Ann. Inst. Statist. Math.
33 (1981), Part A, 91-100

INVARIANT PREDICTION RULES AND AN ADEQUATE STATISTIC

YOSHIKAZU TAKADA

(Received Aug. 1, 1980; revised Jan. 5, 1981)

1. Introduction

In many statistical decision problems it is reasonable to confine attention to rules that are invariant with respect to a certain group of transformations. If a given decision problem admits a sufficient statistic, it is well known that the class of invariant rules based on the sufficient statistic is essentially complete in the class of all invariant rules under several assumptions; for example see Theorems 5.4.4 and 5.4.5 in Nabeya [9], p. 192. This result may be used to show that if there exists a minimax invariant rule among invariant rules based on sufficient statistic, it is minimax among all invariant rules.

In this paper we consider statistical prediction problems which are invariant with respect to a certain group of transformations. In prediction problems, an adequate statistic plays important roles as a sufficient statistic does in ordinary statistical decision problems. For the details, see papers by Skibinsky [10] and Takeuchi and Akahira [14]. The purpose of this paper is to show that the class of invariant prediction rules based on the adequate statistic is essentially complete in the class of all invariant prediction rules under several assumptions.

In Section 2, some results on invariant prediction rules and an adequate statistic are stated. Our aimed results are stated in Section 3. Applying these results in Section 4, we obtain minimax invariant predictors in some examples.

2. Invariant prediction rules and an adequate statistic

Let X be an observable random variable and Y a future random variable. Let $(\mathcal{X}, \mathcal{B})$ and $(\mathcal{Y}, C)$ be sample spaces of X and Y, respectively. Let $(\mathcal{Z}, \mathcal{A})=(\mathcal{X}\times\mathcal{Y}, \mathcal{B}\times C)$ and $\mathcal{P}=\{P_\theta:\ \theta\in\Theta\}$ be a family of probability measures on $(\mathcal{Z}, \mathcal{A})$ and Θ a parameter space. Let $\mathcal{G}$ be a group of measurable, one-to-one transformations from $\mathcal{Z}$ onto itself.

ASSUMPTION 1. Each $g\in\mathcal{G}$ induces a one-to-one transformation $\bar{g}$ from Θ onto itself defined by $P_{\bar{g}\theta}(gA)=P_\theta(A)$, $A\in\mathcal{A}$, $\theta\in\Theta$.

Let $(\mathcal{X}, \mathcal{B})=(\mathcal{X}_1\times\mathcal{X}_2, \mathcal{B}_1\times\mathcal{B}_2)$ and $\mathcal{G}_1$, $\mathcal{G}_2$ and $\mathcal{G}_3$ be groups of measurable, one-to-one transformations from $\mathcal{X}_1$, $\mathcal{X}_2$ and $\mathcal{Y}$ onto themselves, respectively.

ASSUMPTION 2. There exist mappings from $\mathcal{G}$ to $\mathcal{G}_1$ and $\mathcal{G}_2$ and a mapping from $\mathcal{G}\times\mathcal{X}_2$ to $\mathcal{G}_3$, whose image at $(g, x_2)\in\mathcal{G}\times\mathcal{X}_2$ is denoted by g_{x_2}, such that for each $g\in\mathcal{G}$, $x_1\in\mathcal{X}_1$, $x_2\in\mathcal{X}_2$ and $y\in\mathcal{Y}$,

$$g(x_1, x_2, y)=(g_1x_1, g_2x_2, g_{x_2}y) \tag{2.1}$$

where $g_1\in\mathcal{G}_1$ and $g_2\in\mathcal{G}_2$ are images of g to $\mathcal{G}_1$ and $\mathcal{G}_2$, respectively.

Remark 1. In cases where the transformation on $\mathcal{Y}$ does not depend on x, there is no need to consider $\mathcal{X}_2$.

Let $(\mathcal{D}, \mathcal{F})$ be a decision space and $\tilde{\mathcal{G}}$ a group of measurable one-to-one transformations from $\mathcal{D}$ onto itself. Let $L(\theta, y, d)$ be a loss function from $\Theta\times\mathcal{Y}\times\mathcal{D}$ to $[0, \infty)$ which is $C\times\mathcal{F}$ measurable for each $\theta\in\Theta$.

ASSUMPTION 3. There exists a mapping from $\mathcal{G}\times\mathcal{X}_2$ to $\tilde{\mathcal{G}}$, whose image at $(g, x_2)\in\mathcal{G}\times\mathcal{X}_2$ is denoted by $\tilde{g}_{x_2}$, such that for each $g\in\mathcal{G}$, $\theta\in\Theta$, $x_2\in\mathcal{X}_2$, $y\in\mathcal{Y}$ and $d\in\mathcal{D}$,

$$L(\bar{g}\theta, g_{x_2}y, \tilde{g}_{x_2}d)=L(\theta, y, d)\,. \tag{2.2}$$

A prediction rule δ will be defined as follows; for each $x\in\mathcal{X}$, $\delta(\cdot|x)$ is a probability measure on $(\mathcal{D}, \mathcal{F})$ and for each $D\in\mathcal{F}$, $\delta(D|\cdot)$ is $\mathcal{B}$ measurable. The risk function of δ is given by

$$R(\theta, \delta)=\int_{\mathcal{F}}\left\{\int_{\mathcal{D}} L(\theta, y, s)\delta(ds|x)\right\}P_\theta(dz)\,, \qquad \theta\in\Theta\,. \tag{2.3}$$

DEFINITION 1. A prediction rule δ is said to be invariant under $\mathcal{G}$ if for all $x\in\mathcal{X}$, $g\in\mathcal{G}$ and $D\in\mathcal{F}$,

$$\delta(\tilde{g}_{x_2}D|g_1x_1, g_2x_2)=\delta(D|x)\,. \tag{2.4}$$

A very important property of an invariant prediction rule is that its risk function is constant on orbits. More precisely we have the following lemma, the proof of which is the same as that of Theorem 1 of Ferguson [4], p. 150, so omitted.

LEMMA 1. *If Assumptions* 1 *through* 3 *hold, then for any invariant rule* δ

$$R(\theta, \delta)=R(\bar{g}\theta, \delta)\,, \quad \theta\in\Theta\,, \quad g\in\mathcal{G}\,.$$

Let t be a measurable mapping from $(\mathcal{X}, \mathcal{B})$ onto $(\mathcal{T}, \mathcal{U})$ and let $T=t(x)$.

DEFINITION 2. A statistic T is said to be adequate for X w.r.t. Y if T is sufficient for X and given T, X and Y are conditionally independent.

Sugiura and Morimoto [12] provided a simple criterion with which we can generally determine an adequate statistic.

LEMMA 2. *If $\mathcal{P}$ is dominated by $\lambda=\lambda_1\times\lambda_2$ where λ_1 and λ_2 are probability measures on $\mathcal{X}$ and $\mathcal{Y}$, respectively, then T is adequate for X w.r.t. Y if and only if*

$$\frac{dP_\theta}{d\lambda}=h(x)f_\theta(t(x), y)$$

where $h(x)$ is $\mathcal{B}$ measurable and $f_\theta(t, y)$ is $\mathcal{U}\times C$ measurable.

3. Main result

In this section we shall show that the class of invariant prediction rules based on an adequate statistic, that is, the class of invariant rules δ for which given any $D\in\mathcal{F}$, $\delta(D|\cdot)$ is a function of the adequate statistic, is essentially complete in the class of all invariant rules.

Let t_1 be a measurable mapping from $(\mathcal{X}_1, \mathcal{B}_1)$ onto $(\mathcal{T}_1, \mathcal{U}_1)$ and $t(x)=(t_1(x_1), x_2)$ for $x=(x_1, x_2)$. Let $X=(X_1, X_2)$.

ASSUMPTION 4. X_1 and X_2 are independent, $T_1=t_1(X_1)$ is sufficient for X_1 and $T=t(X)$ is adequate for X w.r.t. Y.

Remark 2. In cases where $\mathcal{X}_2$ is not needed (see Remark 1), this Assumption means that T_1 is adequate.

ASSUMPTION 5. There exists a real valued function Q on $\mathcal{B}_1\times\mathcal{X}_1$ such that

(i) for any $x_1\in\mathcal{X}_1$, $Q(\cdot|x_1)$ is a probability measure on $(\mathcal{X}_1, \mathcal{B}_1)$;

(ii) for any $B\in\mathcal{B}_1$, $Q(B|\cdot)$ is a version of conditional probability of B given $\mathcal{B}_{t_1}=\{t_1^{-1}(U)\,;\ U\in\mathcal{U}_1\}$;

(iii) for any $x_1\in\mathcal{X}_1$, $B\in\mathcal{B}_1$ and $g_1\in\mathcal{G}_1$,

$$Q(g_1B|g_1x_1)=Q(B|x_1)\,.$$

THEOREM 1. *If Assumptions* 1 *through* 5 *hold, then for any invariant prediction rule δ, there exists an invariant prediction rule δ_0 based on T such that*

$$R(\theta, \delta)=R(\theta, \delta_0)\,, \qquad \theta\in\Theta\,. \tag{3.1}$$

PROOF. Define for $D\in\mathcal{F}$ and $x\in\mathcal{X}$, (3.2) $$\delta_0(D|x)=\int_{\mathcal{X}_1}\delta(D|s,\, x_2)Q(ds|x_1)\,.$$

Then from Theorem 1 of Takeuchi and Akahira [14], we have (3.1). Therefore the proof is complete if δ_0 is an invariant rule. From Assumption 5 for any $g\in\mathcal{G}$, $x\in\mathcal{X}$ and $D\in\mathcal{F}$,

$$\begin{aligned}\delta_0(\tilde{g}_{x_2}D|g_1x_1,\, g_2x_2)&=\int_{\mathcal{X}_1}\delta(\tilde{g}_{x_2}D|s,\, g_2x_2)Q(ds|g_1x_1)\\&=\int_{\mathcal{X}_1}\delta(\tilde{g}_{x_2}D|g_1s,\, g_2x_2)Q(g_1ds|g_1x_1)\\&=\int_{\mathcal{X}_1}\delta(D|s,\, x_2)Q(ds|x_1)\\&=\delta_0(D|x)\,,\end{aligned}$$

which completes the proof.

Assumption 5 is not easy to verify. The following lemma which is given by Hall, et al. [6] (see Theorem 7.1, p. 608) may be used to verify Assumption 5.

LEMMA 3. *$\mathcal{X}_1$ is a Borel subset of n-dimensional Euclidean space and $\mathcal{B}_1$ the Borel σ-field of $\mathcal{X}_1$. Let f_θ be the density function of X_1 with respect to Lebesgue measure such that*

$$f_\theta(x_1)=h(x_1)g_\theta(t_1(x_1))\,,\quad x_1\in\mathcal{X}_1\,,\quad \theta\in\Theta$$

where t_1 is a measurable function from $\mathcal{X}_1$ into k-dimensional Euclidean space ($k<n$) with range $\mathcal{T}_1$, h and g_θ are positive real-valued measurable functions on $\mathcal{X}_1$ and $\mathcal{T}_1$, respectively. Let $\mathcal{B}_{t_1}=\{t_1^{-1}(U)\,;\, U\in\mathcal{U}_1\}$ where $\mathcal{U}_1$ is the Borel σ-field of $\mathcal{T}_1$. Suppose that there is an invariant open set $B\in\mathcal{B}_{t_1}$ of $\mathcal{P}$-measure 1 such that on B

(i) each $g_1\in\mathcal{G}_1$ is continuously differentiable and the Jacobian depends only on $t_1(x_1)$;

(ii) for each $g_1\in\mathcal{G}_1$, $t_1(x_1)=t_1(x_1')$ implies $t_1(g_1x_1)=t_1(g_1x_1')$;

(iii) $t_1(x_1)$ is continuously differentiable and the matrix $\|[\partial t_{1j}(x_1)/\partial x_{1i}]: j=1,\cdots,k,\ i=1,\cdots,n\|$ is of rank k where $x_1=(x_{11},\cdots,x_{1n})$ and $t_1(x_1)=(t_{11}(x_1),\cdots,t_{1k}(x_1))$;

(iv) for each $g_1\in\mathcal{G}_1$, $h(g_1x_1)/h(x_1)$ depends only on $t_1(x_1)$. Then Assumption 5 holds.

Now we consider the problem of the essential completeness of the class of nonrandomized invariant prediction rules.

ASSUMPTION 6. $\mathcal{D}$ is a convex Borel subset of p-dimensional Euclidean space, $\mathcal{F}$ is the Borel σ-field of $\mathcal{D}$, $L(\theta, y, d)$ is a convex function of $d\in\mathcal{D}$ for all $\theta\in\Theta$ and $y\in\mathcal{Y}$ and $L(\theta, y, d)\to\infty$ as $\|d\|\to\infty$ where $\|d\|^2=d'd$.

ASSUMPTION 7. $\tilde{\mathcal{G}}$ contains only linear transformations, that is, transformations of the form $\tilde{g}d=Bd+c$ where B is a $p\times p$ nonsingular matrix and c is a p-dimensional vector.

ASSUMPTION 8. There exists a nonrandomized invariant prediction rule based on T.

THEOREM 2. *If Assumptions* 1 *through* 8 *hold, then for any invariant prediction rule* δ *there exists a nonrandomized invariant prediction rule* ϕ *based on* T *such that*

$$R(\theta, \delta)\geqq R(\theta, \phi)\,, \qquad \theta \in \Theta\,.$$

PROOF. Define ϕ by

$$\phi(x)=\int s\delta_0(ds|x) \text{ if } \int \|s\|\delta_0(ds|x)<\infty\,, \tag{3.3}$$
$$=\psi(x)\,, \qquad \text{otherwise}\,,$$

where δ_0 is given by (3.2), ψ is a nonrandomized invariant prediction rule based on T which exists by Assumption 8 and

$$\int s\delta_0(ds|x)=\left(\int s_1\delta_0(ds|x), \cdots, \int s_p\delta_0(ds|x)\right).$$

If $R(\theta, \delta_0)<\infty$, then from (2.3)

$$\int L(\theta, Y, s)\delta_0(ds|X)<\infty \text{ a.e. } [P_\theta]\,.$$

Therefore from Theorem 1 and Remark in Ferguson [4], p. 78 and Assumption 6, we have

$$\int \|s\|\delta_0(ds|X)<\infty \text{ a.e. } [P_\theta]\,,$$

hence by Jensen's inequality

$$L(\theta, Y, \phi(X))\leqq\int L(\theta, Y, s)\delta_0(ds|X) \text{ a.e. } [P_\theta]\,,$$

which implies that

$$R(\theta, \phi)\leqq R(\theta, \delta_0)\,. \tag{3.4}$$

If $R(\theta, \delta_0)=\infty$, it is clear that (3.4) holds. Hence from (3.1) the proof is complete if ϕ is invariant, that is, $\phi(g_1x_1, g_2x_2)=\tilde{g}_{x_2}\phi(x)$, for all $g \in \mathcal{G}$ and $x=(x_1, x_2) \in \mathcal{X}$. From Assumption 7 and the invariance of δ_0 we have

$$\int s\delta_0(ds|g_1x_1, g_2x_2) = \int \tilde{g}_{x_2} s\delta_0(\tilde{g}_{x_2} ds|g_1x_1, g_2x_2) = \tilde{g}_{x_2} \int s\delta_0(ds|x)\,, \tag{3.5}$$

which implies that $\int \|s\|\delta_0(ds|g_1x_1, g_2x_2) < \infty$ if and only if $\int \|s\|\delta_0(ds|x) < \infty$. Then from (3.3) and (3.5) it is easy to see that ϕ is invariant.

Remark 3. From Theorem 2 if a nonrandomized invariant prediction rule based on T is minimax among all nonrandomized invariant prediction rules based on T, then it is minimax among all invariant prediction rules.

4. Examples

Now we apply the previous result to find a minimax invariant predictor. In this section the space $\mathcal{D}$ is equal to $\mathcal{Y}$.

4.1. *Multivariate normal distribution*

Let X_i, $i=1, \cdots, n+1$, be independently normally distributed $(p+q)$-dimensional random vectors with unknown mean μ and unknown non-singular covariance matrix Σ. Let $n > p+q$.

The following partitions are used in the sequel:

$$X_i = \begin{pmatrix} X_i^1 \\ X_i^2 \end{pmatrix}, \quad i=1, \cdots, n+1\,, \quad \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix},$$

where X_i^1 and μ_1 are $p\times 1$ and Σ_{11} is $p\times p$.

We can observe $X_1, \cdots, X_n, X_{n+1}^1$ but can not observe X_{n+1}^2. The purpose here is to predict X_{n+1}^2. This problem was considered in Ishii [8], p. 482. Let $X=(X_1, \cdots, X_n, X_{n+1}^1)$ and $Y=X_{n+1}^2$. Ishii proposed as a predictor of Y,

$$r(X) = \bar{X}_2 + S_{21}S_{11}^{-1}(X_{n+1}^1 - \bar{X}_1)\,, \tag{4.1}$$

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 $p\times p$. But the justification of (4.1) has not appeared in literatures as far as the author knows. We shall show that (4.1) is a minimax invariant predictor under the following transformations. Let $G(m)$ be 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 $B \in G(p+q)\}$, which operates on x_i $(i=1, \cdots, n+1)$ as follows; for $g=(b, B)$

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

Let $\mathcal{X}_1$ and $\mathcal{X}_2$ be sample spaces of $(X_1,\cdots, X_n)$ and X^1_{n+1}, respectively. Let $\mathcal{G}_1=\mathcal{G}$, $\mathcal{G}_2=\{(b_1, B_1) ; b_1$ is $p\times 1$ and $B_1 \in G(p)\}$ and $\mathcal{G}_3=\tilde{\mathcal{G}}=\{(b_2, B_2) ;$ b_2 is $q\times 1$ and $B_2 \in G(q)\}$. We partition $g=(b, B) \in \mathcal{G}$ as

$$b=\begin{pmatrix} b_1 \\ b_2 \end{pmatrix}, \qquad B=\begin{pmatrix} B_{11} & 0 \\ B_{21} & B_{22} \end{pmatrix}$$

where b_1 is $p\times 1$ and B_{11} is $p\times p$. Then the mapping from $\mathcal{G}$ to $\mathcal{G}_2$ by $(b, B) \to (b_1, B_{11})$ and the mappings from $\mathcal{G}\times\mathcal{X}_2$ to $\mathcal{G}_3$ and $\tilde{\mathcal{G}}$ by $((b, B), x^1_{n+1}) \to (b_2+B_{21}x^1_{n+1}, B_{22})$ are defined. From (4.2) these mappings satisfy (2.1).

Let $\theta=(\mu, \Sigma)$ and we take the loss function defined by

(4.3) $$L(\theta, y, d)=(y-d)'(\Sigma_{22}-\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})^{-1}(y-d) .$$

The induced transformation on the parameter space Θ corresponding to $g \in \mathcal{G}$ is defined by

(4.4) $$(\mu, \Sigma) \to (B\mu+b, B\Sigma B') , \qquad g=(b, B) .$$

Then it is easy to see that (4.3) satisfies (2.2). Let $T_1=(\bar{X}, S)$ and $T=(T_1, X^1_{n+1})$. Then from Lemma 2 T is an adequate statistic and Assumption 4 holds. All parts of conditions of Lemma 3 can be verified to hold (see [6], p. 611). Therefore Assumption 5 holds. It follows easily that the other Assumptions hold. From Remark 3 we can confine our attention to nonrandomized invariant prediction rules based on T. The transformation $g \in \mathcal{G}$ induces on the space of the adequate statistic T the transformation

$$(\bar{x}, S, x^1_{n+1}) \to (B\bar{x}+b, BSB', B_{11}x^1_{n+1}+b_1)$$

where $g=(b, B)$. Therefore the invariance of a predictor ϕ based on T means that for all (b, B) and $(\bar{x}, S, x^1_{n+1})$

(4.5) $$\phi(B\bar{x}+b, BSB', B_{11}x^1_{n+1}+b_1)=B_{22}\phi(\bar{x}, S, x^1_{n+1})+B_{21}x^1_{n+1}+b_2 .$$

Putting $b=-B\bar{x}$ in (4.5), we get

(4.6) $$\phi(0, BSB', B_{11}(x^1_{n+1}-\bar{x}_1))=B_{22}\phi(\bar{x}, S, x^1_{n+1})+B_{21}(x^1_{n+1}-\bar{x}_1)-B_{22}\bar{x}_2 .$$

Let $S=AA'$ where $A \in G(p+q)$. Put

$$A=\begin{pmatrix} A_{11} & 0 \\ A_{21} & A_{22} \end{pmatrix}$$

where A_{11} is $p\times p$. Then $S_{11}=A_{11}A'_{11}$, $S_{22}-S_{21}S_{11}^{-1}S_{12}=A_{22}A'_{22}$ and $S_{21}S_{11}^{-1}=A_{21}A_{11}^{-1}$. Put $B=A^{-1}$ in (4.6). Then we get

$$\phi(0, I, A_{11}^{-1}(x_{n+1}^1-\bar{x}_1))=A_{22}^{-1}[\phi(\bar{x}, S, x_{n+1}^1)-\bar{x}_2]-A_{22}^{-1}S_{21}S_{11}^{-1}(x_{n+1}^1-\bar{x}_1)$$

where I is the $p\times p$ identity matrix. Therefore we get

$$\phi(\bar{x}, S, x_{n+1}^1)=r(x)+A_{22}h(A_{11}^{-1}(x_{n+1}^1-\bar{x}_1)) \tag{4.7}$$

where $r(x)$ is given by (4.1) and $h(\cdot)$ is some measurable function from $(R^p, \mathcal{B}^p)$ to $(R^q, \mathcal{B}^q)$.

From Lemma 1, without loss of generality we assume $(\mu, \Sigma)=(0, I)$. Then

$$\begin{aligned} R((0, I), \phi)=&\mathrm{E}\,\|Y-r(X)\|^2+2\,\mathrm{E}\,(Y-r(X))'A_{22}h(A_{11}^{-1}(X_{n+1}^1-\bar{X}_1)) \\ &+\mathrm{E}\,\|A_{22}h(A_{11}^{-1}(X_{n+1}^1-\bar{X}_1))\|^2\,. \end{aligned} \tag{4.8}$$

It is well known that $(\bar{X}, X_{n+1}^1, Y)$, $S_{22}-S_{21}S_{11}^{-1}S_{12}$ and (S_{21}, S_{11}) are mutually independent and under $(\mu, \Sigma)=(0, I)$ conditional means of $Y-\bar{X}_2$ given $(\bar{X}_1, X_{n+1}^1)$ and of $S_{21}S_{11}^{-1}$ given S_{11} are zero, respectively (e.g. see Theorem 6.4.1 in Giri [5], p. 120). Therefore we get

$$\mathrm{E}\,(Y-r(X))'A_{22}h(A_{11}^{-1}(X_{n+1}^1-\bar{X}_1))=0\,.$$

Thus from (4.8) it turns out that (4.1) is a minimax invariant predictor.

Remark 4. When $q=1$, it is known that (4.1) is inadmissible if $p\geqq 3$. This fact is first proved by Stein [11]. Superior predictors can be constructed using estimators given by Baranchik [1] and Takada [13].

4.2. *Exponential distribution*

Let $X_1<X_2<\cdots<X_n$ be order statistics of size n from the exponential distribution with the density $\theta^{-1}\exp(-x/\theta)$, $x>0$, $\theta>0$.

We shall consider the prediction problem of X_n for the situation where the first r observations $X_1<X_2<\cdots<X_r$, $1\leqq r<n$, have been observed. Let $\mathcal{G}$ be the group of transformations $x_i\to cx_i$ $(i=1, \cdots, r, n, c>0)$. We are now concerned with the minimax invariant predictor under the loss function $(y-d)^2/\theta^2$. In this case the space $\mathcal{X}_2$ is not needed. So we denote x_1 in Section 3 by x.

Let $X=(X_1, \cdots, X_r)$ and $Y=X_n$. Then the joint probability density of X and Y is given by

$$\frac{n!}{(n-r-1)!}\theta^{-(r+1)}\exp\left[-\left(\sum_{i=1}^r x_i+y\right)\Big/\theta\right][\exp(-x_r/\theta)-\exp(-y/\theta)]^{n-r-1}$$

for $0<x_1<\cdots<x_r<y$ and zero, otherwise. It follows from Lemma 2 that $t(X)=\left(\sum_{i=1}^r X_i, X_r\right)$ is an adequate statistic. Let $\mathcal{X}=\{x\,;\,0<x_1<\cdots<x_r\}$. Then the probability density of X is given by

$$\frac{n!}{(n-r)!}\theta^{-r}\exp\left[-\left(\sum_{i=1}^{r}x_i+(n-r)x_r\right)\Big/\theta\right]$$

on $\mathcal{X}$. According to Lemma 3, let $h(x)=1$ and

$$g_\theta(t(x))=\frac{n!}{(n-r)!}\theta^{-r}\exp\left[-(t_1(x)+t_2(x))/\theta\right]$$

where $t_1(x)=\sum_{i=1}^{r}x_i$ and $t_2(x)=(n-r)x_r$. Then it follows easily that the matrix $D(x)=\|[\partial t_i(x)/\partial x_j]\,;\ i=1, 2,\ j=1, \cdots, r\|$ is of rank 2 and all the other assumptions in Lemma 3 hold. It is easily checked that assumptions in Theorem 2 hold. Hence we confine our attention to non-randomized invariant predictors based on $T=t(X)$. Let $S=\sum_{i=1}^{r}X_i+(n-r)X_r$ and $T'=(S, X_r)$. Since T and T' are equivalent, we consider invariant predictors based on T'. It follows easily that they are given by

$$\delta(X)=h(X_r/S)S\,, \tag{4.9}$$

where $h(\cdot)$ is some measurable function. Then the risk function of (4.9) becomes

$$R(\theta, \delta)=\mathrm{E}_\theta\left[(Y-\delta(X))^2/\theta^2\right]. \tag{4.10}$$

Using Lemma 3 in Epstein and Sobel [3], p. 375, we have that

$$\mathrm{E}_\theta\,(Y\,|\,X_r=x_r)=x_r+\theta\sum_{i=r+1}^{n}(n-i+1)^{-1}\,.$$

We denote it by $x_r+a\theta$ with $a=\sum_{i=r+1}^{n}(n-i+1)^{-1}$. Then (4.10) is equal to

$$\mathrm{E}_\theta\,(Y-X_r-a\theta)^2/\theta^2+\mathrm{E}_\theta\,(X_r-h(X_r/S)S+a\theta)^2/\theta^2\,. \tag{4.11}$$

Let $g(t)=t-h(t)$. Then

$$\mathrm{E}_\theta\,(X_r-h(X_r/S)S+a\theta)^2/\theta^2=\mathrm{E}_\theta\,(g(X_r/S)S+a\theta)^2/\theta^2\,. \tag{4.12}$$

Since S is complete and sufficient (see Theorem 2 in [3], p. 377), from Theorem of Basu [2] S and X_r/S are independent. Therefore (4.12) becomes

$$r(r+1)\,\mathrm{E}_\theta\,g^2(X_r/S)+2ra\,\mathrm{E}_\theta\,g(X_r/S)+a^2\,.$$

This is minimized by $g(\cdot)=-a/(r+1)$. Therefore from (4.11) a minimax invariant predictor is given by

$$\delta(X)=X_r+\left(\sum_{i=r+1}^{n}(n-i+1)^{-1}\right)S\Big/(r+1)\,.$$

Remark 5. This problem can be solved by direct application of Corollary of Hora and Buehler [7], p. 798.

Acknowledgement

The author wishes to thank the referee for his helpful comments in revising paper. Especially, the proof of Theorem 2 is owing to his suggestion.

UNIVERSITY OF TSUKUBA*

REFERENCES

[1] Baranchik, A. J. (1973). Inadmissibility of maximum likelihood estimators in some multiple regression problems with three or more independent variables, *Ann. Statist.*, **1**, 312-321.
[2] Basu, D. (1955). On statistics independent of a complete sufficient statistics, *Sankhyā*, **15**, 377-380.
[3] Epstein, B. and Sobel, M. (1954). Some theorems relevant to life testing from an exponential distribution, *Ann. Math. Statist.*, **25**, 373-381.
[4] Ferguson, T. S. (1967). *Mathematical Statistics: A Decision Theoretic Approach*, Academic Press, New York.
[5] Giri, N. C. (1977). *Multivariate Statistical Inference*, Academic Press, New York.
[6] 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.
[7] Hora, R. B. and Buehler, R. J. (1967). Fiducial theory and invariant prediction, *Ann. Math. Statist.*, **38**, 795-801.
[8] Ishii, G. (1969). Optimality of unbiased predictors, *Ann. Inst. Statist. Math.*, **21**, 471-488.
[9] Nabeya, S. (1978). *Mathematical Statistics* (in Japanese), Kyōritu Press, Tokyō.
[10] Skibinsky, M. (1967). Adequate subfields and sufficiency, *Ann. Math. Statist.*, **38**, 155-161.
[11] Stein, C. (1960). Multiple regression, *Contribution to Probability and Statistics*, Stanford Univ. Press.
[12] Sugiura, M. and Morimoto, H. (1969). Factorization theorem for adequate σ-field (in Japanese), *Sūgaku*, **21**, 286-289.
[13] Takada, Y. (1979). A family of minimax estimators in some multiple regression problems, *Ann. Statist.*, **7**, 1144-1147.
[14] Takeuchi, K. and Akahira, M. (1975). Characterizations of prediction sufficiency (Adequacy) in terms of risk functions, *Ann. Statist.*, **3**, 1018-1024.

* Now at the Kumamoto University.