Kumamoto J. Sci. (Math.)
Vol. 15, 79～89 March (1983)

INVARIANT PREDICTION REGION OF FUTURE OBSERVATIONS

Yoshikazu TAKADA

(Received November 10, 1982)

1. Introduction

We consider a statistical prediction problem invariant under a certain group of transformations. Hora and Buehler [2] treated the problem of the point prediction and obtained a representation of the best invariant predictor by using the Haar measure on the group. Takada [6] extended the assumptions used by them and obtained an alternative expression of the best invariant predictor which is more suitable for applications.

In this paper we consider the problem of the prediction region. We are concerned with two risks for each prediction regions, the size and average volume of the region. The best invariant prediction region is defined as the one which has the smallest average volume among the class of all invariant prediction regions with the preassigned size $1-\varepsilon$. The purpose of this paper is to obtain a representation of the best invariant prediction region by using the Haar measure on the group. The method based on an adequate statistic is considered by Ishii [3] and Takada [7].

In Section 2, we define the statistical prediction problem. Under several assumptions, we express the best invariant prediction region by using the Haar measure on the group in Section 3. In Section 4, we obtain an alternative expression of the best invariant prediction region. In Section 5, some application is considered.

2. Group invariant structure of the prediction problem

Let X be an observable random variable and Y a future (therefore unobservable) random variable. Let $(\mathfrak{X}, \mathfrak{B})$ and $(\mathfrak{Y}, \mathfrak{C})$ be spaces of X and Y, respectively. Let $(\mathfrak{Z}, \mathfrak{A})=(\mathfrak{X}\times\mathfrak{Y}, \mathfrak{B}\times\mathfrak{C})$ and $\mathfrak{P}=\{P_\theta;\ \theta\varepsilon\Theta\}$ be a family of probability measures on $(\mathfrak{Z}, \mathfrak{A})$ such that $Z=(X, Y)$ is distributed according to P_θ, $\theta\varepsilon\Theta$, and Θ

a parameter space. Let $\mathfrak{G}$ be a group of one-to-one transformations acting on the space $\mathfrak{X}$, $\mathfrak{Z}$ and Θ, mapping each onto itself, and let $\tilde{\mathfrak{G}}$ be a group of transformations on $\mathfrak{Y}$.

ASSUMPTION 1. $\mathfrak{P}$ is invariant under $\mathfrak{G}$, that is,

$$P_{g\theta}(gA)=P_\theta(A), \qquad A\varepsilon\,\mathfrak{A},\ g\varepsilon\mathfrak{G},\ \theta\varepsilon\Theta$$

and $\mathfrak{G}$ satisfies that

$$g(x,y)=(gx,[g;x]y), \qquad g\varepsilon\mathfrak{G},\ x\varepsilon\mathfrak{X},\ y\varepsilon\mathfrak{Y}, \tag{2. 1}$$

where $[g;x]\ \varepsilon\tilde{\mathfrak{G}}$

After observing $X=x$, we want to construct the region in which the value of Y is contained. We consider a randomized prediction region, which is defined as follows. Let ϕ be a measurable function on $\mathfrak{Z}$ with the value in $[0,1]$. After observing $X=x$, the prediction region is given by $\{y;\ \phi(x,y)\geq u\}$, where u is a realized value of uniform random variable on $[0,1]$ which is independent of (X,Y).

DEFINITION 1. A prediction region is said to have a size $1-\varepsilon$ if

$$E_\theta\{\phi(X,Y)\}\geq 1-\varepsilon, \qquad \theta\varepsilon\Theta.$$

Another risk of a prediction region that we adopt here is the average volume of the region,

$$E_\theta\{\int\phi(X,y)\xi(dy)\}, \tag{2. 2}$$

where ξ is some σ-finite measure on $(\mathfrak{Y},\mathfrak{C})$.

DEFINITION 2. A prediction region is said to be invariant under $\mathfrak{G}$ if

$$\phi(g(x,y))=\phi(x,y), \qquad g\varepsilon\mathfrak{G},\ x\varepsilon\mathfrak{X},\ y\varepsilon\mathfrak{Y}.$$

An invariant prediction region is said to be best if it minimizes (2.2) for all $\theta\varepsilon\Theta$

among the class of all invariant prediction regions with size $1-\varepsilon$.

ASSUMPTION 2. ξ is a relatively invariant measure with modulus J, i. e.,

$$\xi(\tilde{g}C)=J(\tilde{g})\xi(C),\ \tilde{g}\varepsilon\tilde{\mathfrak{G}},\ C\varepsilon\mathfrak{C},$$

and for any $g\varepsilon\mathfrak{G}$, $J([g;x])$ does not depend on $x\varepsilon\mathfrak{X}$.

Therefore for simplicity, we shall write $J(g)$ instead of $J([g;x])$.

ASSUMPTION 3. Θ is isomorphic to $\mathfrak{G}$.

Let $\theta_0\varepsilon\Theta$ be the point to corresponding to the identity element e of $\mathfrak{G}$. The isomorphism is established by $\theta=g\theta_0$ if $\theta\varepsilon\Theta$ corresponds to $g\varepsilon\mathfrak{G}$. We shall identify the group element g with the parameter value θ and simplify the notation by letting θ designate g, so that we shall consider $\mathfrak{G}=\Theta$

ASSUMPTION 4. $\tilde{\mathfrak{G}}$ acts freely on $\mathfrak{Y}$, i. e., if $\tilde{g}\neq\tilde{e}$, $\tilde{g}y\neq y$ for any $y\varepsilon\mathfrak{Y}$ and $\tilde{g}\varepsilon\tilde{\mathfrak{G}}$ where $\tilde{e}$ is the identity element of $\tilde{\mathfrak{G}}$.

The following lemma states the basic property of the transformation $[g;x]$ introduced in (2.1). For the proof, see Lemma 2 in [6].

LEMMA 1. *If Assumption 4 holds, then for any g, $g'\varepsilon\mathfrak{G}$ and $x\varepsilon\mathfrak{X}$,*

$$[g';gx]=[g';gx]\,[g;x], \tag{2.3}$$

$$[g;x]^{-1}=[g^{-1};gx]. \tag{2.4}$$

For a prediction region ϕ, let

$$\gamma_1(\theta,\phi)=E_\theta\{1-\phi(X,Y)\}, \tag{2.5}$$

$$\gamma_2(\theta,\phi)=J(\theta^{-1})E_\theta\{\int\phi(X,y)\xi(dy)\}. \tag{2.6}$$

LEMMA 2. *If Assumptions 1 through 4 hold, then for any invariant prediction region ϕ, $\gamma_1(\theta,\phi)$ and $\gamma_2(\theta,\phi)$ does not depend on $\theta\varepsilon\Theta$.*

PROOF. It is easy to see that $\gamma_1(\theta,\phi)$ does not depend on $\theta\varepsilon\Theta$. By the invariance of ϕ and Assumption 1,

$$\gamma_2(\theta,\phi)=J(\theta^{-1})E_\theta\{\int\phi(\theta^{-1}X,[\theta^{-1};X]y)\xi(dy)\}$$
$$=E_\theta\{\int\phi(\theta^{-1}X,y)\,\xi(dy)\}$$

Since the family of probability distributions of X is invariant under $\mathfrak{G}$ by (2.1),

$$\gamma_2(\theta,\phi)=E_{\theta_0}\{\int\phi(X,y)\,\xi(dy)\},$$

which implies the result.

3. Representation of the best invariant prediction region

In this section, we shall express the best invariant prediction region by using the Haar measure on the group.

ASSUMPTION 5. $\mathfrak{G}$ is a locally compact topological group with a σ-field $\mathfrak{L}$.

Let μ and ν, respectively, denote the left and right Haar measures on $(\mathfrak{G},\mathfrak{L})$ and Δ denote the modular function. For the details, see Chapter 2 of Nachbin [4].

ASSUMPTION 6. There exists a space $\mathfrak{M}$ and a one-to-one bimeasurable map π from $\mathfrak{X}$ onto $\mathfrak{G}\times\mathfrak{M}$ such that if $\pi(x)=(h,a)$, then $\pi(gx)=(gh,a)$.

To simlify the presentation, we shall put $x=(h,a)$ and $gx=(gh,a)$. Assumptions 3 and 6 imply that the probability measure on $\mathfrak{M}$ induced from X does not depend on $\theta\varepsilon\Theta$. Hence we shall denote it by λ.

ASSUMPTION 7. The density function of X with respect to $\mu\times\lambda$ can be expressed in the form

$$f_1(\theta^{-1}h,a),\qquad h\varepsilon\mathfrak{G},\ a\varepsilon\mathfrak{M},\ \theta\varepsilon\Theta, \tag{3.1}$$

whereas, given $X=x$, the conditional density function of Y with respect to ξ can be expressed in the form

(3.2) $$f_2([\theta^{-1};\ x]y|\theta^{-1}x)J(\theta^{-1}), \qquad y\varepsilon\mathfrak{Y},\ x\varepsilon\mathfrak{X},\ \theta\varepsilon\Theta,$$

where $f_1(h,a)$ and $f_2(y|x)$ are the density function and conditional density function under P_{θ_0}, respectively.

Now we shall express the best invariant prediction region by using the right Haar measure ν. For this we need the following lemma.

LEMMA 3. *If Assumptions* 1 *through* 7 *hold and if* ϕ *is an invariant prediction region, then for any* $h\varepsilon\mathfrak{G}$,

(3.3) $$\gamma_1(\theta_0,\phi)=\Delta(h)\iiint[1-\phi(h,a,y)]f_1(\theta^{-1}h,a)f_2([\theta^{-1};h,a]y|\theta^{-1}h,a)J(\theta^{-1}) \times \xi(dy)\nu(d\theta)\lambda(da),$$

(3.4) $$\gamma_2(\theta_0,\phi)=\Delta(h)\iiint\phi(h,a,y)f_1(\theta^{-1}h,a)J(\theta^{-1})\xi(dy)\nu(d\theta)\lambda(da).$$

PROOF. By Assumptions 6 and 7,

$$\gamma_1(\theta_0,\phi)=\iiint[1-\phi(g,a,y)]f_1(g,a)f_2(y|g,a)\xi(dy)\mu(dg)\lambda(da)$$
$$=\Delta(h)\iiint[1-\phi(g'h,a,y)]f_1(g'h,a)f_2(y|g'h,a)\xi(dy)\mu(dg')\lambda(da),$$

where the second equality follows from the transformation $g=g'h$ and the fact that $\mu(dg)=\Delta(h)\mu(dg')$.

By the transfromation $y'=[g';\ h,a]^{-1}y$ and the invariance of ϕ,

$$\gamma_1(\theta_0,\phi)=\Delta(h)\iiint[1-\phi(h,a,y')]f_1(g'h,a)f_2([g';\ h,a]y'|g'h,a)J(g') \times\xi(dy')\mu(dg')\lambda(da).$$

Then by the transformation $\theta=g^{-1}$ and the fact that $\nu(d\theta)=\mu(dg')$, we have (3.3). By the similar way, (3.4) is obtained.

On the basis of Lemma 3, we sall prove that the following prediction region is the best invariant prediction region;

$$\text{(3.5)}\qquad \begin{aligned}\phi^*(x,y)&=1 && \text{if } F(x,y)>c,\\ &=r && \text{if } F(x,y)=c,\\ &=0 && \text{if } F(x,y)<c,\end{aligned}$$

where

$$\text{(3.6)}\qquad F(x,y)=\frac{\int f_1(\theta^{-1}h,a)f_2([\theta^{-1};\,x]y\,|\,\theta^{-1}x)J(\theta^{-1})\nu(d\theta)}{\int f_1(\theta^{-1}h,a)J(\theta^{-1})\nu(d\theta)}$$

and c and r are constants such that $c>0$ and $0\leq r\leq 1$.

THEOREM 1. *If Assumptions* 1 *through* 7 *hold, then the prediction region* ϕ^* *given by* (3.5) *is the best invariant prediction region where c and r are chosen such that* $E_{\theta_0}\{\phi^*(X,Y)\}=1-\varepsilon$.

PROOF. First we shall show that ϕ^* is invariant. Substituting $g(x,y)=(gx,[g;\,x]y)$ in place of (x,y), and using the transformation $\theta=g\theta'$ and the fact that $\nu(d\theta)=\Delta(g^{-1})\nu(d\theta')$, we have that

$$\text{(3.7)}\qquad \begin{aligned}&\int f_1(\theta^{-1}gh,a)f_2([\theta^{-1};\,gh,a][g;h,a]y\,|\,\theta^{-1}gh,a)J(\theta^{-1})\nu(d\theta)\\ &=\Delta(g^{-1})\int f_1(\theta'^{-1}h,a)f_2([(g\theta')^{-1};gh,a][g;h,a]y\,|\,\theta'^{-1}h,a)J((g\theta')^{-1})\nu(d\theta').\end{aligned}$$

Since by (2.3) and (2.4)

$$\begin{aligned}[(g\theta')^{-1};gh,a][g;h,a]&=[\theta'^{-1};h,a]\,[g^{-1};gh,a][g;h,a]\\ &=[\theta'^{-1};h,a]\end{aligned}$$

and $J((g\theta')^{-1})=J(g^{-1})J(\theta'^{-1})$, (3.7) becomes

$$\Delta(g^{-1})J(g^{-1})\int f_1(\theta^{-1}h,a)f_2([\theta^{-1};h,a]y\,|\,\theta^{-1}h,a)J(\theta^{-1})\nu(d\theta).$$

By the similar way, we have

$$\int f_1(\theta^{-1}gh,a)J(\theta^{-1})\nu(d\theta)=\Delta(g^{-1})J(g^{-1})\int f_1(\theta^{-1}h,a)J(\theta^{-1})\nu(d\theta),$$

so that $F(g(x,y))=F(x,y)$, which implies that ϕ^* is invariant.

Now we shall show that ϕ^* is the best invariant. Let ϕ be any invariant prediction region withe size $1-\varepsilon$. Then from (3.5) and (3.6),

$$\{\phi(x,y)-\phi^*(x,y)\}\left\{\int f_1(\theta^{-1}h,a)f_2([\theta^{-1};x]y|\theta^{-1}x)J(\theta^{-1})\nu(d\theta)\right.$$
$$\left.-c\int f_1(\theta^{-1}h,a)J(\theta^{-1})\nu(d\theta)\right\}\leq 0,$$

so that by Lemma 3,

$$\gamma_1(\theta_0,\phi^*)-\gamma_1(\theta_0,\phi)\leq c\{\gamma_2(\theta_0,\phi)-\gamma_2(\theta_0,\phi^*)\}.$$

Hence by Lemma 2 and $c>0$, we have the result.

4. Alternative expression of the best invariant prediction region

The main difficulty in applying Theorem 1 to a specific prediction problem is to verify Assumptions 6 and 7. Therefore we shall present sufficient conditions for them, assuming always Assumptions 1 to 5. This enables us to rewrite the best invariant prediction region in a form which is more tractable for some applications.

CONDITION 1. There exists a relatively invariant measure η on $(\mathfrak{X},\mathfrak{B})$ w!th modulus J_1 and $\mathfrak{P}$ is dominated by $\eta\times\xi$ and the density function of $Z=(X,Y)$ can be expressed by

$$J_1(\theta^{-1})J(\theta^{-1})p(\theta^{-1}z),\qquad z\varepsilon\mathfrak{Z},\ \theta\varepsilon\Theta. \tag{4.1}$$

Then by (2.1) the density function of X with respect to η is given by

$$J_1(\theta^{-1})p_1(\theta^{-1}x),\qquad x\varepsilon\mathfrak{X},\ \theta\varepsilon\Theta, \tag{4.2}$$

where $p_1(x)=\int p(x,y)\xi(dy)$.

CONDITION 2. There exists a Borel set $B\varepsilon\mathfrak{B}$ (Borel cross section) which intersects each orbit $\mathfrak{G}x=\{gx;\ g\varepsilon\mathfrak{G}\}$ precisely once.

CONDITION 3. $\mathfrak{X}$ is a separable complete metrizable locally compact space and $\mathfrak{G}$ is a separable complete metrizable locally compact topological group acting freely and continuously on $\mathfrak{X}$.

Under these conditions, it follows from Lemma 4 in [6] that Assumptions 6 and 7 are satisfied by taking B as $\mathfrak{M}$, and that

$$f_1(\theta^{-1}h, a)=J_1(\theta^{-1})p_1(\theta^{-1}x) \Big/ \int J(g)p_1(ga)\nu(dg) \tag{4.3}$$

and

$$f_2([\theta^{-1}; x]y|\theta^{-1}x)=p(\theta^{-1}(x, y))/p_1(\theta^{-1}x), \tag{4.4}$$

where $x=ha$. Hence, we have the following result from (3.6), noticing that $J_1(\theta^{-1}h)=J_1(\theta^{-1})J_1(h)$.

THEOREM 2. *If Assumptions* 1 *through* 5 *and Conditions* 1 *through* 3 *hold, then the best invariant prediction region is constructed from* ϕ^* *given by* (3.5) *with*

$$F(x, y)=\int J_1(\theta^{-1})J(\theta^{-1})p(\theta^{-1}(x, y))\nu(d\theta) \Big/ \int J_1(\theta^{-1})J(\theta^{-1})p_1(\theta^{-1}x)\nu(d\theta), \tag{4.5}$$

where c and r are chosen such that $E_{\theta_0}\{\phi^*(X, Y)\}=1-\varepsilon$.

REMARK 1. By applying this result to location, scale, and location-scale parameter family, we have the result obtained by Takeuchi [8].

5. Example

In this section we consider the example in [6]. Let $X_1, \ldots, X_n, X_{n+1}$ be independently and identically distributed p-dimensional random vectors with the density function with respect to Lebesgue measure on E^p (p-dimensional Euclidean space),

$$|\Lambda|^{-1}f(||\Lambda^{-1}(x-\mu)||^2),$$

where f is some known function, Λ is a lower triangular matrix of order p with positive diagonal elements and $|\Lambda|$ denotes the determinant.

Suppose that $\theta=(\mu, \Lambda)$ is unknown. We shall denote by $G(m)$ the set of all lower triangular matrices of order m with positive diagonal elements. Then $\Theta=\{\theta=(\mu, \Lambda);\ \mu\varepsilon E^p,\ \Lambda\varepsilon G(p)\}$. The following partitions are used in the sequel:

$$(5.1)\qquad X_i=\begin{pmatrix}X_i^1\\ X_i^2\end{pmatrix},\ i=1,\ \dots,\ n+1,\qquad \mu=\begin{pmatrix}\mu_1\\ \mu_2\end{pmatrix},\qquad \Lambda=\begin{pmatrix}\Lambda_{11} & 0\\ \Lambda_{21} & \Lambda_{22}\end{pmatrix},$$

where X_i^1 and μ_1 are $p_1\times 1$, $\Lambda_{11}\varepsilon G(p_1)$ and $\Lambda_{22}\varepsilon G(p_2)$ $(p_1+p_2=p)$.

We consider the problem of predicting $Y=X_{n+1}^2$ after observing $X=(X_1,\dots, X_n, X_{n+1}^1)$. Define the following transformation $g\varepsilon\ \mathfrak{G}$ on $\mathfrak{Z}$,

$$g(x_1,\dots, x_n, x_{n+1})=(b+Cx_1,\dots, b+Cx_n, b+Cx_{n+1}),\ g=(b, C),$$

where $b\varepsilon E^p$ and $C\varepsilon G(p)$. We shall view $\mathfrak{G}$ as the Cartesian product $E^P\times G(p)$ with the following group operation;

$$(b_1, C_1)(b_2, C_2)=(b_1+C_1b_2, C_1C_2),\qquad (b, C)^{-1}=(-C^{-1}b, C^{-1}).$$

Then it is well known that $\mathfrak{G}$ is a locally compact topological group and that the right Haar measure is given by

$$(5.2)\qquad \nu(d\theta)=\prod_{i=1}^{p}(\lambda_{ii})^{-(p+1-i)}d\mu d\Lambda,$$

where λ_{ii} $(i=1,\dots,p)$ are the diagonal elements of Λ, $d\mu$ and $d\Lambda$ denote Lebesgue measures on E^p and $G(p)$, respectively (see Fraser [1], p. 148).

Taking Lebesgue measure on E^{p_1} as ξ, we [6] showed that Assumptions and Conditions in Theorem 2 are satisfied, and that for $\theta=(\mu, \Lambda)$

$$J_1(\theta^{-1})=|\Lambda|^{-n}|\Lambda_{11}|^{-1},\quad J(\theta^{-1})=|\Lambda_{22}|^{-1}$$

and

$$p(\theta^{-1}(x, y))=\prod_{=1}^{n+} f(||\Lambda^{-1}(x_i-\mu)||^2),$$

where the same partition as (5.1) is used for Λ.

Hence by Theorem 2, the best invariant prediction region is constructed from ϕ^* given by (3.5) with

$$(5.3)\qquad F(x,y)=\frac{\int |A|^{-(n+1)}\prod_{i=1}^{n+1} f(||A^{-1}(x_i-\mu)||^2)\nu(d\theta)}{\int |A|^{-(n+1)}\prod_{i=1}^{n} f(||A^{-1}(x_i-\mu)||^2)f^*(||A_{11}^{-1}(x_{n+1}^1-\mu_1)||^2)\nu(d\theta)},$$

where $f^*(||u||^2)=\int f(||u||^2+||v||^2)dv$ ($u\varepsilon E^{p_1}$ and $v\varepsilon E^{p_2}$) and ν is (5.2).

Suppose that the random vectors are normal. Set

$$\bar{X}=\sum_{i=1}^{n} X_i/n, \qquad S=\sum_{i=1}^{n}(X_i-\bar{X})(X_i-\bar{X})',$$

and

$$W=(w_1,\ldots,w_p)'=A^{-1}(X_{n+1}-\bar{X}),$$

where $AA'=S$ and $A\varepsilon G(p)$.

Then the tedious but straightforward calculation shows that $F(x,y)$ defined by (5.3) is proportional to the conditional density function of $W_2=(w_{p_1+1},\ldots,w_p)'$ given $W_1=(w_1,\ldots,w_{p_1})'$, which is given by

$$h(W_2|W_1)=\prod_{\rho=p_1+1}^{p}\frac{\Gamma\left(\frac{n+1-\rho}{2}\right)(\sigma_\rho^2)^{-1/2}}{\pi^{1/2}\Gamma\left(\frac{n-\rho}{2}\right)(1+\sigma_\rho^{-2}w_\rho^2)^{\frac{n+1-\rho}{2}}},$$

where $\sigma_\rho^2=1+\frac{1}{n}+\sum_{i=1}^{\rho-1}w_i^2$.

REMARK. In the case of the normal distribution, the same conclusion can be obtained by using an adequate statistic. For the detailes, see [3] and [7], though the conditional density function of W_2 given W_1 which is calculated in [3] is not correct.

The best invariant prediction is very complicated. Schervish [5] proposed an invariant prediction region which is more useful than the best invariant prediction region from the practical point of view.

References

[1] FRASER, D. A. S., The Structure of Inference, Wiley, New York, 1968.
[2] HORA, R. B. and BUEHLER, R. J., Fiducial theory and invariant prediction, Ann. Math. Statist., **38** (1967), 795-801.
[3] ISHII, G., Best invariant prediction region based on an adequate statistic, in Recent Developments in Statistical Inference and the Data Analysis, K. Matusita, editor, North-Holland, 1980.
[4] NACHBIN, L., The Haar Integral, Van Nostrand, Princeton, 1965.
[5] SCHERVISH, M. J., Tolerance regions for random vectors and best linear predictors, Commun. Statist. -Theor. Math., **A9** (1980), 1177-1183.
[6] TAKADA, Y., A comment on best invariant predictors, Ann. Statist., **10** (1982), 971-978.
[7] TAKADA, Y., Application of an adequate statistic to the invariant prediction region, to appear in Ann. Inst. Statist. Math., 1982.
[8] TAKEUCHI, K., Statistical Prediction Theory (in Japanese), Baihūkan, Tōkyo, 1975.

Department of Mathematics
Faculty of Science
Kumamoto University