Ann. Inst. Statist. Math.
34 (1982), Part A, 491-503

APPLICATION OF AN ADEQUATE STATISTIC TO THE INVARIANT PREDICTION REGION

YOSHIKAZU TAKADA

(Received Oct. 27, 1981; revised Feb. 16, 1982)

1. Introduction

This paper deals with the problem of invariant prediction regions of a future observation on the basis of a past observation when the family of the joint probability distributions is invariant under a certain group of transformations. In prediction problems Skibinsky [10] and Takeuchi and Akahira [15] showed that the class of procedures based on an adequate statistic is essentially complete among all procedures. But it does not necessarily mean the essential completeness of the class of invariant procedures based on the adequate statistic among the class of all invariant procedures. If it is shown, we may confine our attention to those based on the adequate statistic for seeking the best invariant procedures.

For a fairly general prediction problem we [13] showed it under several assumptions. But the result can not be immediately applied to the problem of the prediction region treated in this paper. Hence, we shall show that the class of invariant prediction regions based on the adequate statistic is essentially complete in the class of all invariant prediction regions.

In Section 2 some result about an adequate statistic is stated. Using this result in Section 3, we shall prove the above result.

Ishii [7] obtained the best invariant prediction region among the class of those based on the adequate statistic. But the result (Theorem 1 in [7]) is not correct without an additional assumption (see Remark 5 in Section 4). So we shall again consider this problem in Section 4, though the assumptions we use are slightly different from those in [7]. In Section 5 we shall consider the examples in [7] and show that these examples satisfy the additional assumption and that they are best not only among the class of invariant prediction regions based on the adequate statistic but also among the class of all invariant prediction regions.

2. Adequate statistic

Let $(\mathcal{X}, \mathcal{B})$ and $(\mathcal{Y}, \mathcal{C})$ be sample spaces of observable random variable X and future random variable Y, respectively, and let $(\mathcal{Z}, \mathcal{A}) = (\mathcal{X}\times\mathcal{Y}, \mathcal{B}\times\mathcal{C})$ be the sample space of $Z=(X, Y)$. Suppose that the distribution of Z belongs to a family of probability measures $\mathcal{P}=\{P_\theta: \theta\in\Theta\}$ on $(\mathcal{Z}, \mathcal{A})$, which is indexed by a set Θ, called a parameter space.

Observing X, we are interested in the construction of a region in which Y will fall. Such a region is called a prediction region. We shall consider a randomized prediction region, which is constructed by the following way. Let ϕ be an $\mathcal{A}$ measurable function defined on $\mathcal{Z}$ such that $0\leqq\phi\leqq 1$. If $X=x$ is observed, the prediction region is given by a set $\{y: \phi(x, y)\geqq u\}$ where u is a realization of a uniform random variable on $[0, 1]$ independent of X. If ϕ takes only zero and one, it is called a non-randomized prediction region. In this case set $R(x)=\{y: \phi(x, y)=1\}$. Then the probability that Y is contained in the region, which is called the size of the prediction region, is $P_\theta(Y\in R(X))$. Let ξ be some σ-finite measure on $(\mathcal{Y}, \mathcal{C})$. We define the desirability of the prediction region by the average volume with respect to ξ, that is, $\mathrm{E}_\theta\, \xi(R(X))$ where E_θ denotes the expectation under P_θ. For a randomized prediction region ϕ it is easily shown that the size and average volume are given by $\mathrm{E}_\theta\, \phi(Z)$ and $\mathrm{E}_\theta \int \phi(X, y)\xi(dy)$, respectively. For the details, see Takeuchi [14], pp. 24–25.

DEFINITION 1. A prediction region ϕ is said to have confidence level $1-\varepsilon$ if

$$\mathrm{E}_\theta\, \phi(Z)\geqq 1-\varepsilon \qquad \text{for all } \theta\in\Theta .$$

We shall compare prediction regions of confidence level $1-\varepsilon$ by the average volumes.

Let t be a measurable mapping from $(\mathcal{X}, \mathcal{B})$ onto $(\mathcal{T}, \mathcal{U})$ and $\mathcal{B}_t = \{t^{-1}(U): U\in\mathcal{U}\}$. Set $\mathcal{B}'=\{B': B'=B\times\mathcal{Y},\ B\in\mathcal{B}\}$, $\mathcal{C}'=\{C': C'=\mathcal{X}\times C,\ C\in\mathcal{C}\}$, and $\mathcal{B}_t'=\{B': B'=B\times\mathcal{Y}, B\in\mathcal{B}_t\}$.

DEFINITION 2. A statistic $T=t(X)$ is said to be adequate for X with respect to Y and $\mathcal{P}$ if $\mathcal{B}_t'$ is sufficient for $\mathcal{B}'$, and $\mathcal{B}'$ and $\mathcal{C}'$ are conditionally independent given $\mathcal{B}_t'$.

ASSUMPTION 1. $\mathcal{P}$ is dominated by $\lambda=\lambda_1\times\lambda_2$ where λ_1 and λ_2 are probability measures on $(\mathcal{X}, \mathcal{B})$ and $(\mathcal{Y}, \mathcal{C})$, respectively.

By this assumption there exists a countable subfamily, $\{P_{\theta_1}, P_{\theta_2}, \cdots\}$, which is equivalent to $\mathcal{P}$ (Halmos and Savage [6]). We define a probability measure μ_1 on $(\mathscr{X}, \mathscr{B})$ whose density function with respect to λ_1 is $\sum_{i=1}^{\infty} c_i \int f_{\theta_i}(x, y)\lambda_2(dy)$ where $c_i>0$ $(i=1, 2, \cdots)$, $\sum_{i=1}^{\infty} c_i=1$ and f_{θ_i} denotes the density function of P_{θ_i} with respect to λ. Set $\lambda_0=\mu_1\times\lambda_2$. Then it is easy to see that $\mathscr{P}$ is dominated by λ_0. The following lemma which is obtained by Sugiura and Morimoto [11] is a factorization theorem for an adequate statistic.

LEMMA 1. *If Assumption* 1 *holds, then a statistic T is adequate if and only if for any $\theta\in\Theta$ the density function of P_θ with respect to λ_0 is $\mathscr{B}'_t\vee C'$ measurable where $\mathscr{B}'_t\vee C'$ denotes the smallest σ-field containing $\mathscr{B}'_t$ and C'.*

For any integrable function f with respect to λ_0 we denote the conditional expectation given $\mathscr{B}'_t\vee C'$ by $\mathrm{E}_{\lambda_0}(f\,|\,\mathscr{B}'_t\vee C')$. In particular, we set

$$\phi_0=\mathrm{E}_{\lambda_0}(\phi\,|\,\mathscr{B}'_t\vee C') \tag{2.1}$$

for a prediction region ϕ.

ASSUMPTION 2. ξ is dominated by λ_2.

LEMMA 2. *If Assumptions* 1 *and* 2 *hold, and if T is adequate, then it holds that for any prediction region ϕ*

$$\mathrm{E}_\theta\,\phi(Z)=\mathrm{E}_\theta\,\phi_0(Z) \tag{2.2}$$

and

$$\mathrm{E}_\theta\int\phi(X, y)\xi(dy)=\mathrm{E}_\theta\int\phi_0(X, y)\xi(dy) \tag{2.3}$$

where ϕ_0 is (2.1).

PROOF. Denote the density function of P_θ with respect to λ_0 by p_θ. Then for any $A\in\mathscr{A}$

$$\begin{aligned}\int_A\phi(z)P_\theta(dz)&=\int_A\phi(z)p_\theta(z)\lambda_0(dz)\\&=\int_A\mathrm{E}_{\lambda_0}(\phi p_\theta\,|\,\mathscr{B}'_t\vee C')(z)\lambda_0(dz)\\&=\int_A\phi_0(z)p_\theta(z)\lambda_0(dz)\end{aligned} \tag{2.4}$$

where the last equality follows from Lemma 1. Hence, by putting $A=\mathscr{Z}$, we have (2.2). Now, we shall prove (2.3). By Assumption 2 we denote the density function of ξ with respect to λ_2 by q. Since the density function of X with respect to μ_1 is given by $h_\theta(x)=\int p_\theta(x, y)\lambda_2(dy)$, we have

$$\text{(2.5)} \qquad \begin{aligned} E_\theta \int \phi(X, y)\xi(dy) &= \int\int \phi(x, y)h_\theta(x)q(y)\mu_1(dx)\lambda_2(dy) \\ &= \int E_{\lambda_0}(\phi h_\theta q \mid \mathcal{B}_t' \vee C')(z)\lambda_0(dz) \\ &= \int\int \phi_0(x, y)h_\theta(x)q(y)\mu_1(dx)\lambda_2(dy) \\ &= E_\theta \int \phi_0(X, y)\xi(dy) \end{aligned}$$

where we used the fact that $\lambda_0=\mu_1\times\lambda_2$ and $h_\theta(x)q(y)$ is $\mathcal{B}_t'\vee C'$ measurable. This completes the proof.

Remark 1. Though Takeuchi announced the result of Lemma 2 (see Theorem 5 in [14], p. 140), we have given the rigorous proof.

3. Invariant prediction region

Suppose that $T=t(X)$ is an adequate statistic with sample space $(\mathcal{T}, \mathcal{U})$. Let $\mathcal{G}$ denote a group of transformations on $\mathcal{Z}$, $\mathcal{T}\times\mathcal{Y}$ and Θ. Here we suppose that if the action space is a measurable space, the transformation is measurable. We define the mapping from $\mathcal{Z}$ onto $\mathcal{T}\times\mathcal{Y}$ by

$$\text{(3.1)} \qquad s(z)=(t(x), y)\ , \qquad z=(x, y)\ .$$

ASSUMPTION 3. $\mathcal{P}$ is invariant under $\mathcal{G}$, that is,

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

and $\mathcal{G}$ satisfies that

$$\text{(3.2)} \qquad gs'=s(gz)\ , \qquad g\in\mathcal{G}\ ,$$

for $s'\in\mathcal{T}\times\mathcal{Y}$ and $z\in\mathcal{Z}$ satisfying $s'=s(z)$.

It follows easily from (3.2) that for any $D\in\mathcal{U}\times C$

$$g(s^{-1}(D))=s^{-1}(gD)\ , \qquad g\in\mathcal{G}\ .$$

Hence, since $\mathcal{B}_t'\vee C'=\{s^{-1}(D);\ D\in\mathcal{U}\times C\}$, we have

$$\text{(3.3)} \qquad g(\mathcal{B}_t'\vee C')=\mathcal{B}_t'\vee C'\ , \qquad g\in\mathcal{G}\ .$$

DEFINITION 3. (1) A prediction region ϕ is said to be invariant with respect to $\mathcal{G}$ if for all $z\in\mathcal{Z}$ and $g\in\mathcal{G}$

$$\phi(gz)=\phi(z)\ .$$

(2) A prediction region ϕ is said to be almost invariant with respect

to $\mathcal{G}$ if for all $g \in \mathcal{G}$

$$\phi(gz)=\phi(z)\ , \qquad z \in \mathcal{Z}-N_g\ ,$$

where N_g is a $\mathcal{P}$ null set and permitted to depend on g.

LEMMA 3. *If Assumptions* 1 *to* 3 *hold and* ϕ *is an invariant prediction region, then* ϕ_0 *given by* (2.1) *is almost invariant.*

PROOF. It follows from (2.4) that ϕ_0 is a version of the conditional expectation of ϕ given $\mathcal{B}_t' \vee C'$ under P_θ. Hence combining (3.3) and Lemma 3.1 in Hall, et al. [5], it turns out that ϕ_0 is almost invariant.

Remark 2. If the mapping $s(z)$ given by (3.1) satisfies that $s(gz_1) = s(gz_2)$ whenever $s(z_1)=s(z_2)$, the transformation of $g \in \mathcal{G}$ on $\mathcal{T} \times \mathcal{Y}$ can be defined $gs'=s(gz)$ for $s' \in \mathcal{T} \times \mathcal{Y}$ and $z \in \mathcal{Z}$ satisfying $s'=s(z)$. Then (3.2) is satisfied if this transformation is measurable.

Now we shall show that the class of invariant prediction regions is essentially complete with respect to average volume among the class of almost invariant prediction regions which have confidence level $1-\varepsilon$ and are based on the adequate statistic.

ASSUMPTION 4. There exist σ-field $\mathcal{L}$ of $\mathcal{G}$ and σ-finite measure ν on $(\mathcal{G}, \mathcal{L})$ such that for any set $E \in \mathcal{U} \times C$ the set $\{(s, g);\ gs \in E\}$ belongs to $\mathcal{U} \times C \times \mathcal{L}$ and $\nu(B)=0$ implies $Bg \in \mathcal{L}$ and $\nu(Bg)=0$ for all $g \in \mathcal{G}$.

Then it follows from Assumption 1 and Theorem 4 in Lehmann ([8], p. 225) that for any almost invariant prediction region ϕ based on T, there exists an invariant prediction region $\tilde{\phi}$ based on T such that

$$\phi(z)=\tilde{\phi}(z) \qquad \text{a.e. } [\mathcal{P}]\ . \tag{3.4}$$

ASSUMPTION 5. There exists an invariant set $A \in \mathcal{B}_t' \vee C'$ ($gA=A$ for all $g \in \mathcal{G}$) such that $P_\theta(A)=1$ for all $\theta \in \Theta$ and $\mathcal{P}$ is equivalent to λ on A.

Set

$$\phi^*(z)=\begin{cases} \tilde{\phi}(z)\ , & \text{if } z \in A\ , \\ 0\ , & \text{otherwise}\ . \end{cases} \tag{3.5}$$

Then it follows from (3.4) and Assumption 5 that ϕ^* is an invariant prediction region based on T and

$$\text{(3.6)} \qquad \mathrm{E}_\theta\, \phi^*(Z) = \mathrm{E}_\theta\, \phi(Z)$$

and that on A

$$\text{(3.7)} \qquad \phi^*(z) = \phi(z) \qquad \text{a.e. } [\lambda] .$$

Hence from (2.5) and (3.7) we have that

$$\begin{aligned} \mathrm{E}_\theta \int \phi(X, y)\xi(dy) &= \int\int \phi(x, y)h_\theta(x)q(y)\mu_1(dx)\lambda_2(dy) \\ &\geqq \int\int_A \phi^*(x, y)h_\theta(x)q(y)\mu_1(dx)\lambda_2(dy) \\ &= \mathrm{E}_\theta \int \phi^*(X, y)\xi(dy) , \end{aligned}$$

where the last equality follows from (3.5). Therefore we have the following result from (3.6) and Lemmas 2 and 3.

THEOREM 1. *If Assumptions* 1 *to* 5 *hold, then the class of all invariant prediction regions based on an adequate statistic is essentially complete with respect the size and average volume among the class of all invariant prediction regions.*

Remark 3. In Assumption 4, the requirement that for all $g \in \mathcal{G}$ and $B \in \mathcal{L}$, $\nu(B)=0$ implies $\nu(Bg)=0$ is satisfied in particular when there exists a right invariant measure on $(\mathcal{G}, \mathcal{L})$. For the existence of such a measure, see Fraser [3].

Remark 4. Assumption 5 is satisfied whenever $\mathcal{P}$ is equivalent to λ.

4. Best invariant prediction region

In this section we shall seek the prediction region which minimizes the average volume among the class of all invariant prediction regions with confidence level $1-\varepsilon$. Such a prediction region is said to be the best invariant. Then Theorem 1 implies that it can be found among those based on the adequate statistic.

Ishii [7] obtained the best invariant prediction region among the class of all invariant prediction regions based on the adequate statistic. But Theorem 1 in [7] can not be obtained without an additional assumption. For the details, see Remark 5 in this section. So we shall consider this problem again.

Now we shall state several assumptions which are slightly different from those in [7], but the method of the proof is almost the same.

Let $\tilde{\mathcal{G}}$ denote a group of transformations on $\mathcal{Y}$.

ASSUMPTION 6. $\mathcal{G}$ is a group of transformations on $\mathcal{T}$ such that

(4.1) $$g(t, y)=(gt, [g; t]y)\ , \qquad g \in \mathcal{G},\ t \in \mathcal{T},\ y \in \mathcal{Y}\ ,$$

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

ASSUMPTION 7. ξ is a relatively invariant measure with respect to $\tilde{\mathcal{G}}$, that is, $\xi(\tilde{g}C)=J(\tilde{g})\xi(C)$, $C \in \mathcal{C}$, $\tilde{g} \in \tilde{\mathcal{G}}$, and $J([g; t])$ does not depend on $t \in \mathcal{T}$.

To simplify the presentation, from now on we shall write $J(g)$ instead of $J([g; t])$.

ASSUMPTION 8. $\mathcal{G}$ is transitive on Θ, that is, for every θ and $\theta' \in \Theta$, there exists a $g \in \mathcal{G}$ such that $g\theta=\theta'$.

In the sequel, a point $\theta_0 \in \Theta$ is fixed and by $g_\theta \in \mathcal{G}$ we shall denote the transformation such that $g_\theta\theta_0=\theta$ for any $\theta \in \Theta$. Then we have the following lemma.

LEMMA 4. *If Assumptions* 3 *and* 6 *to* 8 *hold, then for any invariant prediction region* ϕ *based on* T, *we have*

(4.2) $$\mathrm{E}_\theta\, \phi(T, Y)=\mathrm{E}_{\theta_0}\, \phi(T, Y)$$

and

(4.3) $$\mathrm{E}_\theta \int \phi(T, y)\xi(dy)=J(g_\theta)\, \mathrm{E}_{\theta_0} \int \phi(T, y)\xi(dy)\ .$$

PROOF. Assumption 3 implies that the family of probability distributions of (T, Y) is invariant under $\mathcal{G}$. By (4.1) ϕ is invariant if

(4.4) $$\phi(gt, [g; t]y)=\phi(t, y)\ , \qquad g \in \mathcal{G},\ t \in \mathcal{T},\ y \in \mathcal{Y}\ .$$

Since the distribution of $g_\theta(T, Y)$ under P_{θ_0} is equal to that of (T, Y) under P_θ, from (4.4) we obtain (4.2).

The distribution of $g_\theta T$ under P_{θ_0} is equal to that of T under P_θ, so that

$$\mathrm{E}_\theta \int \phi(T, y)\xi(dy)=\mathrm{E}_{\theta_0} \int \phi(g_\theta T, y)\xi(dy)$$
$$=J(g_\theta)\, \mathrm{E}_{\theta_0} \int \phi(T, y)\xi(dy)$$

where the last equality follows from Assumption 7 and (4.4). Hence we obtain (4.3).

This lemma implies that we have only to consider the size and average volume of invariant prediction regions at $\theta=\theta_0$.

Suppose that $t(x)$ is decomposed so that $t(x)=(t_1(x), t_2(x))$. Let $(\mathcal{T}_1,$

$\mathcal{U}_1$) and $(\mathcal{T}_2, \mathcal{U}_2)$ be the sample spaces of $T_1=t_1(X)$ and $T_2=t_2(X)$, respectively.

ASSUMPTION 9. $\mathcal{G}$ is a group of transformations on $\mathcal{T}_1$ and $\mathcal{T}_2$ and satisfies that for any $g \in \mathcal{G}$

$$gt=(gt_1, gt_2)\ , \qquad t=(t_1, t_2)\ . \tag{4.5}$$

ASSUMPTION 10. $\mathcal{G}$ is transitive on $\mathcal{T}_1$.

In the sequel, a point $t_0 \in \mathcal{T}_1$ is fixed and by $g_{t_1} \in \mathcal{G}$ we denote the transformation such that $t_1=g_{t_1}t_0$ for any $t_1 \in \mathcal{T}_1$.

Setting $g=g_{t_1}^{-1}$ in (4.4), from (4.5) we have that for any invariant prediction region ϕ

$$\phi(t_0, g_{t_1}^{-1}t_2, [g_{t_1}^{-1}; t]y)=\phi(t_1, t_2, y)\ . \tag{4.6}$$

Let w_1 be the mapping from $\mathcal{T}$ to $\mathcal{T}_2$ such that

$$w_1(t)=g_{t_1}^{-1}t_2\ , \qquad t=(t_1,t_2) \tag{4.7}$$

and let w_2 be the mapping from $\mathcal{T}\times\mathcal{Y}$ to $\mathcal{Y}$ such that

$$w_2(t, y)=[g_{t_1}^{-1}; t]y\ .$$

Since $g_{t_1}(t_0, w_1, w_2)=(t_1, t_2, y)$, $[g_{t_1}; t_0, w_1]w_2=y$. Therefore

$$w_2(t, y)=[g_{t_1}; t_0, w_1]^{-1}y\ . \tag{4.8}$$

From (4.6) ϕ is a function of w_1 and w_2. So we set

$$\phi(t, y)=\Phi(w_1, w_2)\ . \tag{4.9}$$

ASSUMPTION 11. There exist σ-finite measures, ξ_1 and ξ_2, on $(\mathcal{T}_1, \mathcal{U}_1)$ and $(\mathcal{T}_2, \mathcal{U}_2)$ such that ξ_2 is relatively invariant with respect to $\mathcal{G}$, that is, $\xi_2(gU_2)=\Delta(g)\xi_2(U_2)$, $g \in \mathcal{G}$, $U_2 \in \mathcal{U}_2$, and the family of probability distributions of (T, Y) is dominated by $\xi_1\times\xi_2\times\xi$.

By $q(t, y)$ we denote the density function of (T, Y) at $\theta=\theta_0$ with respect to $\xi_1\times\xi_2\times\xi$.

ASSUMPTION 12. The mappings, $w_1(t)$, $w_2(t, y)$, $J(g_{t_1})$ and $\Delta(g_{t_1})$, are measurable.

Setting $W_1=w_1(T)$ and $W_2=w_2(T, Y)$, from (4.7) and (4.8) the density function of (T_1, W_1, W_2) is given by

$$q(t_1, g_{t_1}w_1, [g_{t_1}; t_0, w_1]w_2)\Delta(g_{t_1})J(g_{t_1}) \tag{4.10}$$

and the marginal density function of (T_1, W_1) becomes

$$q_1(t_1, g_{t_1}w_1)\Delta(g_{t_1}) , \tag{4.11}$$

where $q_1(t_1, t_2)=\int q(t_1, t_2, y)\xi(dy)$.

Hence, by (4.9) we obtain that

$$\begin{aligned} \mathrm{E}_{\theta_0}\phi(T, Y)&=\mathrm{E}_{\theta_0}\Phi(W_1, W_2) \\ &=\int\int \Phi(w_1, w_2)h(w_1, w_2)\xi_2(dw_1)\xi(dw_2) \end{aligned} \tag{4.12}$$

and

$$\begin{aligned} \mathrm{E}_{\theta_0}\int \phi(T, y)\xi(dy)&=\int\int\int \phi(t_1, t_2, y)q_1(t_1, t_2)\xi_1(dt_1)\xi_2(dt_2)\xi(dy) \\ &=\int\int\int \Phi(w_1, w_2)q_1(t_1, g_{t_1}w_1)\Delta(g_{t_1})J(g_{t_1})\xi_1(dt_1) \\ &\qquad\times \xi_2(dw_1)\xi(dw_2) \\ &=\int\int \Phi(w_1, w_2)f(w_1)\xi_2(dw_1)\xi(dw_2) , \end{aligned} \tag{4.13}$$

where h is the density function of (W_1, W_2) and from (4.10) it becomes

$$h(w_1, w_2)=\int q(t_1, g_{t_1}w_1, [g_{t_1}; t_0, w_1]w_2)\Delta(g_{t_1})J(g_{t_1})\xi_1(dt_1) \tag{4.14}$$

and

$$f(w_1)=\int q_1(t_1, g_{t_1}w_1)\Delta(g_{t_1})J(g_{t_1})\xi_1(dt_1) . \tag{4.15}$$

Using (4.11), we can write

$$f(w_1)=\mathrm{E}_{\theta_0}(J(g_{T_1})|W_1)(w_1)h_1(w_1) , \tag{4.16}$$

where $\mathrm{E}_{\theta_0}(J(g_{T_1})|W_1)$ is the conditional expectation of $J(g_{T_1})$ given W_1 and $h_1(w_1)$ is the marginal density function of W_1.

ASSUMPTION 13. $J(g_{T_1})$ and W_1 are mutually independent.

Then we obtain from (4.13) and (4.16) that

$$\mathrm{E}_{\theta_0}\int \phi(T, y)\xi(dy)=a\int\int \Phi(w_1, w_2)h_1(w_1)\xi_2(dw_1)\xi(dw_2)$$

with $a=\mathrm{E}_{\theta_0}J(g_{T_1})$.

Hence, from (4.12) we have the following result by the method of the proof of Neyman Pearson's Lemma, Theorem 1 and Lemma 4.

THEOREM 2. *If Assumptions* 1 *to* 13 *hold, then the best invariant prediction region with confidence level* $1-\varepsilon$ *is given by*

$$\phi(z)=\begin{cases} 1 & \text{if } h(w_2|w_1)>c\,, \\ r & \text{if } h(w_2|w_1)=c\,, \\ 0 & \text{if } h(w_2|w_1)<c\,, \end{cases}$$

where $h(w_2|w_1)$ *is the conditional density function of* W_2 *given* W_1*, and* r *and* c *are determined by* $\mathrm{E}_{\theta_0}\phi(Z)=1-\varepsilon$.

Remark 5. Ishii considered $f(w_1)$ given by (4.15) as the density function of W_1 (see [7], (3)). But it is evident from (4.11) that it is not the density function of W_1. So from (4.16), to prove Theorem 1 in [7] it is necessary to add Assumption 13. If T_2 is null set, then it is not necessary to treat W_1 and Assumption 13 is trivial.

5. Examples

Now we shall consider some examples in [7] again and show that these examples satisfy Assumption 13, since other assumptions in Theorem 2 are easy to verify. In the sequel, we suppose that ξ_1, ξ_2 and ξ are Lebesgue measures.

5.1. *Multivariate normal distribution*

Let X_i, $i=1,\cdots,n+1$, be independently distributed $(p+q)$-dimensional normal random vectors with unknown mean μ and unknown non-singular covariance matrix Σ. Suppose $n>p+q$ and let $X'_{n+1}=(X^{1\prime}_{n+1}, X^{2\prime}_{n+1})$ where X^1_{n+1} is $p\times 1$. We can observe $X=(X_1,\cdots,X_n,X^1_{n+1})$ but can not observe $Y=X^2_{n+1}$ until a later time. Therefore on the basis of X, we want to construct a prediction region of Y.

By $G(m)$ we denote the group of $m\times m$ lower triangular matrices with positive diagonal elements. Let $\mathcal{G}=\{(b,B);\ b$ is $(p+q)\times 1$ and $B\in G(p+q)\}$ and $g=(b,B)\in\mathcal{G}$ operates on $\mathcal{Z}$ as follows

$$g(x_1,\cdots,x_{n+1})=(b+Bx_1,\cdots,b+Bx_{n+1})\,.$$

Using Lemma 1, it is easy to see that an adequate statistic is given by

$$T=(\bar{X}, S, X^1_{n+1})\,,$$

where

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

and $\bar{X}_1$ is $p\times 1$ and S_{11} is $p\times p$.

Set $S=AA'$ with $A\in G(p+q)$ and partition A as

$$A=\begin{pmatrix} A_{11} & 0 \\ A_{21} & A_{22} \end{pmatrix}, \qquad A_{11} \in G(p),\ A_{22} \in G(q) .$$

Then by putting $T_1=(\bar{X}, S)$ and $T_2=X^1_{n+1}$, it can be shown that

$$(5.1) \qquad \begin{aligned} W_1 &= A_{11}^{-1}(X^1_{n+1}-\bar{X}_1) , \\ W_2 &= A_{22}^{-1}\{X^2_{n+1}-\bar{X}_2-A_{21}A_{11}^{-1}(X^1_{n+1}-\bar{X}_1)\} \end{aligned}$$

and

$$(5.2) \qquad J(g_{T_1})=|A_{22}| .$$

Since $S_{11}=A_{11}A'_{11}$ and $S_{22}-S_{21}S_{11}^{-1}S_{12}=A_{22}A'_{22}$, A_{11} and A_{22} are mutually independent (e.g. see Theorem 6.4.1 in Giri [4], p. 120), so that from (5.1) and (5.2) Assumption 13 is satisfied. Therefore the best invariant prediction region is obtained from the conditional density function of W_2 given W_1, which is calculated by Ishii (see [7], p. 151).

5.2. *Exponential distribution*

Let $X_1<X_2<\cdots<X_n$ be order statistics of size n from the exponential distribution with density function $\sigma^{-1}\exp\{-(x-\mu)/\sigma\}$, $x>\mu$, $\sigma>0$. Here we suppose that $\theta=(\mu, \sigma)$ is unknown.

We shall consider the prediction problem of $Y=X_n$ for the situation where the first r $(1<r<n)$ observations $X=(X_1,\cdots, X_r)$ have been observed.

Let $\mathcal{G}=\{(a, b);\ b>0\}$ and $g=(a, b)\in\mathcal{G}$ operates on $\mathcal{Z}$ as follows

$$g(x_1,\cdots, x_r, y)=(a+bx_1,\cdots, a+bx_r, a+by) .$$

Using Lemma 1, it is easy to see that an adequate statistic is given by

$$T=(X_1, S_r, X_r) ,$$

where $S_r=\sum_{i=2}^{r}(X_i-X_1)+(n-r)(X_r-X_1)$. Set $T_1=(X_r, S_r)$ and $T_2=X_1$. Then it can be shown that

$$(5.3) \qquad W_1=(X_1-X_r)/S_r , \qquad W_2=(Y-X_r)/S_r , \qquad J(g_{T_1})=S_r .$$

Putting $Z_i=(n-i+1)(X_i-X_{i-1})$, $i=1,\cdots, n$, with $X_0=\mu$, Z_i's are mutually independent and have exponential distributions with $\mu=0$ (see Lemma 3 in Epstein and Sobel [2]). Using Z_i's,

$$(5.4) \qquad S_r=\sum_{i=2}^{r} Z_i , \qquad W_1=-\left\{\sum_{i=2}^{r} Z_i/(n-i+1)\right\}\Big/ S_r .$$

It is easy to see that S_r is complete and sufficient for the family of probability distribution of Z_i, $i=2,\cdots,r$, so that from Basu's Theorem (see [1]) and (5.4), it turns out that S_r and W_1 are mutually independent. Hence by (5.3) Assumption 13 is satisfied.

Since $Y-X_r=\sum_{i=r+1}^{n} Z_i/(n-i+1)$, $Y-X_r$, S_r and W_1 are mutually independent, so that from (5.3) W_2 is independent of W_1. Therefore from Theorem 2 the best invariant prediction region is given by

$$\phi(x, y)=\begin{cases} 1, & \text{if } h(w_2)>c, \\ 0, & \text{otherwise}, \end{cases}$$

where h is the density function of W_2 and c is determined such that its size becomes $1-\varepsilon$. The exact formula of h is obtained by Likěs [9].

Remark 6. This problem can be also solved by the result given by Takeuchi ([14], pp. 63–65) and it can be shown that the best invariant prediction region becomes interval (cf. [12]).

Acknowledgment

The author wishes to thank Professor H. Kudō for his helpful suggestions.

KUMAMOTO UNIVERSITY

REFERENCES

[1] Basu, D. (1955). On statistics independent of a complete sufficient statistics, *Sankhyā*, **15**, 377-380.

[2] Epstein, B. and Sobel, M. (1954). Some theorems relevant to life testing from an exponential distribution, *Ann. Math. Statist.*, **25**, 373-381.

[3] Fraser, D. A. S. (1968). *The Structure of Inference*, Wiley, New York.

[4] Giri, N. C. (1977). *Multivariate Statistical Inference*, Academic Press, New York.

[5] Hall, W. J., Wijsman, R. A. and Ghosh, J. K. (1965). The relationship between sufficiency and invariance with application in sequential analysis, *Ann. Math. Statist.*, **36**, 575-614.

[6] Halmos, P. R. and Savage, L. J. (1949). Application of Radon-Nikodym theorem to the theory of sufficient statistics, *Ann. Math. Statist.*, **20**, 225-241.

[7] Ishii, G. (1980). Best invariant prediction region based on an adequate statistic, *Recent Developments in Statistical Inference and the Data Analysis* (ed. K. Matusita), North-Holland.

[8] Lehmann, E. L. (1953). *Testing Statistical Hypotheses*, Wiley, New York.

[9] Likěs, J. (1974). Prediction of *s*th ordered observation for the two-parameter exponential distribution, *Technometrics*, **16**, 241-244.

[10] Skibinsky, M. (1967). Adequate subfields and sufficiency, *Ann. Math. Statist.*, **38**, 155-161.

[11] Sugiura, M. and Morimoto, H. (1969). Factorization theorem for adequate σ-field, *Sūgaku*, **21**, 286-289 (in Japanese).

[12] Takada, Y. (1979). The shortest invariant prediction interval from the largest observation from the exponential distribution, *J. Japan Statist. Soc.*, **9**, 87–91.

[13] Takada, Y. (1981). Invariant prediction rules and an adequate statistic, *Ann. Inst. Statist. Math.*, **33**, A, 91–100.

[14] Takeuchi, K. (1975). *Statistical Prediction Theory*, Baihūkan, Tokyō (in Japanese).

[15] Takeuchi, K. and Akahira, M. (1975). Characterizations of prediction sufficiency (adequacy) in terms of risk function, *Ann. Statist.*, **3**, 1018–1024.