*The Annals of Statistics*
1982, Vol. 10, No. 3, 971–978

# A COMMENT ON BEST INVARIANT PREDICTORS

BY YOSHIKAZU TAKADA

*Kumamoto University*

A statistical prediction problem invariant under a certain group of transformations is considered. Under several assumptions it is shown that the best invariant predictor is obtained by using the invariant measure on the group. This result is an extension of that of Hora and Buehler.

**1. Introduction.** In this paper we discuss a statistical prediction problem invariant under a certain group of transformations. Takada (1981b) obtained the expression of the best invariant predictor (that is, an invariant predictor that is as good as any other invariant predictor) on the basis of the best unbiased predictor. The purpose of this paper is to represent the best invariant predictor by using the invariant measure on the group.

The representation by using the invariant measure was first treated by Hora and Buehler (1967). In Section 2 we shall extend the assumptions used by them and obtain the representation of the best invariant predictor by using the invariant measure. In Section 3 we shall discuss conditions required to satisfy the assumptions. Under the conditions, we can represent the best invariant predictor in more suitable form for applications. In Section 4 an example is given.

**2. Representation of the best invariant predictor.** Suppose $X$ is an observable random vector with sample space $\mathscr{X}$ and $Y$ a future random vector with sample space $\mathscr{Y}$. Let $\mathscr{Z} = \mathscr{X} \times \mathscr{Y}$ and $\mathscr{P} = \{P_\theta : \theta \in \Theta\}$ be a family of probability measures on $\mathscr{Z}$ such that $Z = (X, Y)$ is distributed according to $P_\theta$, $\theta \in \Theta$ and $\Theta$ a parameter space. Let $G$ be a group of one-to-one transformations acting on the spaces $\mathscr{X}$, $\mathscr{Z}$ and $\Theta$, mapping each onto itself, and let $\tilde{G}$ be a group of transformations on $\mathscr{Y}$.

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

$$P_{g\theta}(gB) = P_\theta(B)$$

for any Borel set $B$ of $\mathscr{Z}$, $g \in G$ and $\theta \in \Theta$, and $G$ is a locally compact topological group such that

$$g(x, y) = (gx, [g; x]y), \quad g \in G, \quad x \in \mathscr{X}, \quad y \in \mathscr{Y}, \tag{2.1}$$

where $[g; x] \in \tilde{G}$.

See Section 4 for the case that the transformation on $\mathscr{Y}$ does depend on $x \in \mathscr{X}$. We shall write $\mu$ and $\nu$ for the left and right invariant measures on $G$, respectively and $\Delta$ for the modulus of $G$; e.g. see Chapter 2 of Nachbin (1965).

After observing $X = x$, we want to predict the value of $Y$. A non-negative loss function $L(d, y, \theta)$ defined on $\mathscr{Y} \times \mathscr{Y} \times \Theta$ represents the loss of erroneously predicting $Y = y$ by $d$ under the true value $\theta$.

ASSUMPTION 2. $L$ is invariant under $G$, that is,

$$L([g; x]\, d, [g; x]y, g\theta) = L(d, y, \theta) \tag{2.2}$$

for all $d$, $x$, $y$, $\theta$.

Received August 1981; revised December 1981.
*AMS* 1980 *subject classification*. Primary 62A05; secondary 62F99.
*Key words and phrases*. Invariant measure, transformation group, invariant predictor.

In this paper we shall confine our attention to non-randomized predictors. Let $\delta$ be a predictor of $Y$ and

$$R(\theta, \delta) = E_\theta L(\delta(X), Y, \theta),$$

where $E_\theta$ denotes the expectation under $P_\theta$. $R(\theta, \delta)$ is called a risk function of the predictor $\delta$.

By (2.1) a predictor $\delta$ is said to be invariant under $G$ if

$$\delta(gx) = [g; x]\delta(x) \tag{2.3}$$

for any $g \in G$ and $x \in \mathscr{X}$.

ASSUMPTION 3. $\Theta$ is isomorphic to $G$.

Let $\theta_0 \in \Theta$ be the point corresponding to the identity element $e$ of $G$. The isomorphism is established by $\theta = g\theta_0$ if $\theta$ corresponds to $g \in G$. We shall identify the group element $g$ with the parameter value $\theta$ and simplify the notation by letting $\theta$ designate $g$, so that we shall consider $G = \Theta$. Under this meaning, it is supposed that $L$ is jointly measurable in its three arguments.

By (2.1) and (2.2) the risk function of a predictor $\delta$ can be written as

$$R(\theta, \delta) = E_{\theta_0} L(\delta(\theta X), [\theta; X]Y, \theta) = E_{\theta_0} L([\theta; X]^{-1}\delta(\theta X), Y, \theta_0),$$

so that it follows from (2.3) that for any invariant predictor $\delta$

$$R(\theta, \delta) = R(\theta_0, \delta), \quad \theta \in \Theta. \tag{2.4}$$

ASSUMPTION 4. There is a one-to-one bimeasurable map $\pi$ from $\mathscr{X}$ onto $G \times A$ such that if $\pi(x) = (h, a)$, then $\pi(gx) = (gh, a)$, where $A$ is some space.

Usually $A$ is a sample space of the maximal invariant statistic defined on $\mathscr{X}$ with respect to $G$. To simplify the presentation, we shall put $x = (h, a)$ and $gx = (gh, a)$ if $\pi(x) = (h, a)$.

By (2.1) it is easy to see that the family of probability distributions of $X$ induced from $\mathscr{P}$ is invariant under $G$, so that Assumptions 3 and 4 imply that the probability measure on $A$ induced from $X$ does not depend on $\theta \in \Theta$. Hence we shall denote it by $\lambda$.

ASSUMPTION 5. There is a relatively invariant measure $\xi$ on $\mathscr{Y}$ with modulus $J$ with respect to $\tilde{G}$, i.e., $\xi(\tilde{g}C) = J(\tilde{g})\xi(C)$ for $\tilde{g} \in \tilde{G}$ and Borel set $C$ of $\mathscr{Y}$, and for any $g \in G$, $J([g; x])$ does not depend on $x \in \mathscr{X}$.

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

ASSUMPTION 6. The density function of $X$ with respect to $\mu \times \lambda$ can be expressed in the form

$$f_1(\theta^{-1}h, a), \quad h \in G, \quad a \in A, \quad \theta \in \Theta, \tag{2.5}$$

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

$$f_2([\theta^{-1}; x]y \mid \theta^{-1}x)J(\theta^{-1}), \quad y \in \mathscr{Y}, \quad \theta \in \Theta, \tag{2.6}$$

where $f_1(h, a)$ and $f_2(y \mid x)$ are the density function and conditional density function under $P_{\theta_0}$, respectively.

Then the risk function of an invariant predictor $\delta$ can be written as

$$R(\theta_0, \delta) = \int\int\int L(\delta(g, a), y, \theta_0) f(g, y \mid a)\lambda(da)\mu(dg)\xi(dy), \tag{2.7}$$

where

$$f(g, y \mid a) = f_1(g, a) f_2(y \mid g, a). \tag{2.8}$$

By (2.4) an invariant predictor is said to be best if it minimizes (2.7) among all invariant predictors.

Suppose $H$ is a group of transformations acting on some space $D$ and let $h_0$ be the identity element of $H$. Then $H$ is said to act freely on $D$ if $h \neq h_0$ implies $hd \neq d$ for any $d \in D$ and $h \in H$.

ASSUMPTION 7. $\tilde{G}$ acts freely on $\mathscr{Y}$.

The following lemma states the property of the transformation of $[g; x]$ introduced in (2.1).

LEMMA 1. *If Assumptions* 1 *and* 7 *hold, then for any* $g, g' \in G$ *and* $x \in \mathscr{X}$,

$$[g'g; x] = [g'; gx][g; x], \tag{2.9}$$

$$[g; x]^{-1} = [g^{-1}; gx]. \tag{2.10}$$

PROOF. By (2.1),

$$g'(g(x, y)) = g'(gx, [g; x]y) = (g'gx, [g'; gx][g; x]y).$$

Since this is equal to $(g'gx, [g'g; x]y)$, we have (2.9) by Assumption 7. Set $g' = g^{-1}$ in (2.9). Then by using the fact that $[e; x] = \tilde{e}$ where $\tilde{e}$ denotes the identity element of $\tilde{G}$, (2.10) is obtained.

Now we shall represent the best invariant predictor by using the right invariant measure $\nu$ on $G$. For this we need the following lemma.

LEMMA 2. *If Assumptions* 1 *to* 7 *hold and if* $\delta$ *is an invariant predictor, then for any* $h \in G$,

$$R(\theta_0, \delta) = \Delta(h) \int\int\int L(\delta(h, a), y, \theta) f(\theta^{-1}h, [\theta^{-1}; h, a]y \mid a) J(\theta^{-1}) \lambda(da) \nu(d\theta) \xi(dy).$$

PROOF. From (2.7) and the transformation $g = g'h$, we obtain

$$R(\theta_0, \delta) = \Delta(h) \int\int\int L(\delta(g'h, a), y, \theta_0) f(g'h, y \mid a) \lambda(da) \mu(dg') \xi(dy),$$

where we used the fact that $\mu(dg) = \Delta(h)\mu(dg')$.

By (2.2) and (2.3),

$$\begin{aligned} L(\delta(g'h, a), y, \theta_0) &= L([g'; h, a]\delta(h, a), y, \theta_0) \\ &= L(\delta(h, a), [g'; h, a]^{-1}y, g'^{-1}), \end{aligned}$$

so that after the transformation $y' = [g'; h, a]^{-1}y$ we obtain

$$R(\theta_0, \delta) = \Delta(h) \int\int\int L(\delta(h, a), y', g'^{-1}) f(g'h, [g'; h, a]y' \mid a) J(g') \lambda(da) \mu(dg') \xi(dy').$$

Then the lemma is obtained by the transformation $\theta = g'^{-1}$ and the fact that $\nu(d\theta) = \mu(dg')$.

On the basis of Lemma 2, we shall prove the following result, which is an extension of Theorem 2 of Hora and Buehler (1967).

THEOREM 1. *If Assumptions 1 to 7 hold and if there exists a predictor $\delta^*$ such that for each $x = (h, a)$, $\delta^*(x)$ is the unique value of d which minimizes*

$$\int\int L(d, y, \theta) f(\theta^{-1}h, [\theta^{-1}; x]y \mid a) J(\theta^{-1}) \nu(d\theta) \xi(dy), \tag{2.11}$$

*then $\delta^*$ is the best invariant predictor.*

PROOF. By Lemma 2, it is enough to show that $\delta^*$ satisfies (2.3). Substituting $gx = (gh, a)$ in place of $x = (h, a)$ in (2.11) and using the transformation $\theta = g\theta'$ and the fact that $\nu(d\theta) = \Delta(g^{-1})\nu(d\theta')$, we can write (2.11) as

$$\Delta(g^{-1}) \int\int L(d, y, g\theta') f(\theta'^{-1}h, [(g\theta')^{-1}; gh, a]y \mid a) J((g\theta')^{-1}) \nu(d\theta') \xi(dy). \tag{2.12}$$

Since by (2.9) and (2.10)

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

and $J((g\theta')^{-1}) = J(g^{-1})J(\theta'^{-1})$, after the transformation $y' = [g; h, a]^{-1}y$, (2.12) becomes

$$\Delta(g^{-1}) \int\int L(d, [g; h, a]y', g\theta') f(\theta'^{-1}h, [\theta'^{-1}; h, a]y' \mid a) J(\theta'^{-1}) \nu(d\theta') \xi(dy')$$

$$= \Delta(g^{-1}) \int\int L([g; x]^{-1}d, y', \theta') f(\theta'^{-1}h, [\theta'^{-1}; x]y \mid a) J(\theta'^{-1}) \nu(d\theta') \xi(dy'),$$

where we used (2.2). Therefore from the definition of $\delta^*$ we obtain that

$$\delta^*(gx) = [g; x]\delta^*(x),$$

which completes the proof of the theorem.

REMARK 1. By an argument similar to that in Kudo (1955) or Kiefer (1957), it is possible to prove that the best invariant predictor is minimax.

**3. Sufficient conditions for the assumptions.** The main difficulty in applying Theorem 1 to a specific problem is to verify Assumptions 4 and 6, so that we shall present a set of sufficient conditions for them, assuming always other assumptions.

CONDITION 1. There exists a relatively invariant measure $\eta$ on $\mathscr{X}$ with modulus $J_1$ with respect to $G$ and $\mathscr{P}$ is dominated by $\eta \times \xi$ and the probability density function of $Z = (X, Y)$ can be expressed by

$$J_1(\theta^{-1})J(\theta^{-1})p(\theta^{-1}z), \quad z \in \mathscr{Z}, \quad \theta \in \Theta, \tag{3.1}$$

where $p(z)$ denotes the probability density function of $Z$ under $P_{\theta_0}$.

Then by (2.1) the density function of $X$ with respect to $\eta$ is given by

$$J_1(\theta^{-1})p_1(\theta^{-1}x), \quad x \in \mathscr{X}, \quad \theta \in \Theta, \tag{3.2}$$

where

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

CONDITION 2. $\mathscr{X}$ is a separable complete metrizable locally compact space and $G$ is a separable complete metrizable locally compact topological group acting freely and continuously on $\mathscr{X}$.

CONDITION 3. There exists a Borel set $A$ of $\mathscr{X}$ which intersects each orbit of $G$ in $\mathscr{X}$ precisely once.

The Borel set $A$ is called Borel cross-section. Then the following lemma holds. For a proof, see Theorem 1 of Bondar (1976).

LEMMA 3. *If Conditions* 2 *and* 3 *hold, then Assumption* 4 *is satisfied by taking A as the Borel cross-section, and if f is a real-valued function which is integrable with respect to* $\eta$, *then*

$$\int_{\mathscr{X}} f(x)\eta(dx) = \int_A \alpha(da) \int_G f(ha)J_1(h)\mu(dh) \tag{3.3}$$

*for some* $\sigma$-*finite measure* $\alpha$ *on* $A$.

Using this lemma, we shall verify Assumption 4.

LEMMA 4. *If Conditions* 1 *to* 3 *hold, then Assumption* 6 *is satisfied and for* $x = (h, a)$

$$f(h, y \mid a) = k(a)^{-1}J_1(h)p(x, y), \tag{3.4}$$

*where*

$$k(a) = \int J_1(g)p_1(ga)\mu(dg). \tag{3.5}$$

PROOF. From (3.2) and (3.3), the density function of $X$ with respect to $\mu \times \alpha$ is given by $J_1(\theta^{-1}h)p_1(\theta^{-1}ha)$. Since

$$\int J_1(\theta^{-1}h)p_1(\theta^{-1}ha)\mu(dh) = \int J_1(g)p_1(ga)\mu(dg)$$

and this is the density function of $\lambda$ with respect to $\alpha$, we have that for $x = (h, a)$

$$f_1(\theta^{-1}h, a) = J_1(\theta^{-1}h)p_1(\theta^{-1}x)/k(a), \tag{3.6}$$

where $k(a)$ is (3.5). Using (3.1) and (3.2), we obtain

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

so that from (3.6) Assumption 6 is satisfied and (3.4) is obtained from (2.8). This completes the proof of the lemma.

Now we shall prove more useful results than Theorem 1.

THEOREM 2. *If Assumptions* 1 *to* 3, 5 *and* 7 *and Conditions* 1 *to* 3 *hold and if there exists a predictor* $\delta^*$ *such that for each* $x \in \mathscr{X}$, $\delta^*(x)$ *is the unique value of d which minimizes*

$$\int\int L(d, y, \theta)J_1(\theta^{-1})J(\theta^{-1})p(\theta^{-1}(x, y))\nu(d\theta)\xi(dy), \tag{3.7}$$

*then* $\delta^*$ *is the best invariant predictor.*

PROOF. From Lemmas 3 and 4, the assumptions in Theorem 1 are satisfied. Hence from (2.1) and (3.4), (2.11) is equal to

$$k(a)^{-1}J_1(h)\int\int L(d,\,y,\,\theta)J_1(\theta^{-1})J(\theta^{-1})p(\theta^{-1}(x,\,y))\nu(d\theta)\xi(dy)$$

for $x = (h,\,a)$ since $J_1(\theta^{-1}h) = J_1(\theta^{-1})J_1(h)$. Therefore we have the result from Theorem 1.

**4. Example.** Let $X_1, \cdots, X_n, X_{n+1}$ be independently, identically distributed $p$-dimensional random vectors with the probability density function with respect to Lebesgue measure on $E^p$ ($p$-dimensional Euclidean space),

$$|\Lambda|^{-1}f(\|\Lambda^{-1}(x-\mu)\|^2), \quad x \in E^p, \tag{4.1}$$

where $f$ is some known function, $\|b\|^2 = b'b$, $\Lambda$ is a $p \times p$ lower triangular matrix 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 $m \times m$ lower triangular matrices with positive diagonal elements. The following partitions are used in the sequel:

$$X_i = \begin{pmatrix} X_i^1 \\ X_i^2 \end{pmatrix}, \quad i = 1, \cdots, n+1, \quad \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \quad \Lambda = \begin{pmatrix} \Lambda_{11} & 0 \\ \Lambda_{21} & \Lambda_{22} \end{pmatrix}, \tag{4.2}$$

where $X_i^1$ and $\mu_1$ are $p_1 \times 1$, $\Lambda_{11} \in G(p_1)$ and $\Lambda_{22} \in G(p_2)$, $p_1 + p_2 = p$.

We consider the problem of predicting $Y = X_{n+1}^2$ after observing $X = (X_1, \cdots, X_n, X_{n+1}^1)$ under the loss function

$$L(d,\,y,\,\theta) = \|\Lambda_{22}^{-1}(d-y)\|^2. \tag{4.3}$$

Let $G = \{(b,\,C) : b \in E^p,\, C \in G(p)\}$. We define the following transformation $g$ of $G$ on $\mathscr{X}$:

$$g(x_1, \cdots, x_n, x_{n+1}) = (b + Cx_1, \cdots, b + Cx_n, b + Cx_{n+1}) \tag{4.4}$$

for $g = (b,\,C) \in G$. Clearly (4.4) satisfies (2.1) by setting

$$gx = (b + Cx_1, \cdots, b + Cx_n, b_1 + C_{11}x_{n+1}^1) \tag{4.5}$$

and

$$[g;\,x]y = C_{22}y + C_{21}x_{n+1}^1 + b_2 \tag{4.6}$$

where the same partitions as (4.2) are used for $(b,\,C)$, $x = (x_1, \cdots, x_n, x_{n+1}^1)$ and $y = x_{n+1}^2$.

We shall view $G$ as the Cartesian product $E^p \times G(p)$ with group operation of $G$ in the following manner:

$$\begin{aligned}(b_1,\,C_1)(b_2,\,C_2) &= (b_1 + C_1b_2,\, C_1C_2),\\ (b,\,C)^{-1} &= (-C^{-1}b,\, C^{-1}).\end{aligned} \tag{4.7}$$

Then it is well known that $G$ is a locally compact topological group and that the right invariant measure is given by

$$\nu(d\theta) = \textstyle\prod_{i=1}^{p} (\lambda_{ii})^{-(p+1-i)}\, d\mu\, d\Lambda \tag{4.8}$$

where $\lambda_{ii}(i = 1, \cdots, p)$ are diagonal elements of $\Lambda$, $d\mu$ and $d\Lambda$ denote Lebesgue measures on $E^p$ and $G(p)$, respectively; see page 148 of Fraser (1968). Hence Assumption 1 holds. It is easy to see that Assumption 2 holds from (4.3), (4.6) and (4.7). From the definition of $G$, Assumption 3 is satisfied.

Let $\xi$ and $\eta$ be Lebesgue measures on $E^{p_2}$ and $E^{np+p_1}$, respectively. Then by (4.6)

$$J([g;\,x]) = |C_{22}|, \quad g = (b,\,C),$$

so that Assumption 5 holds. It is easy to see that Assumption 7 is satisfied. By (4.1) and

(4.7) Condition 1 is satisfied and for $\theta = (\mu, \Lambda)$,

$$J_1(\theta^{-1}) = |\Lambda|^{-n}|\Lambda_{11}|^{-1}, \quad J(\theta^{-1}) = |\Lambda_{22}|^{-1}, \tag{4.9}$$

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

It is easy to see that Condition 2 holds.

Let $\mathscr{X}_1$ be the sample space of $(X_1, \cdots, X_n)$ and define the action $g$ of $G$ on $\mathscr{X}_1$ by

$$g(x_1, \cdots, x_n) = (b + Cx_1, \cdots, b + Cx_n), \quad g = (b, C).$$

For this action, there exists a Borel cross-section for the orbits in $\mathscr{X}_1$; see page 145 of Fraser (1968). Hence from (4.5) and Proposition 2 of Bondar (1976), there exists a Borel cross-section for the orbits in $\mathscr{X}$, which implies Condition 3. Therefore by Theorem 2 we can obtain the best invariant predictor. By (4.3) and (4.9), (3.7) is equal to

$$\int\int \|\Lambda_{22}^{-1}(d - y)\|^2 |\Lambda|^{-(n+1)} \prod_{i=1}^{n+1} f(\|\Lambda^{-1}(x_i - \mu)\|^2)\nu(d\theta)\, dy,$$

where $\nu$ is (4.8). Hence the best invariant predictor is given by

$$\begin{aligned}\delta^*(x) = &\left\{\int\int (\Lambda_{22}\Lambda_{22}')^{-1}|\Lambda|^{-(n+1)} \prod_{i=1}^{n+1} f(\|\Lambda^{-1}(x_i - \mu)\|^2)\nu(d\theta)\, dy\right\}^{-1} \\ &\times \left\{\int\int (\Lambda_{22}\Lambda_{22}')^{-1}y|\Lambda|^{-(n+1)} \prod_{i=1}^{n+1} f(\|\Lambda^{-1}(x_i - \mu)\|^2)\nu(d\theta)\, dy\right\},\end{aligned} \tag{4.10}$$

where $x = (x_1, \cdots, x_n, x_{n+1}^1)$ and $y = x_{n+1}^2$.

Suppose the random vectors are normal. Then tedious and straightforward calculations show that (4.10) becomes

$$\delta^*(x) = \bar{x}_2 + S_{21}S_{11}^{-1}(x_{n+1}^1 - \bar{x}_1) \tag{4.11}$$

where

$$\bar{x} = \begin{pmatrix}\bar{x}_1 \\ \bar{x}_2\end{pmatrix} = \frac{1}{n}\sum_{i=1}^n x_i, \quad S = \begin{pmatrix}S_{11} & S_{12} \\ S_{21} & S_{22}\end{pmatrix} = \sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})'$$

and the same partitions as (4.2) are used for $(\bar{x}, S)$.

REMARK 2. The predictor given by (4.11) was proposed in Example 2 of Ishii (1969) as an unbiased predictor, though any optimality was not proved. It is also possible to derive the predictor on the basis of an adequate statistic introduced by Skibinsky (1967). For the details, see Example 4.1 of Takada (1981a).

Lee and Geisser (1975) treated the same problem as ours in a growth curve model and proposed a family of predictors which includes (4.11) as a special case.

**Acknowledgment.** The author wishes to thank the associate editor for his helpful comments. He also wishes to thank the referee for his valuable suggestions and for pointing out the paper of Lee and Geisser (1975).

## REFERENCES

BONDAR, J. (1976). On Borel cross-section and maximal invariants. *Ann. Statist.* 4 866–877.

FRASER, D. A. S. (1968). *The Structure of Inference.* Wiley, New York.

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

ISHII, G. (1969). Optimality of unbiased predictors. *Ann. Inst. Statist. Math.* **21** 471–488.

KIEFER, J. (1957). Invariance, minimax sequential estimation, and continuous time processes. *Ann. Math. Statist.* **28** 573–601.

KUDO, H. (1955). On minimax invariant estimates of transformation parameter. *Natural Science Report* **6** 31–73, Ochanomizu University.

LEE, J. C. and GEISSER, S. (1975). Applications of growth curve prediction. *Sankhyā* **37** 239–256.

NACHBIN, L. (1965). *The Haar Integral.* D. Van Nostrand, Princeton.

SKIBINSKY, M. (1967). Adequate subfields and sufficiency. *Ann. Math. Statist.* **38** 155–161.

TAKADA, Y. (1981a). Invariant prediction rules and an adequate statistic. *Ann. Inst. Statist. Math.* **33** 91–100.

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

DEPARTMENT OF MATHEMATICS
KUMAMOTO UNIVERSITY
KUMAMOTO 860
JAPAN