ON STATISTICAL DECISION PROBLEMS WITH NUISANCE PARAMETERS

Okuma, Akimichi
Department of Mathematics, Kyushu University

https://doi.org/10.5109/13091

出版情報：統計数理研究. 16 (3/4), pp.19-26, 1975-03. Research Association of Statistical Sciences
バージョン：
権利関係：

ON STATISTICAL DECISION PROBLEMS WITH NUISANCE PARAMETERS

By

Akimichi OKUMA*

(Received June 22, 1973)

§ 1. Introduction and Summary.

In the present paper, we consider a statistical decision problem as follows. We observe n samples $X_1, \cdots, X_n$ from a population and choose an action a according to the samples, where the action a is prescribing a decision for some population characteristic θ, and the loss for a is given by $L(\theta, a)$. And we assume that the distribution function of $X=(X_1, \cdots, X_n)^t$ is not only depending on θ but also on some other population characteristic ξ. Since our decision problem is just for parameter θ but not for ξ, so ξ is called a nuisance parameter usually.

In such a decision problem, the average loss function and Bayes risk function depend on the value of ξ. Therefore, in the paper, what we want to study in this case is to find out relations between the risk functions for the decision problem with a fixed nuisance parameter ξ and one with a prior distribution of ξ, and also find out relations between optimal (in some sense) decision functions for the above two different statistical decision problems. For some special case, the above problems were approached in the way of finding sufficient conditions for that the optimal decision function for fixed value ξ be independent of the value of ξ.

In § 2 we introduce the necessary notations and preliminaries, and also have lemmas which give the relations of the risk functions. In § 3 we define a sufficient statistic for a class of distributions with nuisance parameters according to D. A. S. Fraser [2] and show the completeness of the class of all decision functions based on the sufficient statistic. In § 4 we consider an invariant decision problem under some transformation group on the sample space. Finally, in § 5, we give the relations between optimal (minimax or Bayes sense) decision functions for the above two different decision problems.

For the sake of avoiding the mathematical complexity we assume appropriate measurability and integrability of functions in the present paper.

§ 2. Notations and Preliminaries.

For n samples $X_1, \cdots, X_n$, let $X^t=(X_1, \cdots, X_n)$, where X^t is the transposed vector of X. Let $\mathscr{X}$ be the sample space of X and $\mathfrak{B}$ be a σ-field of subsets of $\mathscr{X}$. Let $P_{\theta,\xi}$

* Department of Mathematics, Kyushu University, Fukuoka.

be the probability distribution of X, $\theta \in \Theta$ and $\xi \in \Xi$, where Θ and Ξ are subsets of some spaces and let $\mathcal{F}$ and $\mathcal{E}$ be σ-fields of Θ and Ξ respectively.

As a usual formulation of a statistical decision problem, let $\mathfrak{A}$ be the action space and $\mathcal{A}$ be its σ-field and the space of decision functions is a convex set of behavioral decision functions, $\mathcal{D}$, that is, any $\delta \in \mathcal{D}$ is a function from the sample space $\mathcal{X}$ to the family of all distribution functions on $(\mathfrak{A}, \mathcal{A})$, and then for any $x \in \mathcal{X}$, δ_x is a distribution function on $(\mathfrak{A}, \mathcal{A})$. Therefore any distribution H on $\mathcal{D}$ (provided a certain σ-field on $\mathcal{D}$) is an element of $\mathcal{D}$ in the following sense; for any $x \in \mathcal{X}$ we consider a distribution function $\bar{\delta}_x = \int \delta_x dH(\delta)$ on $\mathfrak{A}$, that is, the *H-mixture* of δ_x.

Now, we shall define the statistical decision problem $\boldsymbol{P}_\xi$, $\xi \in \Xi$,

$$\boldsymbol{P}_\xi: \quad (\Theta, \mathfrak{A}, L, P_{\theta,\xi}) . \tag{2.1}$$

That is, we choose an action a which prescribe some decision for $\theta \in \Theta$, according to the value of the sample X whose distribution is given by $P_{\theta,\xi}$, and its loss is given by $L(\theta, a)$.

Next, we suppose a prior distribution G of ξ, and let P_θ be the marginal distribution of X,

$$P_\theta = \int_\Xi P_{\theta,\xi} dG(\xi) . \tag{2.2}$$

And then we consider new statistical decision problem $\boldsymbol{P}$,

$$\boldsymbol{P}; \quad (\Theta, \mathfrak{A}, L, P_\theta) . \tag{2.3}$$

For any $\delta \in \mathcal{D}$ the risk functions of δ for $\boldsymbol{P}_\xi$ and $\boldsymbol{P}$ are denoted by $R(\theta, \xi, \delta)$ and $\hat{R}(\theta, \delta)$ respectively, where

$$R(\theta, \xi, \delta) = \int_\mathcal{X} \hat{L}(\theta, \delta_X) dP_{\theta,\xi} , \tag{2.4}$$

$$\hat{R}(\theta, \delta) = \int_\mathcal{X} \hat{L}(\theta, \delta_X) dP_\theta \tag{2.5}$$

and

$$\hat{L}(\theta, \delta_x) = \int_\mathfrak{A} L(\theta, a) d\delta_x(a) . \tag{2.6}$$

We restrict our attention to the class of all decision functions whose risk functions (2.4) (or (2.5)) exist and finite for all $\theta \in \Theta$ and $\xi \in \Xi$, in the sequel. We have the following lemmas without proofs (see T. Ferguson [1] or A. Okuma [3]).

LEMMA 1. *For each $\theta \in \Theta$ and $\delta \in \mathcal{D}$*

$$\hat{R}(\theta, \delta) = \int_\Xi R(\theta, \xi, \delta) dG(\xi) .$$

LEMMA 2. *For any distribution H on $\mathcal{D}$, let $\bar{\delta} = \int_\mathcal{D} \delta dH(\delta)$, then*

(i) $\hat{L}(\theta, \bar{\delta}) = \int_\mathcal{D} \hat{L}(\theta, \delta) dH(\delta)$,

(ii) $R(\theta, \xi, \bar{\delta}) = \int_\mathcal{D} R(\theta, \xi, \delta) dH(\delta)$

and

(iii) $\hat{R}(\theta, \bar{\delta}) = \int_\mathcal{D} \hat{R}(\theta, \delta) dH(\delta)$.

§ 3. Sufficient statistics and an essentially complete class.

In this section we propose a sufficient statistic for the family of distributions $\{P_{\theta,\xi}\,;\ \theta\in\Theta,\ \xi\in\Xi\}$ and prove the completeness of the family of all decision functions based on the sufficient statistic in $\mathscr{D}$.

First, let define a sufficient statistic for θ with a nuisance parameter ξ according to D. A. S. Fraser [2]. Let $T(x)$ be a measurable transformation on $(\mathscr{X}, \mathfrak{B})$ taking values in $(\mathscr{T}, \mathscr{B})$ where $\mathscr{B}$ is a σ-field on $\mathscr{T}$.

DEFINITION 1. *T is said to be a sufficient statistic of θ for $\{P_{\theta,\xi}\,;\theta\in\Theta, \xi\in\Xi\}$ if there exists a measurable function $P_\xi(\cdot\,|t)$ such that $P_\xi(\cdot\,|t)$ does not depend on θ and for any $A\in\mathfrak{B}$ and any $B\in\mathscr{B}$*

$$P_{\theta,\xi}(A\cap T^{-1}(B))=\int_B P_\xi(A\,|\,t)dP_\theta^T(t)\,, \tag{3.1}$$

where $P_\theta^T(\cdot)$ is the induced probability measure of $T(x)$ on $(\mathscr{T}, \mathscr{B})$ and is independent of ξ.

Some remarks for the above deffinition are mentioned.

1° The definition above is the ordinary one when no nuisance parameter is there.

2° On the other hand, if there is a nuisance parameter, $T(X)=X$ is not sufficient by the definition above but by the ordinary definition $T(X)=X$ is always sufficient.

3° The reason for 2° is that in the above definition we require the following two conditions;

- (i) The conditional distribution of X given $T=t$ is independent of θ.
- (ii) The induced probability measure on $(\mathscr{T}, \mathscr{B})$ is independent of the nuisance parameter ξ.

Since the second condition is rather restrictive, the existence of the sufficient statistic is not alwayes guaranteed.

Now, we consider the statistical decision problem $\boldsymbol{P}_\xi$, $\xi\in\Xi$, defined in (2.1). and we assume that there exists a sufficient statistic T. Let $\mathscr{D}_0$ be the class of all decision functions in $\mathscr{D}$ based on the sufficient statistic T, that is, for any $\delta\in\mathscr{D}_0$ if $x_1, x_2\in\mathscr{X}$ and $T(x_1)=T(x_2)$ then δ_{x_1} and δ_{x_2} are identical distributions on $(\mathfrak{A}, \mathscr{A})$.

THEOREM 1. *For any fixed $\xi\in\Xi$, $\mathscr{D}_0$ is an essentially complete class in $\mathscr{D}$ for the decision problem $\boldsymbol{P}_\xi$.*

PROOF. For any $\delta\in\mathscr{D}$ let $\delta_t^*=E_\xi(\delta_X\,|\,T(X)=t)$ i.e. δ_t^* is the $P_\xi(\cdot\,|t)$ mixture of δ. Then $\delta^*\in\mathscr{D}_0$ is clear. And the risk function of δ^* is, for any $\theta\in\Theta$,

$$\begin{aligned}R(\theta, \xi, \delta^*)&=\int_{\mathscr{T}}\hat{L}(\theta, \delta_t^*)dP_\theta^T(t)\\&=\int_{\mathscr{T}}\int_{\mathscr{X}}\hat{L}(\theta, \delta_X)dP_\xi(X\,|\,t)dP_\theta^T(t)\,,\end{aligned}$$

by Lemma 2 (i). And also

$$\begin{aligned}R(\theta, \xi, \delta)&=\int_{\mathscr{X}}\hat{L}(\theta, \delta_X)dP_{\theta,\xi}\\&=\int_{\mathscr{T}}\int_{\mathscr{X}}\hat{L}(\theta, \delta_X)dP_\xi(X\,|\,t)dP_\theta^T(t)\,.\end{aligned}$$

Then $R(\theta, \xi, \delta^*)=R(\theta, \xi, \delta)$, which shows that $\mathscr{D}_0$ is an essentially complete class in $\mathscr{D}$.

§ 4. Invariance.

On some statistical decision problems we may restrict our attention to invariant decision functions in the sense that the class of all invariant decision functions constitutes a complete class (or essentially complete class).

First we denote the invariance of the statistical decision problems $\boldsymbol{P}_\xi$, $\xi \in \Xi$ defined in (2.1). Let g be a measurable transformation on $(\mathfrak{X}, \mathfrak{B})$ and $\mathfrak{G}$ be a group of such measurable transformations.

DEFINITION 2. *$\{P_\xi ; \xi \in \Xi\}$ is said to be invariant under a transformation group $\mathfrak{G}$ if the following* (4.1) *and* (4.2) *are statisfied.*

(4.1) *For any $g \in \mathfrak{G}$, there exist measurable transformations $\bar{g}$ on $(\Theta, \mathcal{F})$ and $\tilde{g}$ on $(\Xi, \mathcal{E})$ such that for any $A \in \mathfrak{B}$*

$$P_{\theta,\xi}(gX \in A) = P_{\bar{g}\theta,\tilde{g}\xi}(X \in A)\,.$$

Let $\bar{\mathfrak{G}} = \{\bar{g} ; g \in \mathfrak{G}\}$ and $\tilde{\mathfrak{G}} = \{\tilde{g} ; g \in \mathfrak{G}\}$, then $\bar{\mathfrak{G}}$ and $\tilde{\mathfrak{G}}$ are also groups.

(4.2) *For any $\bar{g} \in \bar{\mathfrak{G}}$, there exists a measurable transformation $\hat{g}$ on $(\mathfrak{A}, \mathcal{A})$ such that for any $\theta \in \Theta$ and $a \in \mathfrak{A}$*

$$L(\theta, a) = L(\bar{g}\theta, \hat{g}a)\,.$$

Let $\hat{\mathfrak{G}} = \{\hat{g} ; g \in \mathfrak{G}\}$, then $\hat{\mathfrak{G}}$ is a group.

Next, we denote strongly invariant as follows.

DEFINITION 3. *$\{\boldsymbol{P}_\xi ; \xi \in \Xi\}$ is said to be strongly invariant under $\mathfrak{G}$ if*

(i) *$\{\boldsymbol{P}_\xi ; \xi \in \Xi\}$ is invariant under $\mathfrak{G}$*

and

(ii) *$\tilde{\mathfrak{G}} = \{\tilde{e}\}$ i. e. $\tilde{\mathfrak{G}}$ in* (4.1) *has only one element, which is identity transformation on $(\Xi, \mathcal{E})$.*

The second condition in Definition 3 is that for any $g \in \mathfrak{G}$ there exists a measurable transformation $\bar{g}$ on $(\Theta, \mathcal{F})$ such that for any $A \in \mathfrak{B}$

$$P_{\theta,\xi}(gX \in A) = P_{\bar{g}\theta,\xi}(X \in A)\,. \tag{4.3}$$

EXAMPLE. If θ is a location parameter of X and ξ is a scale parameter and if $\mathfrak{G}$ is a translation group on $\mathfrak{X}$ then (4.3) is satisfied. Therefore if the loss function for this problem is given by a function of $\theta - a$ then this statistical decision problem is strongly invariant. For example, X has normal distribution with mean θ and variance ξ, and we consider the estimation problem of θ with squared error loss function. Then the estimation problem is strongly invariant under the translation group.

DEFINITION 4. *A family of distributions $\{P_{\theta,\xi} ; \theta \in \Theta, \xi \in \Xi\}$ is said to be invariant under $\mathfrak{G}$ if* (4.1) *holds, and strongly invariant under $\mathfrak{G}$ if* (4.3) *holds for any $g \in \mathfrak{G}$, any $\theta \in \Theta$, any $\xi \in \Xi$ and any $A \in \mathfrak{B}$.*

DEFINITION 5. *A decision function $\delta \in \mathcal{D}$ is said to be invariant under $\mathfrak{G}$ if for any $g \in \mathfrak{G}$, any $x \in \mathfrak{X}$ and any $E \in \mathcal{A}$*

$$\delta_x(E) = \delta_{gx}(\hat{g}E)\,, \tag{4.4}$$

where $\hat{g}E = \{\hat{g}a ; a \in E\}$.

Let δ^g be defined by $\delta_x^g(E)=\delta_x(\hat{g}E)$ for any $x\in\mathfrak{X}$ and any $E\in\mathcal{A}$, then δ is invariant under $\mathfrak{G}$ if and only if $\delta^g=\delta$ for any $g\in\mathfrak{G}$.

THEOREM 2. *If $\{\boldsymbol{P}_\xi\,;\,\xi\in\Xi\}$ is invariant under $\mathfrak{G}$ and G, a distribution on $(\Xi,\mathcal{E})$, is invariant under $\mathfrak{G}$, that is, for any $g\in\mathfrak{G}$ and any $E\in\mathcal{E}$, $G(\hat{g}^{-1}E)=G(E)$, then $\boldsymbol{P}$ is invariant under $\mathfrak{G}$.*

PROOF. It is sufficient to show that for any $g\in\mathfrak{G}$ and any $A\in\mathfrak{B}$

$$P_\theta(gX\in A)=P_{\bar{g}\theta}(X\in A)\,. \tag{4.5}$$

By the invariance of $P_{\theta,\xi}$

$$\begin{aligned}P_\theta(gX\in A)&=\int_\Xi P_{\theta,\xi}(gX\in A)dG(\xi)\\&=\int_\Xi P_{\bar{g}\theta,\bar{g}\xi}(X\in A)dG(\xi)\,.\end{aligned}$$

By taking the transformation, $\tilde{g}\xi=\eta$, and using the invariance of G we have

$$\begin{aligned}P_\theta(gX\in A)&=\int_\Xi P_{\bar{g}\theta,\eta}(X\in A)\,dG(\eta)\\&=P_{\bar{g}\theta}(X\in A)\,.\end{aligned}$$

In this proof we may see that if $\{\boldsymbol{P}_\xi\,;\,\xi\in\Xi\}$ is strongly invariant then $\boldsymbol{P}$ is invariant without any conditions for G.

In the rest of this section we are going to see some relation between the invariance and sufficiency defined in § 3. Let S be a statistic on $(\mathfrak{X},\mathfrak{B})$ taking values in $(\mathcal{S},\mathcal{C})$ where $\mathcal{C}$ is a σ-field for which S is measurable. And let g^* be the induced transformation of g on $(\mathcal{S},\mathcal{C})$ i. e.

$$g^*S(X)=S(gX)\,. \tag{4.6}$$

Then if $P_{\theta,\xi}$ is invariant under $\mathfrak{G}$ then the induced probability measure of S on $(\mathcal{S},\mathcal{C})$, $P^S_{\theta,\xi}$, is invariant under $\mathfrak{G}^*$ where $\mathfrak{G}^*$ is the set of all induced transformations by $\mathfrak{G}$, i.e. $\mathfrak{G}^*=\{g^*\,;\,g\in\mathfrak{G}\}$, which is a group. Since $P^S_{\theta,\xi}$ is the induced probability measure, for any $g\in\mathfrak{G}$ and any $C\in\mathcal{C}$

$$P^S_{\theta,\xi}(g^*S\in C)=P_{\theta,\xi}(S(gX)\in B)=P_{\theta,\xi}(gX\in S^{-1}B)\,.$$

Then by the invariance of $P_{\theta,\xi}$,

$$\begin{aligned}P_{\theta,\xi}(gX\in S^{-1}B)&=P_{\bar{g}\theta,\bar{g}\xi}(X\in S^{-1}B)=P_{\bar{g}\theta,\bar{g}\xi}(S(X)\in B)\\&=P^S_{\bar{g}\theta,\bar{g}\xi}(S\in B)\,.\end{aligned}$$

Especially, if the statistic T is sufficient then

$$P^T_\theta(g^*T\in B)=P^T_{\bar{g}\theta}(T\in B)\,. \tag{4.7}$$

Therefore, when $P_{\theta,\xi}$ is invariant under $\mathfrak{G}$, a necessary and sufficient condition for that a statistic T be sufficient is given by the following theorem.

THEOREM 3. *Suppose that $P_{\theta,\xi}$ is invariant under $\mathfrak{G}$ and $\tilde{\mathfrak{G}}$, induced by $\mathfrak{G}$, is a transitive transformation group on Ξ, i. e. for any ξ, $\xi'\in\Xi$ $\exists\tilde{g}\in\tilde{\mathfrak{G}}$ such that $\xi'=\tilde{g}\xi$, then a statistic T is sufficient if and only if*

(i) *the conditional distribution of X given T does not depend on θ, and*

(ii) *the induced distribution of T, $P^T_{\theta,\xi}$, is strongly invariant under the induced transformation group $\mathfrak{G}^*$ on $\mathfrak{T}$ by $\mathfrak{G}$.*

PROOF. Only if part is trivial. To show if part it is sufficient to prove that for any $\xi, \xi' \in \Xi$, $P^T_{\theta,\xi}$ and $P^T_{\theta,\xi'}$ are identical. By the assumption, for any $\xi, \xi' \in \Xi$ there exists $\tilde{g} \in \tilde{\mathfrak{G}}$ such that $\xi' = \tilde{g}\xi$, and then for any $C \in \mathcal{C}$

$$\begin{aligned} P^T_{\theta,\xi'}(T \in C) &= P^T_{\theta,\tilde{g}\xi}(T \in C) = P_{\theta,\tilde{g}\xi}(X \in T^{-1}C) \\ &= P_{\bar{g}^{-1}\theta,\xi}(gX \in T^{-1}C) = P^T_{\bar{g}^{-1}\theta,\xi}(T(gX) \in C) \\ &= P^T_{\bar{g}^{-1}\theta,\xi}(g^*T \in C)\,. \end{aligned}$$

Then by (ii), $P^T_{\bar{g}^{-1}\theta,\xi}(g^*T \in C) = P^T_{\theta,\xi}(T \in C)$.

§ 5. Optimal decision functions.

As we have mentioned in § 1, the purpose of this paper is to have some relations between the optimal decision functions for $\boldsymbol{P}_\xi$ and $\boldsymbol{P}$. Before we discuss the main results, let prove the following lemma.

LEMMA 3. *For any decision function $\delta \in \mathfrak{D}_0$, the risk function $R(\theta, \xi, \delta)$ is constant in ξ and then*

$$R(\theta, \xi, \delta) = \hat{R}(\theta, \delta) \tag{5.1}$$

for any $\xi \in \Xi$.

PROOF. For any $\xi \in \Xi$

$$\begin{aligned} R(\theta, \xi, \delta) &= \int_{\mathfrak{X}} L(\theta, \delta_{T(X)}) dP_{\theta,\xi}(X) \\ &= \int_{\mathfrak{T}} \int_{\mathfrak{X}} L(\theta, \delta_{T(X)}) dP_\xi(X \mid T = t) dP^T_\theta(t) \\ &= \int_{\mathfrak{T}} L(\theta, \delta_t) dP^T_\theta(t)\,. \end{aligned}$$

Therefor $R(\theta, \xi, \delta)$ is independent of ξ. Furthermore

$$\begin{aligned} \hat{R}(\theta, \delta) &= \int R(\theta, \xi, \delta) dG(\xi) \\ &= R(\theta, \xi, \delta)\,. \end{aligned}$$

The above lemma playes the essential part of the proofs of the following main results.

THEOREM 4. *Let δ^ξ be a minimax decision function for the decision problem $\boldsymbol{P}_\xi$, for any $\xi \in \Xi$, and we suppose that δ^ξ is based on a sufficient statistic T of θ for $P_{\theta,\xi}$. And let $\delta^* = \int_\Xi \delta^\xi dG(\xi)$. If there exists a minimax decision function for $\boldsymbol{P}$ in $\mathfrak{D}$ then δ^* is also a minimax decision function.*

PROOF. By lemma 2,

$$\hat{R}(\theta, \delta^*) = \int \hat{R}(\theta, \delta^\xi) dG(\xi)\,. \tag{5.2}$$

By the assumption we have a minimax decision function, and by Theorem 1 we have a minimax decision function $\bar{\delta}$ in $\mathcal{D}_0$. Then from Lemma 3, (5.2) and the mimaxity of δ^{ξ} for $\boldsymbol{P}_{\xi}$, we have

$$\sup_{\theta\in\boldsymbol{\Theta}} \hat{R}(\theta, \delta^*) \leqq \int \sup_{\theta\in\boldsymbol{\Theta}} \hat{R}(\theta, \delta^{\xi}) dG(\xi) = \int \sup_{\theta\in\boldsymbol{\Theta}} R(\theta, \xi, \delta^{\xi}) dG(\xi)$$

$$\leqq \int \sup_{\theta\in\boldsymbol{\Theta}} R(\theta, \xi, \bar{\delta}) dG(\xi) = \int \sup_{\theta\in\boldsymbol{\Theta}} \hat{R}(\theta, \bar{\delta}) dG(\xi) = \sup_{\theta\in\boldsymbol{\Theta}} R(\theta, \bar{\delta}) .$$

Therefore δ^* is minimax.

Next we shall discuss for Bayes decision function. Let τ be a probability measure on $(\Theta, \mathcal{T})$ and r be the *Bayes risk function*, i. e. for any $\delta \in \mathcal{D}$

$$r(\tau, \xi, \delta) = \int R(\theta, \xi, \delta) d\tau(\theta) , \tag{5.3}$$

$$\hat{r}(\tau, \delta) = \int \hat{R}(\theta, \delta) d\tau(\theta) . \tag{5.4}$$

THEOREM 5. *For each $\xi \in \Xi$, let δ^{ξ} be a Bayes decision function with respect to* (wrt) τ *and $\delta^{\xi} \in \mathcal{D}_0$. Then δ^* defined in Theorem* 4 *is a Bayes decision function* wrt τ *for* $\boldsymbol{P}$.

PROOF. Since δ^{ξ} is a Bayes wrt τ for $\boldsymbol{P}_{\xi}$, for each $\xi \in \Xi$

$$r(\tau, \xi, \delta^{\xi}) \leqq r(\tau, \xi, \delta) \qquad \text{for all} \quad \delta \in \mathcal{D} . \tag{5.5}$$

On the other hand for any $\delta \in \mathcal{D}$, by Fubini's Theorem,

$$\begin{aligned} \hat{r}(\tau, \delta) &= \int \hat{R}(\theta, \delta) d\tau(\theta) \\ &= \iint R(\theta, \xi, \delta) dG(\xi) d\tau(\theta) = \iint R(\theta, \xi, \delta) d\tau(\theta) dG(\xi) \\ &= \int r(\tau, \xi, \delta) dG(\xi) . \end{aligned} \tag{5.6}$$

Therefore by the diffinition of δ^*, (5.5) and (5.6), for any $\delta \in \mathcal{D}$

$$\begin{aligned} \hat{r}(\tau, \delta^*) &= \iint R(\theta, \xi, \delta^{\xi}) d\tau \, dG(\xi) \\ &= \int r(\tau, \xi, \delta^{\xi}) dG(\xi) \leqq \int r(\tau, \xi, \delta) dG(\xi) \\ &= \hat{r}(\tau, \delta) . \end{aligned}$$

Then δ^* is a Bayes decision function wrt τ for $\boldsymbol{P}$.

In Theorem 4 and Theorem 5 we assume that δ^{ξ} is based on a sufficient statistic T, but this assumption is not essential when a sufficient statistic T exists, because of Theorem 1.

References

[1] T.S. FERGUSON, *Mathematical Statistics.* Academic Press, (1967) New York.
[2] D.A.S. FRASER, *Sufficient statistics with nuisance parameters.* Ann. Math. Stat., **27** (1956), 836-842.
[3] A. OKUMA, *On invariant sequential decision problems.* Mem. Fac. Sci., Kyushu Univ., **27** (1973), 121-134.