CONSTRUCTION METHODS OF THE BEST INVARIANT PREDICTOR

1982

YOSHIKAZU TAKADA

Contents

1. Introduction

We consider a statistical prediction problem which is invariant under a certain group of transformations and present three different methods to construct the best invariant predictor; the first method uses the invariant measure on the group, the second uses an adequate statistic and the third goes via the best unbiased predictor.

In Section 2, we define the statistical prediction problem and state some results which are used in subsequent sections.

In Section 3, we consider the problem treated by Hora and Buehler [9], which gives a representation of the best invariant predictor by using the right Haar measure on the group. We extend the assumptions used by them and discuss conditions required to satisfy the assumptions. Under the conditions, we present an alternative expression of the best invariant predictor which is more suitable for applications.

In Section 4, we express the best invariant predictor by using an adequate statistic. Such a statistic is known to play the similar important role in prediction problems as a sufficient statistic does in ordinary statistical decision problem (see e.g. Skibinsky [18] or Takeuchi and Akahira [27]). We show that the class of invariant predictors based on the adequate statistic is essentially complete in the class of all invariant predictors under some mild assumptions. This result enables us to obtain the best invariant predictor based on the adequate statistic.

In Section 5, we show that the best invariant predictor can be expressed by a linear combination of the best unbiased predictor and the uniformly minimum variance (U.M.V.) unbiased estimator under several assumptions. This result may be useful to obtain the best invariant predictor, provided that the best unbiased predictor and U.M.V. unbiased estimator are easily found.

Throughout this thesis, same examples are considered and the best invariant predictor for each case is obtained by each of three methods except for the combination of the second example and the third method.

The contents of this thesis are extentions of those in [22], [23] and [24].

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 $(\mathcal{Y}, \mathcal{C})$ be sample spaces of X and Y, respectively. Let $(\mathcal{Z}, \mathcal{A}) = (\mathfrak{X} \times \mathcal{Y}, \mathfrak{B} \times \mathcal{C})$ and $\mathcal{P} = \{P_\theta;\ \theta \varepsilon \Theta\}$ be a family of probability measures on $(\mathcal{Z}, \mathcal{A})$ such that $Z=(X,Y)$ is distributed according to P_θ, $\theta \varepsilon \Theta$, and Θ a parameter space. Let $\mathcal{G}$ be a group of one-to-one transformations acting on the spaces $\mathfrak{X}$, $\mathcal{Z}$ and Θ, mapping each onto itself, and let $\tilde{\mathcal{G}}$ be a group of transformations on $\mathcal{Y}$.

Assumption 1. $\mathcal{P}$ is invariant under $\mathcal{G}$, that is,

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

and $\mathcal{G}$ satisfies that

(2.1) $$g(x,y) = (gx, [g;x]y), \quad g \varepsilon \mathcal{G},\ x \varepsilon \mathfrak{X},\ y \varepsilon \mathcal{Y},$$

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

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

Assumption 2. L is invariant under $\mathcal{G}$, that is,

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

for all d, x, y, θ.

A randomized predictor δ will be defined as follows: for each $x \varepsilon \mathfrak{X}$, $\delta(.|x)$ is a probability measure on $(\mathcal{Y}, \mathcal{C})$

and for each $C\varepsilon\mathcal{C}$, $\delta(C|.)$ is $\mathcal{B}$-measurable. The risk function of δ is given by

$$R(\theta,\delta) = E_{\theta}\{\int L(s,Y,\theta)\delta(ds|X)\}, \quad \theta\varepsilon\Theta. \tag{2.3}$$

Definition 1. A predictor δ is said to be invariant under $\mathcal{G}$ if for any $x\varepsilon\mathcal{X}$, $g\varepsilon\mathcal{G}$ and $C\varepsilon\mathcal{C}$,

$$\delta([g;x]C|gx) = \delta(C|x). \tag{2.4}$$

On the other hand, a non-randomized predictor δ is invariant under $\mathcal{G}$ if for any $x\varepsilon\mathcal{X}$ and $g\varepsilon\mathcal{G}$,

$$\delta(gx) = [g;x]\delta(x). \tag{2.5}$$

A very important property of an invariant predictor is that its risk function is constant on each orbit. More precisely, we have the following lemma, which can be proved similarly as Theorem 1 of Ferguson [5], p.190.

Lemma 1.([22]) *If Assumptions* 1 *and* 2 *hold, then for any invariant randomized predictor* δ,

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

Definition 2. An invariant predictor is said to be best if it minimizes (2.3) among all invariant randomized predictors for each $\theta\varepsilon\Theta$.

Assumption 3. Θ is isomorphic to $\mathcal{G}$.

Let $\theta_0\varepsilon\Theta$ be the point corresponding to the identity element e of $\mathcal{G}$. The isomorphism is established by $\theta=g\theta_0$ if $\theta\varepsilon\Theta$ corresponds to $g\varepsilon\mathcal{G}$. It will often be notationally convenient to index $\mathcal{G}$ by the corresponding point of Θ. Thus in the expression "$\theta^{-1}s$", θ^{-1} stands for the point g for which $g\theta=\theta_0$, where θ and θ_0 just written have their ordinary meaning. Which meaning is to be attached to a symbol will be clear from the context.

Then from Lemma 1, the risk function of an invariant predictor is constant on Θ, that is,

(2.6) $$R(\theta,\delta) = R(\theta_0,\delta), \quad \theta\varepsilon\Theta.$$

Suppose $\mathcal{H}$ is a group of transformations acting on some space $\mathcal{D}$ and let h_0 be the identity element of $\mathcal{H}$.

Definition 3. $\mathcal{H}$ is said to act freely on $\mathcal{D}$ if $h\neq h_0$ implies $hd\neq d$ for any $d\varepsilon\mathcal{D}$ and $h\varepsilon\mathcal{H}$.

Assumption 4. $\tilde{\mathcal{G}}$ acts freely on $\mathcal{Y}$.

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

Lemma 2. ([24]) *If Assumption 4 holds, then for* g, $g'\varepsilon\mathcal{G}$ *and* $x\varepsilon\mathcal{X}$,

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

(2.8) $$[g;x]^{-1} = [g^{-1};gx].$$

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 (2.7). Set $g'=g^{-1}$ in (2.7). Then by using the fact that $[e;x]=\tilde{e}$ where

$\tilde{e}$ denotes the identity element of $\tilde{\mathcal{G}}$, (2.8) is obtained.

In the subsequent sections we shall present three methods to construct the best invariant predictor under Assumptions 1 to 4.

3. Construction method by Haar measure

In this section, we shall extend the assumptions used by Hora and Buehler [9] so as to be able to treat such a structure as (2.1) and express the best invariant predictor by using Haar measure on the group. We also discuss a set of sufficient conditions for the assumptions and obtain an alternative expression of the best invariant predictor.

3.1 Construction of the best invariant predictor

Assumption A1. $\mathcal{G}$ is a locally compact topological group with a σ-field $\mathcal{L}$.

Let μ and ν, respectively, denote the left and right Haar measures on $(\mathcal{G}, \mathcal{L})$ and Δ denote the modular function, which is a continuous homomorphism of $\mathcal{G}$ into the multiplicative group of real numbers such that for all $B \varepsilon \mathcal{L}$ and $g \varepsilon \mathcal{G}$,

$$\mu(gB) = \mu(B), \qquad \nu(Bg) = \nu(B),$$

$$\nu(B) = \mu(B^{-1}), \qquad \Delta(g)\mu(B) = \mu(Bg)$$

(see e.g. Nachbin [17]).

Assumption A2. There exist a space $\mathcal{M}$ and a one-to-one bimeasurable map π from $\mathfrak{X}$ onto $\mathcal{G} \times \mathcal{M}$ such that if $\pi(x)=(h,a)$, then $\pi(gx)=(gh,a)$.

Usually, $\mathcal{M}$ is the sample space of the maximal invariant statistic defined on $\mathfrak{X}$ with respect to $\mathcal{G}$.

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 $\mathcal{P}$ is invariant under $\mathcal{G}$, so that Assumptions 3 and A2 imply that the probability measure on $\mathcal{M}$ induced from X does not depend on $\theta\varepsilon\Theta$. Hence we shall denote it by λ.

Assumption A3. There exists a relatively invariant measure ξ on $(\mathcal{Y},\mathcal{C})$ with modulus J, i.e.

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

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

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

Assumption A4. The density function of X with respect to $\mu\times\lambda$ can be expressed in the form

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

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

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

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 predictor by using the Haar measure ν. For this we need the following lemma.

Lemma 3.([24]) *If Assumptions* 1 *to* 4 *and* A1 *to* A4 *hold and if* δ *is an invariant predictor, then for any* $h \in \mathcal{G}$,

$$R(\theta_0,\delta) = \Delta(h)\int\int\int\{\int L(s,y,\theta)\delta(ds|h,a)\}f_1(\theta^{-1}h,a)$$
$$\times f_2([\theta^{-1};h,a]y|\theta^{-1}h,a)J(\theta^{-1})\lambda(da)\nu(d\theta)\xi(dy).$$

Proof. By setting $\theta=\theta_0$ in (3.1) and (3.2), it follows from (2.3) that

$$R(\theta_0,\delta) = \int\int\int\{\int L(s,y,\theta_0)\delta(ds|g,a)\}f_1(g,a)f_2(y|g,a)$$
$$\times\ \lambda(da)\mu(dg)\xi(dy)$$
$$= \Delta(h)\int\int\int\{\int L(s,y,\theta_0)\delta(ds|g'h,a)\}f_1(g'h,a) \tag{3.3}$$
$$\times f_2(y|g'h,a)\lambda(da)\mu(dg')\xi(dy),$$

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

The invariance of L and δ (see (2.2) and (2.4)) implies that

$$\int L(s,y,\theta_0)\delta(ds|g'h,a) = \int L([g';h,a]s,y,\theta_0)\delta(ds|h,a)$$
$$= \int L(s,[g';h,a]^{-1}y,g'^{-1})\delta(ds|h,a),$$

so that after the transformation $y'=[g';h,a]^{-1}y$, we have from (3.3) that

$$R(\theta_0,\delta) = \Delta(h)\int\int\int\{\int L(s,y',g'^{-1})\delta(ds|h,a)\}f_1(g'h,a)$$
$$\times f_2([g';h,a]y'|g'h,a)J(g')\lambda(da)\mu(dg')\xi(dy')$$

Then by the transformation $\theta=g'^{-1}$ and the fact that $\nu(d\theta)=\mu(dg')$, the theorem has been proved.

On the basis of Lemma 3, we shall prove the following result, which is an extension of Theorem 2 of Hora and Buehler [9].

Theorem 1.([24]) *If Assumptions* 1 *to* 4 *and* A1 *to* A4 *hold and if there exists a non-randomized predictor* δ^* *such that for each* $x=(h,a)$, $\delta^*(x)$ *is the unique value of* d *which minimizes*

$$(3.4)\qquad \iint L(d,y,\theta)f_1(\theta^{-1}h,a)f_2([\theta^{-1};x]y|\theta^{-1}x)J(\theta^{-1})\nu(d\theta)\xi(dy),$$

then δ^* *is the best invariant predictor.*

Proof. First we shall show that δ^* is an invariant predictor. Substituting $gx=(gh,a)$ in place of $x=(h,a)$, and using the transformation $\theta=g\theta'$ and the fact that $\nu(d\theta)=\Delta(g^{-1})\nu(d\theta')$, we can write (3.4) as

$$(3.5)\qquad \Delta(g^{-1})\iint L(s,y,g\theta')f_1(\theta'^{-1}h,a)f_2([(g\theta')^{-1};gh,a]y|\theta'^{-1}h,a)$$
$$\times J((g\theta')^{-1})\nu(d\theta')\xi(dy).$$

Since by (2.7) and (2.8)

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

and $J((g\theta')^{-1})=J(g^{-1})J(\theta'^{-1})$, after the transformation $y'=[g;h,a]^{-1}y$, (3.5) becomes

$$\Delta(g^{-1})\iint L([g;h,a]^{-1}s,y',\theta')f_1(\theta'^{-1}h,a)f_2([\theta'^{-1};h,a]y'|\theta'^{-1}h,a)$$
$$\times J(\theta'^{-1})\nu(d\theta')\xi(dy'),$$

where we used (2.2). Hence from the definition of δ^* we obtain that $\delta^*(gx)=[g;x]\delta^*(x)$, which implies that δ^* is an invariant predictor.

Now, we shall show that δ^* is the best invariant predictor. From Lemma 3 and Fubini's theorem, it follows that for any invariant predictor δ,

$$R(\theta_0,\delta) = \Delta(h)\int[\{\int\int L(s,y,\theta)f_1(\theta^{-1}h,a)f_2([\theta^{-1};h,a]y|\theta^{-1}h,a)$$
$$\times J(\theta^{-1})\nu(d\theta)\xi(dy)\}\delta(ds|h,a)]\lambda(da)$$
$$\geq \Delta(h)\int[\int\int L(\delta^*(h,a),y,\theta)f_1(\theta^{-1}h,a)f_2([\theta^{-1};h,a]y|\theta^{-1}h,a)$$
$$\times J(\theta^{-1})\nu(d\theta)\xi(dy)]\lambda(da)$$
$$= R(\theta_0,\delta^*)$$

Hence from (2.6) we have the result.

Remark 1. From this theorem it turns out that the best invariant predictor ia non-randomized (cf. Kiefer [13], p.579).

3.2 Alternative expression of the best invariant predictor

The main difficulty in applying Theorem 1 to a specific prediction problem is to verify Assumptions A2 and A4, so that we shall present a set of sufficient conditions for them, assuming always Assumption A3. This enables us to rewrite the best invariant predictor in a form which is more tractable for some applications.

Condition 1. There exists a relatively invariant measure η on $(\mathcal{X},\mathcal{B})$ with modulus γ and $\mathcal{P}$ is dominated by $\eta\times\xi$ and the density function of $Z=(X,Y)$ can be expressed by

(3.6) $$\gamma(\theta^{-1})J(\theta^{-1})p(\theta^{-1}z), \quad z\in\mathcal{Z},\ \theta\in\Theta.$$

Then from (2.1) the density function of X with respect to η

is given by

$$\gamma(\theta^{-1})p_1(\theta^{-1}x), \quad x\varepsilon\mathfrak{X}, \ \theta\varepsilon\Theta, \tag{3.7}$$

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

Definition 4. $B\varepsilon\mathcal{B}$ is said to be a Borel cross-section if it intersects each orbit $\mathcal{G}x=\{gx;\ g\varepsilon\mathcal{G}\}$ precisely once.

Condition 2. There exists a Borel cross-section $B\varepsilon\mathcal{B}$.

Condition 3. $\mathfrak{X}$ is a separable complete metrizable locally compact space and $\mathcal{G}$ is a separable complete metrizable locally compact topological group acting continuously on $\mathfrak{X}$ (i.e., the mapping $(g,x)\to gx$ is continuous on $\mathcal{G}\times\mathfrak{X}$).

Then the following lemma holds. For a proof, see Theorem 1 of Bondar [3].

Lemma 4. *If* $\mathcal{G}$ *acts freely on* $\mathfrak{X}$ *and Conditions* 2 *and* 3 *holds, then Assumption* A2 *is satisfied with* $\mathcal{M}=B$, *and if* f *is a real-valued function which is integrable with respect to* η, *then*

$$\int_{\mathfrak{X}} f(x)\eta(dx) = \int_B \alpha(da)\int_{\mathcal{G}} f(ha)\gamma(h)\mu(dh) \tag{3.8}$$

for some σ*-finite measure* α *on* B.

Using this result, we shall show Assumption A4.

Lemma 5. ([24]) *If Conditions* 1 *to* 3 *holds, then Assumption* A4 *is satisfied by taking* B *as* $\mathcal{M}$, *and*

$$f_1(\theta^{-1}h,a) = \gamma(\theta^{-1}h)p_1(\theta^{-1}x)/\int\gamma(g)p_1(ga)\mu(dg) \tag{3.9}$$

and

$$(3.10) \qquad f_2([\theta^{-1};x]y|\theta^{-1}x) = p(\theta^{-1}(x,y))/p_1(\theta^{-1}x),$$

where x=ha.

Proof. From (3.7) and (3.8), the density function of X with respect to $\mu\times\alpha$ is given by $\gamma(\theta^{-1}h)p_1(\theta^{-1}ha)$. Since

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

and this is the density function of λ with respect to α, we have (3.9). From (3.6) and (3.7), (3.10) is obtained.

Theorem 2. ([24]) *If Assumptions* 1 *to* 4, A3 *and Conditions* 1 *to* 3 *hold and if there exists a non-randomized predictor* δ^* *such that for each x,* $\delta^*(x)$ *is the unique value of* d *which minimizes*

$$(3.11) \qquad \int\int L(d,y,\theta)\gamma(\theta^{-1})J(\theta^{-1})p(\theta^{-1}(x,y))\nu(d\theta)\xi(dy),$$

then δ^* *is the best invariant predictor.*

Proof. From Lemmas 4 and 5, Assumptions in Theorem 1 are satisfied. Therefore from (3.9) and (3.10), (3.4) is equal to

$$\int\int L(d,y,\theta)\gamma(\theta^{-1}h)p(\theta^{-1}(x,y))J(\theta^{-1})\nu(d\theta)\xi(dy)/\int\gamma(g)p_1(ga)\mu(dg)$$
$$= \{\gamma(h)/\int\gamma(g)p_1(ga)\mu(dg)\}\int\int L(d,y,\theta)\gamma(\theta^{-1})J(\theta^{-1})$$
$$\times p(\theta^{-1}(x,y))\nu(d\theta)\xi(dy),$$

since $\gamma(\theta^{-1}h)=\gamma(\theta^{-1})\gamma(h)$. Therefore we have the result.

Condition 4. $\mathcal{Y}$ is p-dimentional Euclidean space and $\tilde{\mathcal{G}}$ is a group of affine transformations on $\mathcal{Y}$ such that $\tilde{g}=(b,B)$

implies $\tilde{g}y=b+By$, $y\varepsilon\mathcal{Y}$, where b is a p-dimensional vector and B is a $p\times p$ non-singular matrix.

The group operation of $\tilde{\mathcal{G}}$ is defined by

$$(3.12)\qquad (b_1,B_1)(b_2,B_2) = (b_1+B_1b_2,B_1B_2),$$
$$(b,B)^{-1} = (-B^{-1}b,B^{-1}).$$

Condition 5. For any $g\varepsilon\mathcal{G}$ and $x\varepsilon\mathcal{X}$,

$$(3.13)\qquad [g;x] = (l(g,x),k(g))$$

and for any d, y $\varepsilon\mathcal{Y}$ and $\theta\varepsilon\Theta$,

$$(3.14)\qquad L(d,y,\theta) = ||k(\theta^{-1})(y-d)||^2,$$

where $||t||^2=t't$.

Lemma 6. *Under Conditions 4 and 5, it holds that*

$$(3.15)\qquad k(gg') = k(g)k(g'),\qquad g,\ g'\varepsilon\mathcal{G},$$

and that L defined by (3.14) *satisfies* (2.2).

Proof. From (2.7) and (3.13) it follows that

$$\begin{aligned}[gg';x] &= [g;g'x][g';x]\\ &= (l(g,g'x),k(g))(l(g',x),k(g'))\\ &= (l(g,g'x)+k(g)l(g',x),k(g)k(g')),\end{aligned}$$

where the last equality follows from (3.12). Since this is equal to $(l(gg',x),k(gg'))$, we have (3.15). From (3.13), (3.14) and (3.15),

$$\begin{aligned}L([g;x]d,[g;x]y,g\theta) &= ||k((g\theta)^{-1})\{k(g)(y-d)\}||^2\\ &= ||k(\theta^{-1})(y-d)||^2.\end{aligned}$$

Therefore L satisfies (2.2).

Theorem 3. *If Assumptions* 1 *to* 4, A3 *and Conditions* 1 *to* 5 *hold, then the best invariant predictor* δ^* *is given by*

$$(3.16)\quad \delta^*(x) = \{\int\int(k(\theta)k(\theta)')^{-1}\gamma(\theta^{-1})J(\theta^{-1})p(\theta^{-1}(x,y))\nu(d\theta)\xi(dy)\}^{-1}$$
$$\times\int\int(k(\theta)k(\theta)')^{-1}y\gamma(\theta^{-1})J(\theta^{-1})p(\theta^{-1}(x,y))\nu(d\theta)\xi(dy).$$

Proof. Since $k(\theta^{-1})=k(\theta)^{-1}$ by (3.15), the result is easily obtained from (3.14) and Theorem 2.

3.3 Examples

3.3.1 The location-scale model

Let R^k denote k-dimensional Euclidean space. Suppose $\mathfrak{X}=R^n$ and $\mathcal{Y}=R^m$, and that the probability density function of $Z=(x,Y)$ with respect to Lebesgue measure on R^{n+m} is

$$\sigma^{-(n+m)}\ f\{(x_1-\mu)/\sigma,\ldots,(x_n-\mu)/\sigma,(y_1-\mu)/\sigma,\ldots,(y_m-\mu)/\sigma\}$$

for some known function f, where (μ,σ) is an unknown location-scale parameter with the parameter space $\Theta=\{\theta=(\mu,\sigma);\sigma>0\}$. $\mathcal{G}$ is the group consisting of linear transformations $g=(b,c)$, $c>0$, on R^{n+m} such that

$$(3.17)\quad g(x_1,\ldots,x_n,y_1,\ldots,y_m) = (b+cx_1,\ldots,b+cx_n,b+cy_1,\ldots,b+cy_m).$$

We adopt the loss function defined by

$$(3.18)\qquad L(d,y,\theta) = \sigma^{-2m}||y-d||^2.$$

Then it is easy to verify Assumptions 1 to 4 and A3 and also Coditions 1, 3 and 4. Condition 2 is satisfied by taking

$$B = \{(\frac{x_1-\bar{x}}{s},\dots,\frac{x_n-\bar{x}}{s})\ ;(x_1,\dots,x_n)\},$$

where $\bar{x}= \sum_{i=1}^{n} x_i/n$ and $s^2= \sum_{i=1}^{n} (x_i-\bar{x})^2$.

From (3.17), $[g;x]=(b1_m,cI_m)$, where $1_m=(1,\dots,1)'$ and I_m is the identity matrix. Then by (3.18) Condition 5 holds. Therefore all assumptions and conditions in Theorem 3 are satisfied. From (3.16), the best invariant predictor is given by

$$(3.18)\quad \delta^*(x)= \frac{\int\int y\sigma^{-(n+m+2)} f(\frac{x_1-\mu}{\sigma},\dots,\frac{x_n-\mu}{\sigma},\frac{y_1-\mu}{\sigma},\dots,\frac{y_m-\mu}{\sigma})\nu(d\theta)dy}{\int\int \sigma^{-(n+m+2)} f(\frac{x_1-\mu}{\sigma},\dots,\frac{x_n-\mu}{\sigma},\frac{y_1-\mu}{\sigma},\dots,\frac{y_m-\mu}{\sigma})\nu(d\theta)dy}$$

with $\nu(d\theta)=d\mu d\sigma/\sigma$ (see Fraser [6], p.63).

Remark 2. It is easy to show that an analogous result holds when there is a location parameter only or scale parameter only.

Example 1. Denote by $X_1<X_2<\dots<X_n$ the order statistics of size n from the exponential distribution with the density function, $(1/\sigma)\exp\{-(x-\mu)/\sigma\}$, $x>\mu$, $\sigma>0$. We consider the problem of predicting X_m after observing only $X_1,\dots,X_r$, where $1\le r<m\le n$.

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

$$(3.20)\quad \frac{n!\sigma^{-(r+1)}}{(m-r-1)!(n-m)!}\exp\{-[\sum_{i=1}^{r}(x_i-\mu)+(n-m+1)(y-\mu)]/\sigma\}$$

$$\times\{\exp[-(x_r-\mu)/\sigma]-\exp[-(y-\mu)/\sigma]\}^{m-r-1}$$

for $\mu<x_1<\ldots<x_r<y$ and zero otherwise. Then the straightforward calculation shows that the best invariant predictor is given by

$$\delta^*(X) = X_r+(S/r)\sum_{i=r+1}^{m} 1/(n-i+1), \tag{3.21}$$

where

$$S = \sum_{i=2}^{r}(X_i-X_1)+(n-r)(X_r-X_1). \tag{3.22}$$

See Appendix for a proof.

3.3.2 The progression model

Let $X_1,\ldots,X_n,X_{n+1}$ be independently and identically distributed $(p+q)$-dimensional random vectors with the probability density function with respect to Lebesgue measure on R^{p+q},

$$|\Lambda|^{-1}f(||\Lambda^{-1}(x-\mu)||^2), \tag{3.23}$$

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

Suppose that $\theta=(\mu,\Lambda)$ is unkown. We shall denote by G(m) the set of all lower triangular matrices of order m with positive diagonal elements. Then $\Theta=\{\theta=(\mu,\Lambda);\ \mu\varepsilon R^{p+q},\ \Lambda\varepsilon G(p+q)\}$.

The following partitions are used in the sequel:

$$X_i=\begin{pmatrix} X_i^1 \\ X_i^2 \end{pmatrix},\ i=1,\ldots,n+1,\ \mu=\begin{pmatrix}\mu_1\\ \mu_2\end{pmatrix},\ \Lambda=\begin{pmatrix}\Lambda_{11} & 0\\ \Lambda_{21} & \Lambda_{22}\end{pmatrix}, \tag{3.24}$$

where X_i^1 and μ_1 are $p\times 1$ and $\Lambda_{11}\varepsilon G(p)$.

We consider the problem of predicting X_{n+1}^2 after observing

$X_1,\ldots,X_n,X_{n+1}^1$. Let $X=(X_1,\ldots,X_n,X_{n+1}^1)$ and $Y=X_{n+1}^2$. Define the following transformation g on $\mathcal{Z}$,

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

where $b\varepsilon R^{p+q}$ and $C\varepsilon G(p+q)$.

We shall view $\mathcal{G}$ as the Cartesian product $R^{p+q}\times G(p+q)$ with such group operations as (3.12). Then it is well known that $\mathcal{G}$ is a locally compact topological group and that the right Haar measure is given by

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

where λ_{ii} $(i=1,\ldots,p+q)$ are diagonal elements of Λ, $d\mu$ and $d\Lambda$ denote Lebesgue measures on R^{p+q} and $G(p+q)$, respectively (see Fraser [6], p.148).

By viewing $\tilde{\mathcal{G}}$ as the group defined in Condition 4, from (3.25) we have for $g=(b,C)$,

(3.27) $$[g;x] = (C_{21}x_{n+1}^1+b_2,C_{22})$$

and

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

where the same partitions as (3.24) are used for (b,C).

We shall adopt the following loss function for this problem:

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

By using (3.12) and (3.27) we can easily show that Conditions 4 and 5 are satisfied, and therefore Assumption 2 is satisfied. From (3.27), $J(g)=|C_{22}|$, which implies Assumption A3. It is clear that Conditions 1 and 3 hold. Therefore we have only to check Condition 2 to apply Theorem 3 to this problem.

Let $\mathfrak{X}_1$ be the sample space of $(X_1,\dots,X_n)$ and the action of $\mathcal{G}$ on $\mathfrak{X}_1$ be

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

For this transformation group, the Borel cross-section for the orbits in $\mathfrak{X}_1$ exists (see Fraser [6], p.145). Then using Proposition 2 of Bondar [3], from (3.28) there exists a Borel cross-section for the orbits in $\mathfrak{X}$. Hence Condition 2 is satisfied.

Since $k(\theta)=\Lambda_{22}$, by (3.16) and (3.23) the best invariant predictor is given by

$$(3.30) \qquad \delta^*(x) = \{\iint(\Lambda_{22}\Lambda_{22}')^{-1}|\Lambda|^{-(n+1)}\prod_{i=1}^{n+1} f(||\Lambda^{-1}(x_i-\mu)||^2)\nu(d\theta)dy\}^{-1}$$

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

where $x=(x_1,\dots,x_{n+1}^1)$, $y=x_{n+1}^2$ and ν is defined in (3.26).

Example 2. Let $X_1,\dots,X_n,X_{n+1}$ be independently normally distributed $(p+q)$-dimensional random vectors with unknown mean μ and unknown non-singular covariance matrix Σ. Let $n>p+q$ and $\theta=(\mu,\Sigma)$. The same partition as in (3.24) is used for Σ.

We want to predict $Y=X_{n+1}^2$ after observing $X=(X_1,\dots,X_n,X_{n+1}^1)$. This problem was considered in Ishii [11], p.482. He proposed a predictor of Y given by

$$(3.31)\qquad \delta(X) = \bar{X}_2 + S_{21}S_{11}^{-1}(X_{n+1}^1-\bar{X}_1),$$

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

and $\bar{X}_1$ is p×1 and S_{11} is p×p. But any justification of (3.31) has not appeared in literature as far as the author knows. We shall show that (3.31) is the best invariant predictor with respect to the loss function

$$(3.32)\qquad L(d,y,\theta) = (y-d)'(\Sigma_{22}-\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})^{-1}(y-d)$$

under the transformation group $\mathcal{G}$ defined by (3.25).

It is well known that $\Sigma=\Lambda\Lambda'$ with $\Lambda\varepsilon G(p+q)$. Therefore the density function is written in the form of (3.23) and (3.32) becomes $||\Lambda_{22}^{-1}(d-y)||^2$, since $\Lambda_{22}\Lambda_{22}'=\Sigma_{22}-\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$. Hence the best invariant predictor is obtained from (3.30), which will be shown to be identical with (3.31) in Appendix.

Remark 3. When q=1, it is well known that (3.31) is inadmissible if $p\geq 3$. This fact was first proved by Stein [19]. They can be improved by using the estimators given in Baranchik [1] and Takada [21].

4. Construction method based on an adequate statistic

It is well known that the class of invariant rules based on a sufficient statistic is essentially complete among the class of all invariant rules under several assumptions. For example, see Theorem 5.4.4 and 5.4.5 in Nabeya [16], p.192. We shall show that the same result holds for the prediction problem, using an adequate statistic in place of a sufficient statistic. Then, among the class of all invariant predictors based on the adequate statistic, we seek the optimal predictor, which is actually the best invariant predictor.

4.1 Adequate statistic

In this section we postulate Assumptions 1 and 2 introduced in Section 2. Let t be a measurable mapping from $(\mathcal{X},\mathcal{B})$ onto $(\mathcal{T},\mathcal{U})$ and let $T=t(X)$.

Definition 4. A statistic T is said to be adequate for X with respect to (w.r.t.) Y if T is sufficient for X and, given T, X and Y are conditionally independent.

Sugiura and Morimoto [20] provided a simple criterion which characterizes an adequate statistic as follows.

Lemma 7. *If* $\mathcal{P}$ *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, then* T *is adequate for* X *w.r.t.* Y *if and only if*

$$dP_\theta/d\lambda = h(x)f_\theta(t(x),y),$$

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

Suppose the space $\mathcal{X}$ is decomposed so that $(\mathcal{X},\mathcal{B})=(\mathcal{X}_1\times\mathcal{X}_2,\mathcal{B}_1\times\mathcal{B}_2)$. Write $X=(X_1,X_2)$ and $x=(x_1,x_2)$.

Assumption B1. $\mathcal{G}$ is a group of one-to-one transformations acting on both $\mathcal{X}_1$ and $\mathcal{X}_2$, mapping each space onto itself, and

(4.1) $$g(x_1,x_2,y) = (gx_1,gx_2,[g;x_2]y), \quad g\varepsilon\mathcal{G},$$

so that $[g;x]$ defined in (2.1) depends on x through x_2.

Let t^* be a measurable mapping from $(\mathcal{X}_1,\mathcal{B}_1)$ onto a measurable space $(\mathcal{T}^*,\mathcal{U}^*)$ and $t(x)=(t^*(x_1),x_2)$ for $x=(x_1,x_2)$. Let $(\mathcal{T},\mathcal{U})$ $=(\mathcal{T}^*\times\mathcal{X}_2,\mathcal{U}^*\times\mathcal{B}_2)$.

Assumption B2. X_1 and X_2 are independent, $T^*=t^*(X_1)$ is sufficient for X_1 and $T=t(X)$ is adequate for X w.r.t. Y.

Remark 4. If the transformation $[g;x]$ on $\mathcal{Y}$ does not depend on x, then there is no need to consider $\mathcal{X}_2$ and Assumption B2 means just that T^* is adequate for X w.r.t. Y.

Assumption B3. There exists a real-valued function Q on $\mathcal{B}_1\times\mathcal{X}_1$ such that

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

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

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

$$Q(gB|gx_1) = Q(B|x_1).$$

Definition 5. A predictor δ is said to be based on T if for any $C\varepsilon \mathcal{C}$, $\delta(C|.)$ is $\mathcal{B}_t$-measurable where $\mathcal{B}_t=\{t^{-1}(U);U\varepsilon\mathcal{U}\}$.

Now, we shall show the essential completeness of the class of all invariant predictors based on an adequate statistic among the class of all invariant predictors.

Theorem 4. ([22]) *If Assumptions* 1, 2 *and* B1 *to* B3 *hold, then for any invariant predictor* δ, *there exists an invariant predictor* δ_0 *based on* T *such that*

$$R(\theta,\delta) = R(\theta,\delta_0),\ \theta\varepsilon\Theta. \tag{4.2}$$

Proof. Define for $C\varepsilon\mathcal{C}$ and $x\varepsilon\mathcal{X}$,

$$\delta_0(C|x) = \int\delta(C|s,x_2)Q(ds|x_1). \tag{4.3}$$

Then by Theorem 1 of Takeuchi and Akahira [27], we have (4.2). Therefore the proof is complete if δ_0 is shown to satisfy (2.4). From Assumptions B1 and B3, for any $g\varepsilon\mathcal{G}$, $x\varepsilon\mathcal{X}$ and $C\varepsilon\mathcal{C}$,

$$\begin{aligned}\delta_0([g;x_2]C|gx_1,gx_2) &= \int\delta([g;x_2]C|s,gx_2)Q(ds|gx_1)\\ &= \int\delta([g;x_2]C|gs,gx_2)Q(gds|gx_1)\\ &= \int\delta(C|s,x_2)Q(ds|x_1)\\ &= \delta_0(C|x)\end{aligned}$$

where the third equality follows from the invariance of δ.

Assumption B3 is not easy to verify. The following lemma which was obtained by Hall et. al. [8], Theorem 7.1, may be useful to verify it.

Lemma 8. $\mathcal{X}_1$ *is a Borel subset of* R^n *and* $\mathcal{B}_1$ *the Borel σ-field* *of* $\mathfrak{X}_1$. *Let* f_θ *be the density function of* X_1 *w.r.t. Lebesgue measure on* R^n *such that*

$$f_\theta(x_1) = h(x_1)g_\theta(t^*(x_1)),\ x_1\varepsilon\mathfrak{X}_1,\ \theta\varepsilon\Theta,$$

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

(i) *for each* $g\varepsilon\,\mathcal{G}$, gx_1 *is continuously differentiable and the Jacobian depends only on* $t^*(x_1)$;

(ii) *for each* $g\varepsilon\mathcal{G}$, $t^*(x_1)=t^*(x_2)$ *implies* $t^*(gx_1)=t^*(gx_2)$;

(iii) $t^*(x_1)$ *is continuously differentiable and the matrix* $||[\partial t^*_j(x_1)/\partial x_{1i}];\ j=1,\ldots,k,\ i=1,\ldots,n||$ *is of rank* k, *where* $x_1=(x_{11},\ldots,x_{1n})$ *and* $t^*(x_1)=(t_1^*(x_1),\ldots,t_k^*(x_1))$;

(iv) *for each* $g\varepsilon\,\mathcal{G}$, $h(gx_1)/h(x_1)$ *depends only on* $t_1(x_1)$.

Then Assumption B3 *holds.*

Next we shall show that the class of all non-randomized invariant predictors is essentially complete among the class of all invariant predictors.

Assumption B4. $\mathcal{Y}$ is R^p and $L(d,y,\theta)$ is a convex function of d for each $y\varepsilon\mathcal{Y}$ and $\theta\varepsilon\Theta$ and $L(d,y,\theta)\to\infty$ as $||d||\to\infty$.

Assumption B5. $\tilde{\mathcal{G}}$ contains only linear transformations, that is, transformations of the form $\tilde{g}d=Bd+c$ where B is a p×p non-singular matrix and $c\varepsilon R^p$.

Assumption B6. There exists a non-randomized invariant predictor based on T.

Theorem 5. ([22]) *If Assumptions* 1, 2 *and* B1 *to* B6 *hold, then for any invariant predictor* δ *there exists a non-randomized invariant predictor* δ^* *based on* T *such that*

$$R(\theta,\delta) \geq R(\theta,\delta^*), \quad \theta\varepsilon\Theta.$$

Proof. Define δ^* by

$$(4.4) \qquad \begin{aligned} \delta^*(x) &= \int s\delta_0(ds|x) \quad \text{if } \int||s||\delta_0(ds|x)<\infty \\ &= \psi(x), \text{ otherwise,} \end{aligned}$$

where δ_0 is given by (4.3), ψ is a non-randomized invariant predictor based on T which exists by Assumption B6 and

$$\int s\delta_0(ds|x) = (\int s_1\delta_0(ds|x),\ldots,\int s_p\delta_0(ds|x)).$$

Suppose first $R(\theta,\delta_0)<\infty$. Then from (2.3)

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

Therefore from Theorem 1 and Remark in Ferguson [5], p.78 and Assumption B4, we have

$$\int||s||\delta_0(ds|x) < \infty \qquad \text{a.e. } [\mathcal{P}],$$

and hence by Jensen's inequality

$$L(\delta^*(X),Y,\theta) \leq \int L(s,Y,\theta)\delta_0(ds|X) \quad \text{a.e. } [\mathcal{P}],$$

which implies that

$$(4.5) \qquad R(\theta,\delta^*) \leq R(\theta,\delta_0), \quad \theta\varepsilon\Theta.$$

On the contrary, if $R(\theta,\delta_0)=\infty$, it is clear that (4.5) holds. Hence from (4.2) the proof is complete if δ^* is invariant, that is, $\delta^*(gx)=[g;x_2]\delta^*(x)$ for all $g\varepsilon\mathcal{G}$ and $x=(x_1,x_2)\varepsilon\mathcal{X}$.

From Assumption B5 and the invariance of δ_0, we have

$$(4.6) \qquad \int s\delta_0(ds|gx_1,gx_2) = \int [g;x_2]s\delta_0([g;x_2]ds|gx_1,gx_2) = [g;x_2]\int s\delta_0(ds|x),$$

which implies that $\int ||s||\delta_0(ds|gx)<\infty$ if and only if $\int ||s||\delta_0(ds|x) <\infty$. Then from (4.4) and (4.6) it follows easily that δ^* is invariant.

Remark 5. Theorem 5 implies that if a non-randomized invariant predictor based on T is best among the class of all non-randomized invariant predictors based on T, then it is the best invariant predictor.

4.2 Construction of the best invariant predictor

In this section we shall construct the best invariant predictor based on an adequate statistic.

Let t be a measurable mapping from $(\mathfrak{X},\mathcal{B})$ onto $(\mathcal{T},\mathfrak{U})$ and suppose that T=t(X) is an adequate statistic and that Assumptions 1 to 4 in Section 2 hold.

Assumption B7. For any $g\varepsilon\mathcal{G}$, $t(gx)=t(gx')$ whenever $t(x)=t(x')$.

Then we can define the action of $\mathcal{G}$ on $\mathcal{T}$ by

$$gt' = t(gx)$$

for $t'\varepsilon\mathcal{T}$ and x satisfying $t'=t(x)$.

Assumption B8. For each $g\varepsilon\mathcal{G}$, $[g;x]$ depends on x through t(x).

Therefore in this section we shall write $[g;t]$ instead of $[g;x]$. Then under Assumptions B7 and B8, a predictor based on T is invariant if and only if

$$(4.7) \qquad \delta([g;t]C|gt) = \delta(C|t), \quad C\varepsilon\mathcal{C}, \ t\varepsilon\mathcal{T}, \ g\varepsilon\mathcal{G} .$$

Suppose that t 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 B9. The action of $\mathcal{G}$ on both $\mathcal{T}_1$ and $\mathcal{T}_2$ can be defined so that

$$(4.8) \qquad gt = (gt_1,gt_2), \quad t=(t_1,t_2)\varepsilon\mathcal{T}, \ g\varepsilon\mathcal{G} .$$

Assumption B10. $\mathcal{T}_1$ is isomorphic to $\mathcal{G}$.

Fix $t_0\varepsilon\mathcal{T}_1$ and we shall denote by $g_{t_1}\varepsilon\mathcal{G}$ the action such that $g_{t_1}t_0=t_1$ for $t_1\varepsilon\mathcal{T}_1$.

Setting $g=g_{t_1}^{-1}$ in (4.7), we have from (4.8) that for any invariant predictor δ based on T,

$$\delta([g_{t_1}^{-1};t]C|t_0,g_{t_1}^{-1}t_2) = \delta(C|t),$$

so that after the transformation $s'=[g_{t_1}^{-1};t]s$ we have

$$(4.9) \qquad \int L(s,y,\theta_0)\delta(ds|t) = \int L(s',[g_{t_1}^{-1};t]y,g_{t_1}^{-1})\delta(ds'|t_0,g_{t_1}^{-1}t_2),$$

where we used the invariance of L (see (2.2)).

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

$$(4.10) \qquad w_1(t) = g_{t_1}^{-1}t_2, \quad t=(t_1,t_2),$$

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

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

Then by (2.8), we have

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

and from (4.9) we obtain that

(4.13) $$\int L(s,y,\theta_0)\delta(ds|t) = \int L(s,w_2,g_{t_1}^{-1})\delta(ds|t_0,w_1).$$

Assumption B11. The mappings, w_1 and w_2, are measurable.

Let $W_1=w_1(T)$ and $W_2=w_2(T,Y)$. Then from (2.5) and (4.13) we have that for any invariant predictor δ based on T,

(4.14) $$R(\theta_0,\delta) = E_{\theta_0}\{\int L(s,W_2,g_{T_1}^{-1})\delta(ds|t_0,W_1)\}.$$

Now, we have the following result.

Theorem 6. *If Assumptions* 1 *to* 4 *and* B1 *to* B3, B7 *to* B11 *hold and if there exists a mapping* h *from* $\mathcal{J}_2$ *to* $\mathcal{Y}$ *such that* $h(w_1)$ *is the value of* d *which minimizes*

$$E_{\theta_0}\{L(d,W_2,g_{T_1}^{-1})|W_1\},$$

where $E_{\theta_0}\{.|W_1\}$ *denotes the conditional expectation given* W_1, *then the predictor* δ^* *defined by*

(4.15) $$\delta^*(T) = [g_{T_1};t_0,W_1]h(W_1)$$

is the best invariant predictor.

Proof. From (4.14) and Fubini's theorem, we have

$$R(\theta_0,\delta) = E_{\theta_0}[\int E_{\theta_0}\{L(s,W_2,g_{T_1}^{-1})|W_1\}\delta(ds|t_0,W_1)]$$

$$\geq E_{\theta_0}[\int E_{\theta_0}\{L(h(W_1),W_2,g_{T_1}^{-1})|W_1\}\delta(ds|t_0,W_1)]$$

$$= E_{\theta_0}\{L(h(W_1),W_2,g_{T_1}^{-1})\}.$$

Combining the invariance of L with (4.12) and (4.15),

$$L(h(W_1),W_2,g_{T_1}^{-1}) = L(\delta^*(T),Y,\theta_0),$$

so that we have $R(\theta_0,\delta) \geq R(\theta_0,\delta^*)$. Therefore from (2.6) and Theorem 4, the proof is complete if it can be shown that δ^* is invariant.

Using the fact that $g_{g't_1}=g'g_{t_1}$ and $w_1(g't)=w_1(t)$ for $g'\varepsilon\mathcal{G}$, we have

$$\begin{aligned}\delta^*(g't) &= [g'g_{t_1};t_0,w_1]h(w_1)\\ &= [g';t][g_{t_1};t_0,w_1]h(w_1)\\ &= [g';t]\delta^*(t),\end{aligned}$$

where the second equality follows from (2.7), which proves the theorem.

Corollary. *Under assumptions in Theorem* 6 *and Conditions* 4 *and* 5 *in Section* 3.2, *the best invariant predictor is given by*

$$\delta^*(T) = k(g_{T_1})h(W_1) + l(g_{T_1},t_0,W_1)$$

and

(4.16) $$h(W_1) = [E_{\theta_0}\{k(g_{T_1})'k(g_{T_1})|W_1\}]^{-1}E_{\theta_0}\{k(g_{T_1})'k(g_{T_1})W_2|W_1\},$$

where W_1 and W_2 are given by (4.10) and (4.11), respectively.

Remark 6. For the problem of prediction region, Ishii [12] obtained the expression of the best invariant prediction region based on an adequate statistic. But his final result (Theorem 1 in [12]) is not correct without an additional assumption. Therefore in [25] we have given the corrected result and also shown the essential completeness of the class of invariant prediction regions based on an adequate statistic among the class of all invariant prediction regions.

4.3 Examples

In this section we shall consider again Examples 1 and 2 in Section 3.3 and construct the best invariant predictor for each case by using the corollary.

Example 3. We consider Example 1 in Section 3.3. Using Lemma 7 and (3.20), an adequate statistic is given by

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

where S is defined in (3.22).

We shall first verify Assumptions B1 to B3 and B7 to B11. In this case $\mathfrak{X}_2$ is not needed since the transformation on $\mathfrak{Y}$ does not depend on x. Therefore Assumptions B1, B2 and B8 are trivial. We shall prove Assumption B3 by using Lemma 8.

Let $\mathfrak{X} = \{x;\ x_1 < \ldots < x_r\}$. Then the probability density function of $X = (X_1, \ldots, X_r)$ is given for $x \varepsilon \mathfrak{X}$ by

$$\frac{n!}{(n-r)!}\sigma^{-r}\exp[-(\sum_{i=1}^{r}(x_i-\mu)+(n-r)(x_r-\mu))/\sigma]$$

for $\mu < x_1$ and zero otherwise. According to Lemma 8, let $h(x)=1$

and

$$g_\theta(t) = \frac{n!}{(n-r)!}\sigma^{-r} I_\mu(x_1)\exp[-(s+n(x_1-\mu))/\sigma],$$

where $I_\mu(x_1)=1$ if $\mu<x_1$ and zero otherwise. Then by defining $(t_1,t_2,t_3)=(x_1,s,x_r)$, it follows easiliy that the matrix $D(x)= ||[\partial t_i(x)/\partial x_j]\ ;i=1,2,3,\ j=1,\ldots,r||$ is of rank 3 and other conditions in Lemma 8 also hold. Therefore Assumption B3 is satisfied.

Define $t(x)=(x_1,s,x_r)$. Then Assumption B7 holds and the action $\mathcal{G}$ on $\mathcal{T}$ becomes

$$g(x_1,s,x_r) = (b+cx_1,cs,b+cx_r), \quad g=(b,c).$$

Taking $T_1=(X_r,S)$ and $T_2=X_1$ in Assumption B9, it follows easily that Assumption B10 is satisfied with $g_{t_1}=(x_r,s)\varepsilon\ \mathcal{G}$ and $t_0=(0,1)$. By (4.10) and (4.11)

$$W_1 = (X_1-X_r)/S$$

and

$$W_2 = (X_m-X_r)/S,$$

so that Assumption B11 holds. Therefore the best invariant predictor can be obtained from Corollary of Theorem 6.

It is well known that for $1\leq i\leq n$ the set of random variables

(4.17) $$Z_i = (n-i+1)(X_i-X_{i-1}), \quad i=1,\ldots,n$$

(where $X_0=\mu$) are mutually independent with pdf, $(1/\sigma)\exp(-x/\sigma)$, $x>0$ (see Lemma 3 in Epstein and Sobel [4]).

Using the Z_i's,

$$W_1 = -\sum_{i=2}^{r} Z_i/(n-i+1)S, \qquad S = \sum_{i=2}^{r} Z_i.$$

It is easily seen that S is complete and sufficient for $(Z_2,\ldots,Z_r)$, so that W_1 and S are independent by Theorem of Basu [2]. Since

$$X_m-X_r = \sum_{i=r+1}^{m} Z_i/(n-i+1),$$

X_m-X_r is independent of (W_1,S), so that W_1 and W_2 are independent. Therefore from (4.16) and $k(g_{T_1})=S$,

$$h(W_1) = \{E_{\theta_0}(S^2|W_1)\}^{-1}E_{\theta_0}\{S^2W_2|W_1\}$$

$$= \{E_{\theta_0}S^2\}^{-1}E_{\theta_0}S^2W_2.$$

Simple calculation at $\theta_0=(0,1)$ shows that

$$E_{\theta_0}S^2 = r(r-1)$$

and

$$E_{\theta_0}S^2W_2 = (r-1)\sum_{i=r+1}^{m} 1/(n-i+1),$$

so that we have

$$h(W_1) = r^{-1}\sum_{i=r+1}^{m} 1/(n-i+1).$$

Hence from Corollary of Theorem 6, the best invariant predictor is given by

$$\delta^*(X) = X_r + (S/r)\sum_{i=r+1}^{m} 1/(n-i+1),$$

since $k(g_{T_1})=S$ and $l(g_{T_1},t_0,W_1)=X_r$.

Example 4. We consider Example 2 in Section 3.3. From Lemma 7, it is easily seen that

$$T = (\bar{X},S,X^1_{n+1}) \tag{4.18}$$

is an adequate statistic. First we shall verify Assumptions B1 to B3.

Let $\mathfrak{X}_1$ and $\mathfrak{X}_2$ be sample spaces of $(X_1,\ldots,X_n)$ and X^1_{n+1}, respectively in Assumption B1. Then from (3.25), (3.27) and (3.28), Assumption B1 holds. Define $T^*=(\bar{X},S)$. Then (4.18) implies Assumption B2. Hall et.al. ([8], p.611) showed that the conditions of Lemma 8 are satisfied, which implies Assumption B3.

Now, we shall verify Assumptions B7 to B11. By setting $T_1=(\bar{X},S)$ and $T_2=X^1_{n+1}$, it can be easily shown that Assumptions B7 to B9 are satisfied and

$$gt = (b+C\bar{x}, CsC', b_1+C_{11}x^1_{n+1}), \quad g=(b,C). \tag{4.19}$$

Assumption B10 holds and

$$g_{T_1} = (\bar{X},A), \qquad T_1=(\bar{X},S)$$

where $S=AA'$ with $A\varepsilon G(p+q)$ and $t_0=(0,I)$. By (3.12),

$$g^{-1}_{T_1} = (-A^{-1}\bar{X}, A^{-1}),$$

so that using

$$A^{-1} = \begin{pmatrix} A^{-1}_{11} & 0 \\ -A^{-1}_{22}A_{21}A^{-1}_{11} & A^{-1}_{22} \end{pmatrix}$$

with the same partition as in (3.24) for A, we obtain that from (4.10) and (4.19)

$$W_1 = A^{-1}_{11}(X^1_{n+1}-\bar{X}_1)$$

and from (3.27) and (4.11)

$$W_2 = A_{22}^{-1}\{X_{n+1}^2-\bar{X}_2-A_{21}A_{11}^{-1}(X_{n+1}^1-\bar{X}_1)\}.$$

Therefore Assumption B11 holds. Hence by applying Corollary of Theorem 6, the best invariant predictor can be obtained.

Since $k(g_{T_1})=A_{22}$,

$$h(W_1) = \{E_{\theta_0}(A_{22}'A_{22}|W_1)\}^{-1}E_{\theta_0}(A_{22}'A_{22}W_2|W_1). \tag{4.20}$$

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 $\theta_0=(0,I)$ the conditional means of $X_{n+1}^2-\bar{X}_2$ given $(\bar{X}_1, X_{n+1}^1)$ and that of $S_{21}S_{11}^{-1}$ given S_{11} are both zero (e.g. see Theorem 6.4.1 in Giri [7]). Since $S_{11}=A_{11}A_{11}'$, $S_{22}-S_{21}S_{11}^{-1}S_{12}=A_{22}A_{22}'$ and $S_{21}S_{11}^{-1}=A_{21}A_{11}^{-1}$,

$$\begin{aligned} E_{\theta_0}\{A_{22}'A_{22}W_2|\bar{X}_1, X_{n+1}^1, A_{11}\} &= (E_{\theta_0}A_{22}')E_{\theta_0}\{X_{n+1}^2-\bar{X}_2-A_{21}A_{11}^{-1} \\ &\quad \times(X_{n+1}^1-\bar{X}_1)|\bar{X}_1, X_{n+1}^1, A_{11}\} \\ &= 0, \end{aligned}$$

which implies $h(W_1)=0$ from (4.20). Hence the best invariant predictor is given by

$$\begin{aligned} \delta^*(X) &= l(g_{T_1}, t_0, W_1) \\ &= \bar{X}_2+S_{21}S_{11}^{-1}(X_{n+1}^1-\bar{X}_1), \end{aligned}$$

since $l(g_{T_1}, t_0, W_1)=A_{21}W_1+\bar{X}_2$ and $A_{21}W_1=S_{21}S_{11}^{-1}$. This result coincides of course with (3.31).

5. Construction method via the best unbiased predictor

In this section we explore the relation of the best invariant predictor and the best unbiased predictor. By using this relation, we can construct the best invariant predictor.

Assume here Assumptions 1 to 4 in Section 2 and Conditions 4 and 5 with $\mathcal{Y} = R^1$, that is,

(5.1) $$[g;x]y = k(g)y + l(g;x),$$

(5.2) $$L(d,y,\theta) = (y-d)^2/k(\theta)^2.$$

Since the loss function is convex, it is easily shown by a similar argument as in the proof of Theorem 5 that the class of non-randomized invariant predictors is essentially complete among the class of all invariant predictors. Therefore we shall confine our attentions to non-randomized predictors.

5.1 Best unbiased predictor

Definition 5. A statistic $\delta(X)$ is said to be an unbiased predictor of Y if

$$E_\theta \delta(X) = E_\theta Y, \qquad \text{for all } \theta \varepsilon \Theta.$$

Definition 6. An unbiased predictor is said to be best if it minimizes $E_\theta(Y-\delta(X))^2$ for all $\theta \varepsilon \Theta$ among the class of all unbiased predictors.

The following lemma due to Takeuchi ([26], p.13) is useful in obtaining the best unbiased predictor.

Lemma 9. *If a statistic* $T(X)$ *is complete and sufficient for* X *and*

$$E_\theta(Y|X) = h(X) + \mu(\theta,T),$$

h *being independent of* θ, *and if an unbiased predictor* δ *of* Y *exists, then* $\delta^*=E(\delta|T)$ *is the best unbiased predictor.*

Let $S_0=\{\gamma;\ E_\theta\gamma(X)=0,\ E_\theta\gamma^2(X)<\infty$ for all $\theta\varepsilon\Theta\}$. Then the following lemma is well known (e.g. see Theorem 3.3.1 of Zacks [28]).

Lemma 10. *An unbiased estimator* $\sigma^*(X)$ *is a uniformly minimum variance (U.M.V.) unbiased estimator of* $k(\theta)$ *if and only if for any* $\theta\varepsilon\Theta$ *and* $\gamma\varepsilon S_0$,

(5.3) $$E_\theta\{\sigma^*(X)\gamma(X)\} = 0.$$

For the prediction problem, Ishii [11] obtained the following result.

Lemma 11. *An unbiased predictor* δ^* *is best if and only if for any* $\theta\varepsilon\Theta$ *and* $\gamma\varepsilon S_0$,

(5.4) $$E_\theta\{(Y-\delta^*(X))\gamma(X)\} = 0.$$

5.2 Construction of the best invariant predictor

In this section, we shall extend slightly the definition of the invariant predictor, that is, we say that a predictor δ is invariant if for any $g\varepsilon \mathcal{G}$

(5.5) $$\delta(gx) = [g;x]\delta(x) \qquad \text{a.e. } [\mathcal{P}],$$

where the exceptional set may depends on g.

It is easy to see that (2.6) holds under this extension.

Theorem 7. ([23]) *If δ^* is the best unbiased predictor, then it is an invariant predictor.*

Proof. Define $\delta_g(X)=[g;X]^{-1}\delta^*(gX)$ for $g\varepsilon\mathcal{G}$. Then from (2.8) $\delta_g(X)=[g^{-1};gx]\delta^*(gX)$, so that

$$\begin{aligned} E_\theta\delta_g(X) &= E_{g\theta}\{[g^{-1};X]\delta^*(X)\} \\ &= E_{g\theta}\{[g^{-1};X]Y\} \\ &= E_\theta\{[g^{-1};gX][g;X]Y\} \\ &= E_\theta Y, \end{aligned}$$

where the second and last equalities follow from (5.1) and (2.8), respectively, and the third follows from the fact that the distribution of $g(X,Y)=(gX,[g;X]Y)$ under P_θ is same as that of (X,Y) under $P_{g\theta}$. Therefore δ_g is an unbiased predictor of Y. By the same way we have that

(5.6) $$\begin{aligned} R(\theta,\delta_g) &= k(\theta)^{-2}E_\theta(Y-\delta_g(X))^2 \\ &= k(g\theta)^{-2}E_\theta([g;X]Y-\delta^*(gX))^2 \\ &= R(g\theta,\delta^*). \end{aligned}$$

Since δ^* is the best unbiased predictor, $R(\theta,\delta^*)\leq R(\theta,\delta_g)$, which implies from (5.6) that $R(\theta,\delta^*)\leq R(g\theta,\delta^*)$ for any $g\varepsilon\mathcal{G}$

and $\theta\varepsilon\Theta$. This shows $R(\theta,\delta^*)$ does not depend on $\theta\varepsilon\Theta$ by Assumption 3. Hence from (5.6) we have that δ_g is the best unbiased predictor, so that from Lemma 11,

$$E_\theta(\delta_g(X)-\delta^*(X))^2 = E_\theta[\{(\delta_g(X)-Y)-(\delta^*(X)-Y)\}(\delta_g(X)-\delta^*(X))]$$
$$= 0,$$

which implies that

$$\delta^*(x) = \delta_g(x) \quad \text{a.e. } [\mathcal{P}].$$

Hence we obtain that

$$\delta^*(gx) = [g;x]\delta^*(x) \quad \text{a.e. } [\mathcal{P}].$$

Remark 7. From this result it turns out that the best invariant predictor is better than or equally good as the best unbiased predictor (cf. Lehmann [14], p.23).

We say that $\gamma(X)$ is scale invariant if for any $g\varepsilon\mathcal{G}$,

$$\gamma(gx) = k(g)\gamma(x) \quad \text{a.e. } [\mathcal{P}],$$

where the exceptional set may depend on g.

Let $S_I=\{\gamma;\ \gamma \text{ is scale invariant and } E_{\theta_0}\gamma^2(X)<\infty\}$.

Theorem 8.([23]) *An invariant predictor* δ^* *is best if and only if for any* $\gamma\varepsilon S_I$,

(5.7) $$E_{\theta_0}\{(Y-\delta^*(X))\gamma(X)\} = 0.$$

Proof. By (2.6) it is enough to compare the risk function of invariant predictors only at $\theta=\theta_0$.

Assume that δ^* is the best invariant predictor. For any $\gamma\varepsilon S_I$, let $\delta(X)=\delta^*(X)+\lambda\gamma(X)$ with a constant λ. Then from (5.5)

$$\begin{aligned}\delta(gx) &= \delta^*(gx)+\lambda\gamma(gx)\\ &= [g;x]\delta^*(x)+\lambda k(g)\gamma(x) \qquad \text{a.e. } [\mathcal{P}]\\ &= [g;x](\delta^*(x)+\lambda\gamma(x)) \qquad \text{a.e. } [\mathcal{P}]\\ &= [g;x]\delta(x) \qquad \text{a.e. } [\mathcal{P}],\end{aligned}$$

where the third equality follows from (5.1), so that δ is an invariant predictor. From the equality that

$$E_{\theta_0}(Y-\delta(X))^2 = E_{\theta_0}(Y-\delta^*(X))^2-2\lambda E_{\theta_0}\{(Y-\delta^*(X))\gamma(X)\} +\lambda^2 E_{\theta_0}\gamma^2(X)$$

and the fact that δ^* is the best invariant predictor, we must have that

$$E_{\theta_0}\{(Y-\delta^*(X))\gamma(X)\} = 0.$$

Conversely, for any invariant predictor δ, let $\gamma(X)=\delta(X)-\delta^*(X)$ where δ^* is the invariant predictor which satisfies (5.7). Then it is easy to see from (5.1) that $\gamma \varepsilon S_I$, and therefore

$$\begin{aligned}E_{\theta_0}(Y-\delta(X))^2 &= E_{\theta_0}(Y-\delta^*(X))^2 + E_{\theta_0}\gamma^2(X)\\ &\geq E_{\theta_0}(Y-\delta^*(X))^2.\end{aligned}$$

Hence δ^* is the best invariant predictor.

Now we consider the relation of the best invariant predictor and the best unbiased predictor. For this we need the following lemma.

Lemma 12. ([23]) *If* $\sigma^*(X)$ *is the U.M.V. unbiased estimator of* $k(\theta)$, *then* $\sigma^*(X)\varepsilon S_I$.

Proof. For any $g\varepsilon \mathcal{G}$, define $\sigma_g(X)=\sigma^*(gX)/k(g)$ and

$$\tau(\theta,\hat\sigma) = E_\theta\{(\hat\sigma(X)-k(\theta))^2/k(\theta)^2\}.$$

Then by a similar argument as in the proof of Theorem 7, we can show that σ_g is an unbiased estimator of $k(\theta)$ and

$$\tau(\theta,\sigma_g) = \tau(g\theta,\sigma^*) \quad \text{for any } \theta\varepsilon\Theta \text{ and } g\varepsilon \mathcal{G}.$$

From this σ_g is an U.M.V. unbiased estimator of $k(\theta)$. Therefore by Lemma 10, we obtain that

$$E_\theta(\sigma_g(X)-\sigma^*(X))^2 = 0 \quad \text{for all } \theta\varepsilon\Theta,$$

which implies that

$$\sigma^*(gx) = k(g)\sigma^*(x) \quad \text{a.e. } [\mathcal{P}].$$

Theorem 9. ([23]) *Let* δ^* *be the best unbiased predictor of* Y *and* σ^* *be the U.M.V. unbiased estimator of* $k(\theta)$. *Let*

$$c_1 = E_{\theta_0}\{(Y-\delta^*(X))\sigma^*(X)\}$$

and

$$c_2 = E_{\theta_0}\sigma^{*2}(X).$$

Put

(5.8) $$\delta_I^*(X) = \delta^*(X) + (c_1/c_2)\sigma^*(X),$$

Then δ_I^* *is the best invariant predictor of* Y.

Proof. From Theorem 7 and Lemma 12, δ_I^* is an invariant predictor. Now we show that δ_I^* satisfies the conditions of Theorem 8.

If $\gamma \varepsilon S_I$, then

$$E_\theta \gamma(X) = E_{\theta_0} \gamma(\theta X) = k(\theta) E_{\theta_0} \gamma(X).$$

Therefore γ must be of the form

$$\gamma(X) = c\sigma(X) \tag{5.9}$$

for some constant c, where $\sigma(X)$ is an unbiased estimator of $k(\theta)$. It follows from Lemmas 10 and 11 that

$$E_{\theta_0}\{(Y-\delta^*(X))(\sigma(X)-\sigma^*(X))\} = 0$$

and

$$E_{\theta_0}\{\sigma^*(X)(\sigma(X)-\sigma^*(X))\} = 0.$$

Therefore from (5.9) and the definition of c_1 and c_2, we have that

$$\begin{aligned} E_{\theta_0}\{(Y-\delta_I^*(X))\gamma(X)\} &= E_{\theta_0}[\{Y-\delta^*(X)-(c_1/c_2)\sigma^*(X)\} \\ &\qquad \times\{c(\sigma(X)-\sigma^*(X))+c\sigma^*(X)\}] \\ &= cE_{\theta_0}[\{Y-\delta^*(X)-(c_1/c_2)\sigma^*(X)\}\sigma^*(X)] \\ &= 0, \end{aligned}$$

which proves the theorem.

Remark 8. For the estimation problem, Mann [15] obtained the relation of the U.M.V. unbiased estimator and the best invariant estimator. But his method is different from ours. If X and Y are independent, let $\psi(\theta)=E_\theta Y$. Then the best unbiased predictor and the best invariant predictor becomes the U.M.V. unbiased estimator and the best invariant estimator of $\psi(\theta)$, respectively, and Theorem 9 coincides with Theorem 1 of Mann [15].

Remark 9. From this result it turns out that the best invariant predictor coincides with the best unbiased predictor if and only if $c_1=0$. For example, if $k(\theta)$ is constant on Θ, then two predictors coincide, provided that the best unbiased predictor exists.

5.3 Example

In this section we shall consider only Example 1 in Section 3.3 because Takeuchi ([26], p.18) showed that a best unbiased predictor does not exist for Example 2 in Section 3.3.

Example 5. We shall obtain the best invariant predictor for Example 1 by using Theorem 9.

Using (4.17), we have that for $X=(X_1,\ldots,X_r)$,

$$E_\theta(X_m|X) = X_r + \sigma\sum_{i=r+1}^{m} 1/(n-i+1).$$

From Theorem 3 of Epstein and Sobel [4], (X_1,S) is complete and sufficient for X where S is defined in (3.22). Then by Lemma 9 it is easy to see that the best unbiased predictor of X is

$$\delta^*(X) = X_r + S(r-1)^{-1}\sum_{i=r+1}^{m} 1/(n-i+1)$$

and the U.M.V. unbiased estimator of σ $(=k(\theta))$ is

$$\sigma^*(X) = S/(r-1).$$

By (4.17) we have

$$E_{\theta_0}\{(X_m-\delta^*(X))\sigma^*(X)\} = E_{\theta_0}[\{\sum_{i=r+1}^{m}(X_i-X_{i-1})-\sigma^*(X)\sum_{i=r+1}^{m} 1/(n-i+1)\}\sigma^*(X)]$$

$$= -\sum_{i=r+1}^{m}(n-i+1)^{-1}V_{\theta_0}(\sigma^*),$$

where $V_{\theta_0}(\sigma^*)$ denotes the variance of $\sigma^*(X)$. Therefore we have that

$$c_1 = -(\sum_{i=r+1}^{m} 1/(n-i+1))/(r-1)$$

and

$$c_2 = r/(r-1).$$

Hence it follows from (5.8) that the best invariant predictor becomes

$$\delta^*_I(X) = X_r + (S/r)\sum_{i=r+1}^{m} 1/(n-i+1),$$

which coincides with (3.21).

6. Appendix

6.1 Proof of (3.21)

From (3.20) we can write the density function as

$$(6.1) \quad \frac{n!\sigma^{-(r+1)}}{(m-r-1)!(n-m)!} \exp\{-\frac{s+n(x_1-\mu)+(n-r)(y-x_r)}{\sigma}\}[\exp\{\frac{y-x_r}{\sigma}\}-1]^{m-r-1}$$

for $\mu<x_1<\ldots<x_r<y$ and zero otherwise,

where s is given by (3.22).

Since by the transformation $u=(y-x_r)/\sigma$

$$\int_{-\infty}^{x_1}[\int_{x_r}^{\infty} y\exp\{-\frac{n(x_1-\mu)+(n-r)(y-x_r)}{\sigma}\}(\exp(\frac{y-x_r}{\sigma})-1)^{m-r-1} dy]d\mu$$

$$= \int_{-\infty}^{x_1}\exp\{-\frac{n(x_1-\mu)}{\sigma}\}d\mu\int_{x_r}^{\infty} y\exp\{-\frac{(n-r)(y-x_r)}{\sigma}\}(\exp(\frac{y-x_r}{\sigma})-1)^{m-r-1} dy$$

$$= \frac{\sigma^2}{n} \int_0^{\infty}(x_r+\sigma u)h(u)du$$

and

$$\int_{-\infty}^{x_1}[\int_{x_r}^{\infty} \exp\{-\frac{n(x_1-\mu)+(n-r)(y-x_r)}{\sigma}\}(\exp(\frac{y-x_r}{\sigma})-1)^{m-r-1} dy]d\mu$$

$$= \frac{\sigma^2}{n}\int_0^{\infty}h(u)du,$$

where

$$h(u) = (e^{-u})^{n-r}(e^{u}-1)^{m-r-1},$$

from (3.19) and (6.1) we have

$$\delta^*(\underset{\sim}{X}) = \{\int_0^{\infty}\sigma^{-(r+2)}\exp(-S/\sigma)d\sigma\int_0^{\infty}h(u)du\}^{-1}[\int_0^{\infty}\sigma^{-(r+2)}\exp(-S/\sigma)$$

$$\times\{X_r\int_0^\infty h(u)du+\sigma\int_0^\infty uh(u)du\}d\sigma]$$

$$= X_r+(S/r)\int_0^\infty uh(u)du/\int_0^\infty h(u)du.$$

Now, we have only to show that

$$\int_0^\infty uh(u)du/\int_0^\infty h(u)du = \sum_{i=r+1}^{m} 1/(n-i+1). \tag{6.2}$$

This may be proved purely analytically but we give a probabilistic proof based on the random variables in (4.17).

Let $X_1<\ldots<X_n$ be order statistics of sample size n from the exponential distribution, e^{-x}, $x>0$. Then the density function of (X_r,X_m) $(r<m)$ is given by

$$\frac{n!}{(r-1)!(m-r-1)!(n-m)!}\, e^{-(x_r+(n-m+1)x_m)}(1-e^{-x_r})^{r-1}(e^{-x_r}-e^{-x_m})^{m-r-1}$$

for $0<x_r<x_m$ and zero otherwise.

Hence the density function of $Z=X_m-X_r$ becomes

$$\frac{n!}{(r-1)!(m-r-1)!(n-m)!}\int_0^\infty e^{-(n-r+1)x_r-(n-r)z}(1-e^{-x_r})^{r-1}(e^z-1)^{m-r-1}dx_r$$

$$= ce^{-(n-r)z}(e^z-1)^{m-r-1},$$

where c is some constant.

Therefore the right hand side of (6.2) is the expectation of $Z=X_m-X_r$, which is equal to $\sum_{i=r+1}^{m} 1/(n-i+1)$ by using (4.17). This shows (6.2).

6.2 Proof of (3.31)

In the sequel we shall use symbols $c_i (i=1,\ldots,6)$ to denote constants which do not depend on θ.

Since

$$||\Lambda^{-1}(x_i-\mu)||^2 = ||\Lambda_{11}^{-1}(x_i^1-\mu_1)||^2+||\Lambda_{22}^{-1}(x_i^2-\mu_2-\Lambda_{21}\Lambda_{11}^{-1}(x_i^1-\mu_1))||^2,$$

we have

$$|\Lambda|^{-1}\int yf(||\Lambda^{-1}(x_{n+1}-\mu)||^2)$$

$$= c_1|\Lambda_{11}|^{-1}(\mu_2+\Lambda_{21}\Lambda_{11}^{-1}(x_{n+1}^1-\mu_1))\exp(-||\Lambda_{11}^{-1}(x_{n+1}^2-\mu_1)||^2/2),$$

so that

$$(6.3)\quad \int\int(\Lambda_{22}\Lambda_{22}')^{-1}y|\Lambda|^{-(n+1)}\prod_{i=1}^{n+1} f(||\Lambda^{-1}(x_i-\mu)||^2)\nu(d\theta)dy$$

$$= c_2\int(\Lambda_{22}\Lambda_{22}')^{-1}(\mu_2+\Lambda_{21}\Lambda_{11}^{-1}(x_{n+1}^1-\mu_1))|\Lambda|^{-n}|\Lambda_{11}|^{-1}$$

$$\times\exp[-\{\sum_{i=1}^{n+1}||\Lambda_{11}^{-1}(x_i^1-\mu_1)||^2+\sum_{i=1}^{n}||\Lambda_{22}^{-1}(x_i^2-\mu_2-\Lambda_{21}\Lambda_{11}^{-1}(x_i^1-\mu_1))||^2\}$$

$$/2]\nu(d\theta).$$

Using the equality that

$$\sum_{i=1}^{n}||\Lambda_{22}^{-1}(x_i^2-\mu_2-\Lambda_{21}\Lambda_{11}^{-1}(x_i^1-\mu_1))||^2$$

$$=\sum_{i=1}^{n}||\Lambda_{22}^{-1}(x_i^2-\bar{x}_2-\Lambda_{21}\Lambda_{11}^{-1}(x_i^1-\bar{x}_1))||^2+n||\Lambda_{22}^{-1}(\mu_2-\bar{x}_2-\Lambda_{21}\Lambda_{11}^{-1}$$

$$\times(\mu_1-\bar{x}_1))||^2,$$

the integration (6.3) by μ_2 yields

$$(6.4)\quad c_3\int\int(\Lambda_{22}\Lambda_{22}')^{-1}(\bar{x}_2+\Lambda_{21}\Lambda_{11}^{-1}(x_{n+1}^1-\bar{x}_1))|\Lambda_{11}|^{-(n+1)}|\Lambda_{22}|^{-(n-1)}$$

$$\times\exp\{-(\sum_{i=1}^{n+1}||\Lambda_{11}^{-1}(x_i^1-\mu_1)||^2+\sum_{i=1}^{n}||\Lambda_{22}^{-1}(x_i^2-\bar{x}_2-\Lambda_{21}\Lambda_{11}^{-1}(x_i^1-\bar{x}_1))||^2)/2\}$$

$$\times\prod_{i=1}^{p+q}\lambda_{ii}^{-(p+q+1-i)}\,d\mu_1 d\Lambda.$$

Since

$$\sum_{i=1}^{n}||\Lambda_{22}^{-1}(x_i^2-\bar{x}_2-\Lambda_{21}\Lambda_{11}^{-1}(x_i^1-\bar{x}_1))||^2$$

$$= \mathrm{tr}(\Lambda_{22}\Lambda_{22}')^{-1}(S_{22}-\Lambda_{21}\Lambda_{11}^{-1}S_{12}-S_{21}\Lambda_{11}'^{-1}\Lambda_{21}'+\Lambda_{21}\Lambda_{11}^{-1}S_{11}\Lambda_{11}'^{-1}\Lambda_{21}')$$

$$= \mathrm{tr}(\Lambda_{22}\Lambda_{22}')^{-1}\{(\Lambda_{21}\Lambda_{11}^{-1}S_{11}^{1/2}-S_{21}S_{11}^{-1/2})(\Lambda_{21}\Lambda_{11}^{-1}S_{11}^{1/2}-S_{21}S_{11}^{-1/2})'$$

$$+ S_{22}-S_{21}S_{11}^{-1}S_{12}\},$$

by transforming $(\Lambda_{11},\Lambda_{21},\Lambda_{22})$ to $(\Lambda_{11},W,\Lambda_{22})$ with $W=\Lambda_{21}\Lambda_{11}^{-1}S_{11}^{1/2}$ (the Jacobian of this transformation is $|S_{11}^{-1/2}\Lambda_{11}|^p$), (6.4) becomes

$$c_4\iiiint(\Lambda_{22}\Lambda_{22}')^{-1}(\bar{x}_2+WS_{11}^{-1/2}(x_{n+1}^1-\bar{x}_1))|\Lambda_{11}|^{-(n-p+1)}|\Lambda_{22}|^{-(n-1)}$$

$$\times\exp[-\{\sum_{i=1}^{n+1}||\Lambda_{11}^{-1}(x_i^1-\mu_1)||^2+\mathrm{tr}(\Lambda_{22}\Lambda_{22}')^{-1}((W-S_{21}S_{11}^{-1/2})(W-S_{21}S_{11}^{-1/2})'$$

$$+ S_{22}-S_{21}S_{11}^{-1}S_{12})\}/2]\prod_{i=1}^{p+q}\lambda_{ii}^{-(p+q+1-i)}\,d\mu_1 d\Lambda_{11}d\Lambda_{22}dW.$$

Noticing that

$$\mathrm{tr}(\Lambda_{22}\Lambda_{22}')^{-1}(W-S_{21}S_{11}^{-1/2})(W-S_{21}S_{11}^{-1/2})'$$

$$= \mathrm{tr}\{\Lambda_{22}^{-1}(W-S_{21}S_{11}^{-1/2})\}\{\Lambda_{22}^{-1}(W-S_{21}S_{11}^{-1/2})\}'$$

and

$$\int W\exp[-\mathrm{tr}\{\Lambda_{22}^{-1}(W-S_{21}S_{11}^{-1/2})\}\{\Lambda_{22}^{-1}(W-S_{21}S_{11}^{-1/2})\}'/2]dW$$

$$= c_5|\Lambda_{22}|^p S_{21}S_{11}^{-1/2},$$

we obtain

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

$$= \{\iiint(\Lambda_{22}\Lambda_{22}')^{-1}g(x,\theta)d\mu_1 d\Lambda_{11}d\Lambda_{22}\}(\bar{x}_2+S_{21}S_{11}^{-1}(x_{n+1}^1-\bar{x}_1)),$$

where

$$g(x,\theta) = c_6|\Lambda_{11}|^{-(n-p+1)}|\Lambda_{22}|^{-(n-p-1)}\exp[-\{\sum_{i=1}^{n+1}||\Lambda_{11}^{-1}(x_i^1-\mu_1)||^2$$

$$+ \mathrm{tr}(\Lambda_{22}\Lambda_{22}')^{-1}(S_{22}-S_{21}S_{11}^{-1}S_{12})\}/2]\prod_{i=1}^{p+q}\lambda_{ii}^{-(p+q+1-i)}.$$

In a similar way we obtain

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

$$= \iiint(\Lambda_{22}\Lambda_{22}')^{-1}g(x,\theta)d\mu_1 d\Lambda_{11}d\Lambda_{22}.$$

Therefore we have (3.31).

Acknowledgment

The author would like to thank Professor M. Okamoto for his valuable advice and suggestions during the preparation of this thesis.

The essential part of this thesis was done while the author was at University of Tsukuba. The author would like to thank Professor N. Sugiura for his valuable suggestions and encouragement.

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 statistic, *Sankhyā*, 15, 377-380.

[3] Bondar, J. (1976). On Borel cross-section and maximal invariants, *Ann. Statist.*, 4, 881-891.

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

[5] Ferguson, T. S. (1967). *Mathematical Statistics: A Decision Theoretic Approach*, Academic Press, New York.

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

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

[8] 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.

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

[10] Ishii, G. (1969). Optimality of unbiased predictors, *Ann. Inst. Statist. Math.*, 21, 471-488.

[11] Ishii, G. (1978). *Statistical prediction* (in Japanese), in Basic Sūgaku, 7, Gendai-Sūgakusha, Tokyō.

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

[13] Kiefer, J. (1959). Invariance, minimax sequential estimation and continuous time process, *Ann. Math. Statist.*, 28, 573-601.

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

[15] Mann, N. R. (1969). Optimum estimators for linear functions of location and scale parameters, *Ann. Math. Statist.*, 40, 2149-2155.

[16] Nabeya, S. (1978). *Mathematical Statistics* (in Japanese), Kyōritu Press, Tokyō.

[17] Nachbin, L. (1965). *The Haar Integral*, Van Nostrand, Princeton.

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

[19] Stein, C. (1960). Multiple regression, *Contribution to Probability and Statistics*, Stanford Univ. Press.

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

[21] Takada, Y. (1979). A family of minimax estimators in some multiple regression problems, *Ann. Statist.*, 7, 1144-1147.

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

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

[24] Takada, Y. (1982). A comment on best invariant predictors, *Ann. Statist.*, 10 (to appear).

[25] Takada, Y. (1981). Application of an adequate statistic to the invariant prediction region (submitted).

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

[27] Takeuchi, K. and Akahira, M. (1975). Characterizations of prediction sufficiency (Adequacy) in terms of risk functions, *Ann. Statist.*, 3, 1018-1o24.

[28] Zacks, S. (1971). *The Theory of Statistical Inference*, Wiley, New York.