-1}(x_{i+1}-\bar{x})(x_i-\bar{x})/s^2(x),$$

and

$$Q=1+\rho^2-\rho^2(a_1^2+a_n^2)-2\rho r$$

(cf. Section 3 of Haq [1]).

For moderate n, a_1 and a_n may be considered small and can be ignored. The expression (3.2) can be written as approximately

$$G(d\rho\,|x)=\frac{(1-\rho^2)[(1+\rho)/(1-\rho)]^{1/2}(1+\rho^2-2\rho r)^{-(n+1)/2}G(d\rho)}{\int_{-1}^{1}(1-\rho^2)[(1+\rho)/(1-\rho)]^{1/2}(1+\rho^2-2\rho r)^{-(n+1)/2}G(d\rho)}$$

where r is (3.3). Hence it follows from Theorem 2 that the Bayes invariant predictor is given by

$$\delta_G(x)=\int_{-1}^{1}\delta_\rho(x)G(d\rho\,|x)$$

where δ_ρ is (3.1). Another aproximation of δ_G is obtained by $\delta_{\hat{\rho}}$ where $\hat{\rho}$ is the mode of $G(d\rho\,|x)$. It follows from Theorem 3 that δ_G is admissible among the class of invariant predictors.

REFERENCES

[1] Haq, M. S., Structual analysis for the first order autoregressive stochastic models, Ann. Math. Statist., **41**(1970), 970–978.

[2] Hora, R. B. and Buehler, R. J., Fiducial theory and invariant prediction, Ann. Math. Statist., **38**(1967), 795–801.

[3] Takada, Y., Relation of the best invariant predictor and the best unbiased predictor in location and scale families, Ann. Statist., **9**(1981), 917–921.

[4] Takada, Y., A comment on best invariant predictors, Ann. Statist., **10** (1982), 971–978.

[5] Zidek, J. V., A representation of Bayes invariant procedures in terms of Haar measure, Ann. Inst. Statist. Math., **21**(1969), 291–308.

Department of Mathematics
Faculty of Science
Kumamoto University