MINIMAX ESTIMATION METHOD FOR THE OPTIMUM DECOMPOSITION OF A SAMPLE SPACE BASED ON PRIOR INFORMATION

KAZUO NODA AND YASUSHI TAGA

(Received May 18, 1970)

Summary

Existences and expressions of minimax estimators based on prior information for the optimum decomposition of a given sample space are studied in a unifying way such that answers can be easily obtained by applying the theorems given in this paper to various statistical problems—optimum selection regions, tolerance regions, prediction regions, optimum stratifications and so on.

Further, ε-approximations to those estimators are given by choosing a finite family of probability measures from the infinite family of them under consideration in such a way that the former may be considered to be a sufficiently good approximation to the latter in the sense of risk function.

Besides, those estimators are proved to be consistent in the sense that the risk decreases to zero as the amount of prior information increases infinitely.

1. Introduction

Those statistical problems cited at the end of the summary may be considered as ones in which some optimum decompositions of given sample spaces are to be searched for in order to minimize losses (or maximize gains) suitably defined under some constraints.

Let $\mathcal{X}$ be a sample space to which a σ-field $\mathfrak{A}$ and certain nonnegative σ-finite measures ν are given, and let X be a random variable in $\mathcal{X}$ subject to a probability measure P (unknown) among a family $\mathcal{P}_1$ of probability measures suitably defined in advance. Here ν may be taken for P or the product measure $P'\nu_0$, where ν_0 is a certain nonnegative measure and P' is an induced measure by P.

A vector-valued measurable function $\phi=(\phi_1, \phi_2, \cdots, \phi_l)$ defined on $\mathcal{X}$ is called an l-decomposition of $\mathcal{X}$ if it satisfies the relation

$$(1.1)\qquad \sum_{j=1}^{l}\phi_j(x)=1\ ,\qquad 0\leqq\phi_j(x)\leqq 1\ ,$$

We shall denote the set of all such functions ϕ by Φ, and its subset of all functions ϕ satisfying some given constraints C by $\Phi(C)$. Further we shall call a vector-valued function $\phi_P=(\phi_{P1},\cdots,\phi_{Pl})$ "an optimum l-decomposition of $\mathcal{X}$ for $P(\in\mathcal{P}_1)$ among $\Phi(C)$" if it attains the infimum (or supremum) of a certain real-valued function $v(\phi, P)$ for a fixed P in $\mathcal{P}_1$ when ϕ varies among $\Phi(C)$.

For example, in the optimum stratification problem $v(\phi, P)$ may be the variance of the stratified estimator of the population mean under a stratification ϕ, and in the classification problem $v(\phi, P)$ may be the probability of misclassification due to a classification ϕ when P is the true probability measure among $\mathcal{P}_1$.

It should be noted here that $\Phi(C)$ may be identical to Φ when no restriction is imposed upon the problem under consideration as in the case of optimum stratification.

If the true probability distribution P of the random variable X under consideration is known, then we could obtain an optimal decomposition ϕ_P for P among $\Phi(C)$, which is not necessarily uniquely determined. Since the true P is not known in general, we have considered Bayes or minimax solutions among $\Phi(C)$ according to whether there is given an a priori distribution ξ over $\mathcal{P}_1$ or not.

Now, let us consider the case where neither true distribution P nor a priori distribution ξ is known but prior information S is given. In this case we could expect to get a better minimax estimator $\tilde{\phi}$ of ϕ_P with consistency property in the sense of risk by the fact that the probability distribution of prior information S may play an important role in taking the expectation of the loss function defined below as if it were an a priori distribution over the family $\mathcal{P}_1$. Here we consider the set $\hat{\Phi}(C)$ of all measurable mappings $\hat{\phi}$ from $\mathcal{S}$ to $\Phi(C)$ which can be expressed by a jointly measurable vector-valued function

$$\hat{\phi}(x, s)=(\hat{\phi}_1(x, s),\cdots,\hat{\phi}_l(x, s))$$

defined on the product space $\mathcal{X}\times\mathcal{S}$ satisfying the restriction C for any fixed s, where $\mathcal{S}$ denotes the space of prior information S. Here $\Phi(C)$ and $\hat{\Phi}(C)$ may be considered as an action space and a space of decision functions respectively. If we take up a $\hat{\phi}(x, s)$ instead of an optimum ϕ_P for true P, then the nonnegative additional loss

$$(1.2)\qquad L(\hat{\phi}, P)=v(\hat{\phi}, P)-v(\phi_P, P)$$

will be caused and must be added to the minimum loss $v(\phi_P, P)$. Therefore $L(\hat{\phi}, P)$ may be considered as a suitable measure for representing the deviation of $\hat{\phi}$ from the optimum ϕ_P.

Further let us define the risk function $r(\hat{\phi}, P)$ by the relation

$$r(\hat{\phi}, P)=E\{L(\hat{\phi}, P)\}, \tag{1.3}$$

where expectation should be taken over the space $\mathcal{S}$ of prior information (see Section 3). Under fairly general assumptions we should search for the minimax estimator $\tilde{\phi}$ of ϕ_P among $\Phi(C)$ such that

$$\sup_{P\in\mathcal{P}_1} r(\tilde{\phi}, P)=\inf_{\hat{\phi}\in\hat{\Phi}(C)}\ \sup_{P\in\mathcal{P}_1} r(\hat{\phi}, P), \tag{1.4}$$

where supremum should be taken among the family $\mathcal{P}_1$ and infimum should be taken among $\hat{\Phi}(C)$.

The main objects of this paper (Sections 4–5) lies first in showing the existence of a minimax estimator $\tilde{\phi}$ among $\hat{\Phi}(C)$ under fairly general conditions imposed upon the family $\mathcal{P}_1$, and second in obtaing the expression of $\hat{\phi}$. However, we shall show existence and expression of a Bayes estimator $\hat{\phi}_\xi$ with respect to a given a priori probability measure ξ over $\mathcal{P}_1$ in order to show that the minimax estimator $\tilde{\phi}$ can be obtained as the limit of a sequence of Bayes estimator $\{\phi_{\xi i}\}$ in the sense of the risk function where the sequence $\{\xi_i\}$ of a priori probability measures should be suitably chosen (see Sections 4–5).

Further, ε-approximations $\hat{\phi}_{\xi_\varepsilon}$ and $\tilde{\phi}_\varepsilon$ for $\hat{\phi}_\xi$ and $\tilde{\phi}$ will respectively, be given by using a finite number of representative probability measures $\mathcal{P}_\varepsilon$ suitably chosen from elements in the ε-covering of $\mathcal{P}_1$ such that

$$|r(\hat{\phi}_{\xi_\varepsilon}, \xi_\varepsilon)-r(\hat{\phi}_\xi, \xi)|\leqq\varepsilon, \tag{1.5}$$

and

$$\left|\sup_{P\in\mathcal{P}_\varepsilon} r(\tilde{\phi}_\varepsilon, P)-\sup_{P\in\mathcal{P}_1} r(\tilde{\phi}, P)\right|\leqq\varepsilon, \tag{1.6}$$

where ξ_ε is the induced probability measure on $\mathcal{P}_\varepsilon$ by ξ (see Section 6).

Weak consistency of a Bayes estimator $\hat{\phi}_\xi$ and a minimax estimator $\tilde{\phi}$ will be shown (in Section 7). Finally, we shall show how our theory can be specialized to various statistical problems to search for the optimum regions (see Section 8), and give proofs of all the lemmas in the first eight sections (in Section 9).

The first half of this paper, from the beginning to Section 5, is included in the foregoing paper submited to the Review of the I.S.I., though only a formulation of the problem and main theorems without proofs are stated there.

So we intend to give proofs to the theorems and add some statements of consistencies and ε-approximation methods in the second half of this paper.

We shall publish succeeding papers in the near future, with the following results:

1) Constructions of $\hat{\phi}$, $\tilde{\phi}$ and $\hat{\phi}_{\varepsilon}$ in some practical situations.
2) Constructions of ε-approximations to them when the family $\mathcal{P}_1$ of distribution functions is totally bounded in some sense.
3) Asymptotic behaviors (the speed of convergence on consistencies) of them when the size of prior information becomes sufficiently large.
4) Specializations in detail of the main theorems to selection regions, tolerance regions, statistical prediction regions, optimum stratifications and so on.

2. Formulation of the problem

First we define the function $v(\phi, P)$ in the following way:

$$v(\phi, P)=h[\psi(\phi, P)]\ , \tag{2.1}$$

where $h(z)$ is a real-valued function defined on a convex set $\mathcal{Z}_1$ in the $(k\times l)$-matrix space $\mathcal{Z}$, the (i, j)th element of $\psi(\phi, P)$ is defined by

$$\psi_{ij}(\phi, P)=\int_{\mathcal{X}} g_i(x)\phi_j(x)dP(x)\ , \qquad (i=1,\cdots, k,\ j=1,\cdots, l) \tag{2.2}$$

and $g=(g_1,\cdots, g_k)$ is a given vector-valued function on $\mathcal{X}$ and integrable with respect to any $P\in\mathcal{P}_1$.

The restriction C may be defined by the inequality*

$$\tau(\phi)\leqq C \qquad (\text{or } =C)\ , \tag{2.3}$$

where C is a given matrix in $\mathcal{Z}$, and the (i, j)th element of $\tau(\phi)$ is defined by

$$\tau_{ij}(\phi)=\int_{\mathcal{X}} f_i(x)\phi_j(x)d\nu(x)\ , \qquad (i=1,\cdots, m\ ;\ j=1,\cdots, l) \tag{2.4}$$

and $f=(f_1,\cdots, f_m)$ is a given vector-valued function on $\mathcal{X}$ and integrable with respect to ν and P in $\mathcal{P}_1$.

We assume that

$$\Phi(C)=\{\phi : \phi\in\Phi,\ \tau(\phi)\leqq C\ (\text{or } =C)\}$$

is nonempty throughout this paper. An optimum l-decomposition ϕ_P

---

* This inequality should be interpreted in the componentwise sense:

$$\tau_{ij}(\phi)\leqq C_{ij} \qquad (i=1,\cdots, m;\ j=1,\cdots, l)\ .$$

for $P \in \mathcal{P}_1$ is one which attains $\inf\{v(\phi, P);\ \phi \in \Phi(C)\}$. $\Phi(C)$ may be taken for an action space in decision theory.

The loss function is defined by

$$L(\phi, P) = v(\phi, P) - v(\phi_P, P)\ . \tag{2.5}$$

There a decision function is defined as a separably-valued and weak $\mathcal{B}$-measurable mapping* from the space of prior informations $(\mathcal{S}, \mathcal{B})$ into $\Phi(C)$, where $\mathcal{B}$ is a σ-field given to $\mathcal{S}$ and $Q_P = Q(\cdot \mid P)$ is a probability measure on $(\mathcal{S}, \mathcal{B})$ induced by P when P is true among $\mathcal{P}_1$. Besides, the family of all such probability measure Q_P's is denoted by $\mathcal{Q}_1$.

For instance, Q_P may be taken for the nth product measure of P if prior information is obtained as the (first) sample $(x_1, \cdots, x_n)$ of size n distributed independently and identically according to the probability measure P.

A decision function $\check{\phi}$ can be expressed by a $\mathfrak{A} \times \mathcal{B}$-measurable vector valued function $\hat{\phi}(x, s) = (\hat{\phi}_1(x, s), \cdots, \hat{\phi}_l(x, s))$ on the product space $\mathcal{X} \times \mathcal{S}$ such that

$$\hat{\phi}_j(x, s) \geqq 0 \qquad \sum_{j=1}^{l} \hat{\phi}_j(x, s) = 1 \qquad (j = 1, 2, \cdots, l) \tag{2.6}$$

(see Proposition 1 in Section 9, Appendix). The space of all such $\hat{\phi}(x, s)$ is denoted by $\hat{\Phi}$, and $\hat{\Phi}(C)$ is the subspace which satisfies the restriction

$$\tau(\hat{\phi}(x, s)) \leqq C\ . \tag{2.7}$$

The risk function $r(\hat{\phi}, P)$ defined by (1.3) may be expressed explicitly as

$$\begin{aligned} r(\hat{\phi}, P) &= \int_{\mathcal{S}} L(\hat{\phi}, P) dQ_P \\ &= \int_{\mathcal{S}} \{h[\psi(\hat{\phi}, P)] - h[\psi(\phi_P, P)]\} dQ_P\ . \end{aligned} \tag{2.8}$$

The main object of this paper lies in obtaining a minimax estimator $\tilde{\phi}$ which attains

$$\inf_{\hat{\phi} \in \hat{\Phi}} \sup_{P \in \mathcal{P}_1} r(\hat{\phi}, P) \tag{2.9}$$

under the restriction (2.7). However we shall state how to obtain Bayes estimators for a certain family of all possible a priori probability meas-

* Let $(\mathcal{S}, \mathcal{B}, m)$ be a measure space, and $x(s)$ a mapping from $\mathcal{S}$ to a B-space $\mathcal{Y}$. $x(s)$ is called weakly $\mathcal{B}$-measurable if for any $f \in \mathcal{Y}^*$, the numerical function $f(x(s))$ of s is $\mathcal{B}$-measurable.

ures, and then how the minimax estimator $\tilde{\phi}$ may be expressed as the limit of a sequence of Bayes estimators corresponding to a sequence of a priori probability measures suitably chosen from the family of a priori probability measures. Let $\tilde{C}$ be a σ-field over $\{(P, Q_P);\ P \in \mathcal{P}_1\}$, and Ξ the family of all possible a priori probability measure ξ's over $(\{(P, Q_P)\}, \tilde{C})$. Then, to each $\xi \in \Xi$ there corresponds a Bayes estimator $\hat{\phi}_\xi \in \hat{\Phi}(C)$ which attains the infimum of

$$r(\hat{\phi}, \xi)=\int_{\{(P, Q_P)\}} r(\hat{\phi}, P)d\xi \tag{2.10}$$

under the restriction (2.7).

In the following sections we use the notation $(\mathcal{P}_1, C)$ for $(\{P, Q_P\}, \tilde{C})$ for the sake of convenience when we consider a priori probability measures which should be distinguished from ones defined by the usual formulation.

Further we shall state some properties of those Bayes estimators which may be useful in showing properties of the minimax estimator.

3. Existence and expression of an optimum decomposition ϕ_P

At first we show the existence of an optimum decomposition ϕ_P which attains $\inf_{\phi \in \Phi(C)} v(\phi, P)$, under the following assumption.

ASSUMPTION 1. The σ-field $\mathfrak{A}$ over $\mathcal{X}$ has a countable number of generators.

THEOREM 3.1 (*Existence of an optimum decomposition*). *Under Assumption* 1, *there exists an optimum l-decomposition* ϕ_P *for any fixed* P *in* $\mathcal{P}_1$ *which attains the infimum of* $v(\phi, P)$ *defined by* (2.1) *among* $\phi(C)$, *if the function h is lower semicontinuous on* $\mathcal{Z}_1$.

Further, under the additional condition that h is concave on $\mathcal{Z}_1$, ϕ_P *becomes explicitly expressible.*

PROOF. We can find a probability measure P_0 on $(\mathcal{X}, \mathfrak{A})$ with respect to which ν is absolutely continuous. Then we define a probability measure P^* on $(\mathcal{X}, \mathfrak{A})$ for $P \in \mathcal{P}_1$ by

$$P^*(E)=(P(E)+P_0(E))/2 \qquad \text{for } E \in \mathfrak{A}\ . \tag{3.1}$$

Now we introduce a topology into Φ, called P^*-weak topology, in the sense of weak convergence, namely $\phi_{(r)} \to \phi_{(0)}$ $(r \to \infty)$ if for every P^*-integrable function f on $\mathcal{X}$,

$$\int f\phi_{(r)}dP^* \to \int f\phi_{(0)}dP^* \qquad (r \to \infty)\ . \tag{3.2}$$

Φ is proved to be compact for this topology as shown in the case of test functions (see Lehmann [7], Appendix). Also as for this topology, $\tau(\phi)-C$ is a continuous mapping from Φ into $\mathcal{Z}_1$, and $v(\phi, P)$ a real-valued lower semicontinuous function. Hence $\Phi(C)$ is the image of a compact set in $\mathcal{Z}_1$ by the continuous function, and so $\Phi(C)$ is also compact. Then there exists a ϕ_P in $\Phi(C)$ which attains the infimum of the lower semicontinuous function $v(\phi, P)$. In order to find the explicit form of ϕ_P, let us see the relation

$$\inf_{\phi\in\Phi(C)} v(\phi, P)=\inf_{\phi\in\Phi(C)} J(\phi, P)\ , \tag{3.3}$$

given by Lemma 3.1 which is stated after this theorem. Then applying Isii's theorem (see [5], Theorem 2.2) to this case, we obtain

$$\inf_{\phi\in\Phi(C)} J(\phi, P)=\sup_{T\in\mathcal{Z}_1^+}\ \inf_{\phi\in\Phi}\ \{J(\phi, P)+T[C-\tau(\phi)]\}\ , \tag{3.4}$$

where $\mathcal{Z}_1^+$ stands for a subset of the conjugate space $\mathcal{Z}_1^*$ of $\mathcal{Z}_1$ such that

$$\mathcal{Z}_1^+=\{T;\ T\in\mathcal{Z}_1^*,\ T(z)\geqq 0\quad\text{for } z\geqq 0\}\ . \tag{3.5}$$

Expanding the right-hand side of (3.4) by Lemma 3.1, we have

$$\begin{aligned}\inf_{\phi\in\Phi(C)} J(\phi, P)=\sup_{(b_{ij})\in\mathcal{Z}}\ \inf_{\phi\in\Phi}\Bigg[\int\sum_{j=1}^{l}\phi_j\Bigg\{\sum_{i=1}^{k}a_{ij}(P^*)g_i\frac{dP}{dP^*}\\ -\sum_{i=1}^{m}b_{ij}f_i\frac{d\nu}{dP^*}\Bigg\}dP^*+\sum_{i,j}b_{ij}C_{ij}\Bigg]\ ,\end{aligned} \tag{3.6}$$

where b_{ij} is the (i, j)th coefficient shown below, in the linear expression $T(z)=\sum_{i,j}b_{ij}z_{ij}$, and $a_{ij}(P^*)$ is the constant given by Lemma 3.1. Thus we obtain ϕ_P with the following expression; for $P\in\mathcal{P}_1$, there exist $a_{ij}(P^*)$ and $b_{ij}(P^*)$ except for a set of P^*-measure zero,

$$\phi_{P_j(x)}=\begin{cases}1 & \text{for } w_j[x;P, a_{ij}(P^*), b_{ij}(P^*)]\\ & \qquad >\sup_{\substack{1\leqq r\leqq l;\\ r\neq j}} w_r[x;P, a_{ij}(P^*), b_{ij}(P^*)]\\ 0 & \text{for } w_j[x;P, a_{ij}(P^*), b_{ij}(P^*)]\\ & \qquad <\sup_{\substack{1\leqq r\leqq l;\\ r\neq j}} w_r[x;P, a_{ij}(P^*), b_{ij}(P^*)]\end{cases}\qquad (j=1, 2,\cdots, l)\ , \tag{3.7}$$

where w_j is given so that

$$\begin{aligned}&w_j[x;P, a_{ij}(P^*), b_{ij}(P^*)]\\ &=\sum_{i=1}^{m}b_{ij}(P^*)f_i(x)\frac{d\nu}{dP^*}-\sum_{i=1}^{k}a_{ij}(P^*)g_i(x)\frac{dP}{dP^*}\qquad (j=1, 2,\cdots, l)\ ,\end{aligned} \tag{3.8}$$

and $a_{ij}(P^*)$ and $b_{ij}(P^*)$ depend only on P^*.

COROLLARY 3.1. *If h is continuous and linear on $\mathcal{Z}_1$, namely there exists a constant a_{ij} $(i=1,\cdots,k;\ j=1,\cdots,l)$ such that*

$$h(z)=\sum_{i,j} a_{ij}z_{ij} \qquad \text{for } z=(z_{ij}) \in \mathcal{Z}_1 , \tag{3.9}$$

$a_{ij}(P^*)$ in (3.8) can be replaced for this a_{ij} which does not depend on P^*.

Hence (3.7) with (3.8) becomes a necessary and sufficient condition that $\phi_P \in \Phi(C)$ be an optimum decomposition.

LEMMA 3.1. *If h is lower semicontinuous and concave on $\mathcal{Z}_1$, there exists $[a_{ij}(P)] \in \mathcal{Z}$ for a optimum decomposition $\phi_P \in \Phi(C)$ such that ϕ_P attains $\inf_{\phi \in \Phi(C)} J(\phi, P)$ for a fixed $P \in \mathcal{P}_1$, where*

$$J(\phi, P)=\sum_{i,j} a_{ij}(P)\psi_{ij}(\phi, P) \qquad \text{for } \phi \in \Phi,\ P \in \mathcal{P}_1 . \tag{3.10}$$

Further, if $\phi^0 \in \Phi(C)$ different from ϕ_P attains the infimum of $J(\phi, P)$ for a fixed $P \in \mathcal{P}_1$, ϕ^0 is also an optimum decomposition for this P.

4. Existence and expression of a Bayes estimator

If an a priori distribution ξ on $(\mathcal{P}_1, C)$ is known, where C is a given σ-field on $\mathcal{P}_1$, we expect to get a Bayes estimator $\hat{\phi}_\xi \in \hat{\Phi}(C)$ with respect to ξ which attains the infimum of the Bayes risk

$$r(\hat{\phi}, \xi)=\int_{\mathcal{P}_1} r(\hat{\phi}, P)d\xi , \tag{4.1}$$

under the restriction

$$\tau(\hat{\phi})\leqq C \qquad (\text{or}=C) . \tag{4.2}$$

In order to get such an estimator we make the following assumptions.

ASSUMPTION 1′. The σ-field $\mathfrak{A}$ over $\mathcal{X}$ and $\mathcal{B}$ over $\mathcal{S}$ each have a countable number of generators.

ASSUMPTION 2. The σ-field C over $\mathcal{P}_1$ has the following two properties:

1) $P(D)$ and $Q_P(E)$ are C-measurable as real-valued functions of P in $\mathcal{P}_1$ for any fixed set D in $\mathfrak{A}$ and E in $\mathcal{B}$, where $\mathcal{B}$ is a certain σ-field over $\mathcal{S}$.
2) Each component $\phi_{Pj}(x)$ of the optimum l-decomposition $\phi_P(x)$ is $\mathfrak{A}\times C$-measurable $(j=1,\cdots,l)$.

ASSUMPTION 3. The family $\mathcal{P}_1$ and $Q_1=\{Q_P; P\in\mathcal{P}_1\}$ are dominated respectively by certain nonnegative σ-finite measures μ and λ.

THEOREM 4.1 (*Existence of a Bayes estimator*). *Let ξ be an a priori probability measure over $(\mathcal{P}_1, C)$. If the function $h(z)$ is continuous on $\mathcal{Z}_1$ and $h\left[\int_E g'(x)\cdot dP(x)\right]$ is ξ-integrable as a function of P in $\mathcal{P}_1$ for any fixed set E in $\mathfrak{A}$, there exists a Bayes estimator $\hat{\phi}_\xi$ in $\hat{\phi}(C)$ under Assumptions* 1′, 2 *and* 3.

Especially, if $h(z)$ is continuous and linear on $\mathcal{Z}_1$ and $\nu=P$ in $\mathcal{P}_1$, the same assertion as above holds only under Assumptions 1′ *and* 2.

PROOF. In the expression

$$r(\hat{\phi}, \xi)=\int_{\mathcal{P}_1}\int_S v(\hat{\phi}, P)dQ_P d\xi-\int_{\mathcal{P}_1} v(\hat{\phi}_P, P)d\xi\ , \tag{4.3}$$

the second term of the right-hand side does not depend on $\hat{\phi}\in\hat{\Phi}(C)$, so it suffices to consider the infimum of the first one

$$\bar{u}(\hat{\phi})=\int_{\mathcal{P}_1}\int_S v(\hat{\phi}, P)dQ_P d\xi\ . \tag{4.4}$$

Further,

$$u(\hat{\phi}, P)=\int_S v(\hat{\phi}, P)dQ_P \tag{4.5}$$

is ξ-integrable by Assumption 2 and the conditions on h.

We can suppose that μ and λ in Assumption 3 are absolute continuous with respect to certain probability measures P_1 and Q_1, respectively. Now we define a new probability measure P^* on $(\mathcal{X}, \mathfrak{A})$ by

$$P^*(E)=[P(E)+P_0(E)]/2 \qquad \text{for } E\in\mathfrak{A}\ , \tag{4.6}$$

where P_0 is the probability measure defined in the proof of Theorem 3.1. Then on $\hat{\Phi}$, $P^*\times Q_1$-weak topology can be introduced just like as P^*-weak topology on Φ in the proof of Theorem 3.1. $u(\hat{\phi}, P)$ is continuous in $\hat{\phi}\in\hat{\Phi}$ for each fixed $P\in\mathcal{P}_1$ with respect to this $P^*\times Q_1$-weak topology. $\tau(\hat{\phi})$ is also continuous.

Since $\hat{\Phi}(C)$ is compact, there exists a $\hat{\phi}\in\hat{\Phi}(C)$ which attains

$$\Sigma(P)=\sup_{\hat{\phi}\in\Phi(C)}|u(\hat{\phi}, P)|\ . \tag{4.7}$$

Therefore $\Sigma(P)$ is also ξ-integrable. If $\hat{\phi}_{(r)}\to\hat{\phi}_{(0)}$ $(r\to\infty)$, it is seen by Lebesgue's convergence theorem that

$$\bar{u}(\hat{\phi}_{(r)}) \to \bar{u}(\hat{\phi}_{(0)}) \qquad (r \to \infty),$$

namely $\bar{u}(\hat{\phi})$ is continuous in $\hat{\phi} \in \hat{\Phi}(C)$. Hence we can see that there exists a $\hat{\phi}_\xi \in \hat{\Phi}(C)$ which attains $\inf_{\hat{\phi} \in \hat{\Phi}(C)} \bar{u}(\hat{\phi})$.

If, in particular, h is linear, then defining a probability measure $\bar{P}$ on $(\mathcal{X} \times \mathcal{S}, \mathfrak{A} \times \mathcal{B})$ by

$$\bar{P}(E) = \int_{\mathcal{P}_1} \left[\iint_E dP dQ_P \right] d\xi \qquad \text{for } E \in \mathfrak{A} \times \mathcal{B}, \tag{4.8}$$

and introducing $\bar{P}$-weak topology on $\hat{\Phi}$, we can show that $\bar{u}(\hat{\phi})$ and $\tau(\hat{\phi})$ are continuous with respect to this topology. So, it is easily seen that there exists a $\hat{\phi}_\xi \in \hat{\Phi}(C)$ in this case. Thus the proof is completed.

The following lemma is necessary to give an expression of a Bayes estimator.

Lemma 4.1. *Suppose that h is concave and continuously differentiable on $\mathcal{Z}_1$ so that $\frac{\partial h}{\partial z_{ij}} \left[\int_E g(x) dP \right]$ is ξ-integrable as a function of $P \in \mathcal{P}_1$ for each fixed $E \in \mathfrak{A}$, under Assumptions 1′, 2 and 3. Then for a Bayes estimator $\hat{\phi}_\xi \in \hat{\Phi}(C)$, there exists a C-measurable $a_{ij}^\xi(P)$ so that $\hat{\phi}_\xi$ attains $\inf_{\phi \in \Phi(C)} r_{J_\xi}(\hat{\phi}, P)$, where*

$$r_{J_\xi}(\hat{\phi}, P) = \int_{\mathcal{P}_1} \int_{\mathcal{S}} [J_\xi(\hat{\phi}, P) - J_\xi(\hat{\phi}_\xi, P)] dQ_P d\xi, \tag{4.9}$$

and

$$J_\xi(\hat{\phi}, P) = \sum_{i,j} a_{ij}^\xi(P) \psi_{ij}(\hat{\phi}, P). \tag{4.10}$$

Further, if another $\hat{\phi}^0 \in \hat{\Phi}(C)$ attains $\inf_{\hat{\phi} \in \hat{\Phi}(C)} r_{J_\xi}(\hat{\phi}, \xi)$, $\hat{\phi}^0$ is also a a Bayes estimator.

Theorem 4.2 (*Explicit expression of a Bayes estimator*). *Under Assumptions 1′, 2 and 3, the Bayes estimator $\hat{\phi}_\xi$ in $\phi(C)$ for ξ becomes explicitly expressible if the function $h(z)$ is concave and continuously differentiable with respect to each component z_{ij} of $z \in \mathcal{Z}_1$.*

Proof. At first we give a proof in the case where h is continuous and linear, namely

$$h(z) = \sum_{i,j} a_{ij} z_{ij} \qquad \text{for } z = (z_{ij}) \in \mathcal{Z}_1. \tag{4.11}$$

Then, introducing a P^*-weak topology on $\hat{\Phi}$ as is defined in the proof of Theorem 4.1, we can find a $\hat{\phi}_\xi \in \hat{\Phi}(C)$ just as $\hat{\phi}_P$ in Corollary 3.1.

Namely $\hat{\phi}_\xi \in \hat{\Phi}(C)$ is a Bayes estimator, if and only if there exists Q_1-integrable $\pi_{ij}(s)$ such that except for a set of $P_1^* \times Q_1$-measure zero,

$$(4.12) \qquad \hat{\phi}_{\xi j}(x, s) = \begin{cases} 1 & \text{for } \hat{w}_j(x, s; a_{ij}, \pi_{ij}(s)) > \sup\limits_{\substack{1 \le r \le l \\ r \ne j}} \hat{w}_r(x, s; a_{ij}, \pi_{ij}(s)) \\ 0 & \text{for } \hat{w}_j(x, s; a_{ij}, \pi_{ij}(s)) < \sup\limits_{\substack{1 \le r \le l \\ r \ne j}} \hat{w}_r(x, s; a_{ij}, \pi_{ij}(s)) \end{cases} \qquad (j=1, \cdots, l),$$

where

$$(4.13) \qquad \hat{w}_j(x, s; a_{ij}, \pi_{ij}(s)) = \sum_{i=1}^{m} \int_{\mathcal{P}_1} \pi_{ij}(s) f_i(x) \frac{d\nu}{dP_1^*} \frac{dQ_P}{dQ_1} d\xi(P) - \sum_{i=1}^{k} \int_{\mathcal{P}_1} a_{ij} g_i(x) \frac{dP}{dP_1^*} \frac{dQ_P}{dQ_1} d\xi(P) \qquad (j=1, \cdots, l).$$

If h is concave on $\mathcal{Z}_1$, by Lemma 4.1 the same method as in the linear case can be applied to finding $\hat{\phi}_\xi$ by using $r_{J_\xi}(\hat{\phi}, \xi)$ instead of $r(\hat{\phi}, \xi)$. However, in this case, a_{ij} must be replaced by $a_{ij}^\xi(P)$ which depends on $\hat{\phi}_\xi$, so (4.12) becomes only a necessary condition.

5. Existence and expression of a minimax estimator $\tilde{\phi}$

In general, it is uncertain whether we can expect to get an a priori probability measure ξ such as in Section 4. Therefore it may be considered, in our case, to find a minimax estimator $\tilde{\phi} \in \hat{\Phi}(C)$ which attains

$$(5.1) \qquad \inf_{\hat{\phi} \in \hat{\Phi}} \sup_{P \in \mathcal{P}_1} r(\hat{\phi}, P)$$

under the restriction

$$(5.2) \qquad \tau(\hat{\phi}) \le C \qquad (\text{or } = C).$$

In this section we show the existence of such a $\tilde{\phi} \in \hat{\Phi}(C)$ and its expression as the limit of a sequence of Bayes solutions for a priori probability measures selected suitably the supports of which are finite subsets in $\mathcal{P}_1$.

At first we show the following lemmas as extensions of Isii's theorem (see [6], Theorem 3) in a sense.

LEMMA 5.1. *Let $\mathcal{U}_0$, $\mathcal{V}_0$ be nonempty subsets in $\mathcal{U}$, $\mathcal{V}$, respectively, and let Ξ be the family of probability measures on $\mathcal{V}_0$ whose supports* *are finite subsets in* $\mathcal{V}_0$. *If* $K(u, v)$ *is a real-valued function defined on* $\mathcal{U}_0 \times v_0$ *and define*

$$K(u, \xi)=\int_{\mathcal{V}_0} K(u, v)d\xi \qquad \text{for } u \in \mathcal{U}_0,\ \xi \in \Xi_0 . \tag{5.3}$$

Then we have

$$\inf_{u \in \mathcal{U}_0} \sup_{v \in \mathcal{V}_0} K(u, v)= \inf_{u \in \mathcal{U}_0} \sup_{\xi \in \Xi_0} K(u, \xi) . \tag{5.4}$$

Here, if there exists $u_0 \in \mathcal{U}_0$ *which attains the infimum of the right-hand side, this* u_0 *also attains the infimum of the left.*

Further, when $\mathcal{U}$ *is a linear space,* $\mathcal{U}_0$ *a convex subset in* $\mathcal{U}$, *suppose that* $K(u, v)$ *satisfies the following conditions:*

1°) *A certain topology can be introduced into* $\mathcal{U}_0$ *such that* $\mathcal{U}_0$ *is compact and* $K(u, v)$ *is lower semicontinuous in* u *for each fixed* v *with respect to it.*

2°) $K(u, v)$ *is a convex function of* u *for each fixed* v.

Then we have

$$\inf_{u \in \mathcal{U}_0} \sup_{v \in \mathcal{V}_0} K(u, v)= \sup_{\xi \in \Xi_0} \inf_{u \in \mathcal{U}_0} K(u, \xi) . \tag{5.5}$$

Further unless the left-hand side of (5.5) *is* $\pm\infty$ *under* 1°), 2°), *there exists a minimax* $u_0 \in \mathcal{U}_0$ *which attains this infimum.*

LEMMA 5.2. *Let* $\mathcal{U}_0$, $\mathcal{V}_0$ *be the sets given in Lemma* 5.1 *and* $K(u, v)$ *a real-valued function on* $\mathcal{U}_0 \times \mathcal{V}_0$, *such that there exists a minimax* $u_0 \in \mathcal{U}_0$ *which attains* $\inf_{u \in \mathcal{U}_0} \sup_{v \in \mathcal{V}_0} K(u, v)$. *Further, suppose that a* σ*-field* $\mathcal{W}_0$ *attached to* $\mathcal{V}_0$ *and a family* Ξ *of probability measures on* $(\mathcal{V}_0, \mathcal{W}_0)$ *are given so that*

3°) *for each fixed* $u \in \mathcal{U}$

$$\sup_{v \in \mathcal{V}_0} K(u, v) \leqq \sup_{\xi \in \Xi} K(u, \xi) , \tag{5.6}$$

and 4°)

$$\inf_{u \in \mathcal{U}_0} \sup_{\xi \in \Xi} K(u, \xi)= \sup_{\xi \in \Xi} \inf_{u \in \mathcal{U}_0} K(u, \xi) , \tag{5.7}$$

where $K(u, \xi)$ *is defined by* (5.3). *Then there exists a countable subset* $\Xi_d \subset \Xi$ *such that the minimax* u_0 *is a Bayes solution in the wide sense for* Ξ_d.

Under the foregoing preparations, we show the existence of a minimax estimator $\hat{\phi} \in \hat{\Phi}$. Necessary assumptions are as follows.

ASSUMPTION 4. A certain topology can be introduced into $\mathcal{P}_1$ so that $\mathcal{P}_1$ is separable and $\int_S v(\hat{\phi}, P)dQ_P$ is a real-valued continuous function in $P \in \mathcal{P}_1$ for each fixed $\hat{\phi} \in \hat{\Phi}(C)$ with respect to that topology.

ASSUMPTION 4′. A certain pseudo-metric ρ can be introduced into $\mathcal{P}_1$ so that $\mathcal{P}_1$ is totally bounded, and the family $\left\{\int v(\hat{\phi}, P)dQ_P ; \hat{\phi} \in \hat{\Phi}(C)\right\}$ is equi-continuous in P with respect to ρ.

THEOREM 5.1 (*Existence of a minimax estimator*). *Assume that Assumptions* 1′ *and* 4 *in the case where h is lower semicontinuous and convex on $\mathcal{Z}_1$, or Assumptions* 1′ *and* 4′ *in the case where h is continuous and not convex, are satisfied. Let Ξ_1 be the family of all a priori probability measures whose supports are finite subsets in $\mathcal{P}_1$. Then the following assertions hold*;

1°)

$$\inf_{\hat{\phi} \in \hat{\Phi}(C)} \sup_{P \in \mathcal{P}_1} r(\hat{\phi}, P) = \inf_{\hat{\phi} \in \hat{\Phi}(C)} \sup_{\xi \in \Xi_1} r(\hat{\phi}, \xi) = \sup_{\xi \in \Xi_1} \inf_{\hat{\phi} \in \hat{\Phi}(C)} (\hat{\phi}, \xi) . \tag{5.8}$$

2°) *Especially if* $\inf_{\hat{\phi} \in \hat{\Phi}(C)} \sup_{P \in \mathcal{P}_1} r(\hat{\phi}, P) < \infty$, *there exists a minimax estimator $\tilde{\phi} \in \hat{\Phi}(C)$ which attains the left-hand side of* (5.8).

3°) *Selecting a countable subset $\Xi_d \subset \Xi_1$ the minimax estimator $\tilde{\phi}$ may be expressed as a Bayes solution in the wide sense for Ξ_d.*

PROOF. Let $\mathcal{P}_0$ be a family of a countable number of generators for $\mathcal{P}_1$ under Assumption 4 (or 4′). Then we have

$$\inf_{\hat{\phi} \in \hat{\Phi}(C)} \sup_{P \in \mathcal{P}_1} r(\hat{\phi}, P) = \inf_{\hat{\phi} \in \hat{\Phi}(C)} \sup_{P \in \mathcal{P}_0} r(\hat{\phi}, P) , \tag{5.9}$$

and $\hat{\phi}_0 \in \hat{\Phi}(C)$ which attains the infimum of the right-hand side of (5.9) attains the infimum of the left-hand side. Hence we can assume, that $\mathcal{P}_1$ itself is countable from the beginning. Then, since $\mathcal{P}_1$ and Q_1 must be dominated by certain probability measures P_1 and Q_1 respectively, the $P_1^* \times Q_1$-weak topology is introduced into $\hat{\Phi}$ so that $\hat{\Phi}(C)$ is compact with respect to this topology as shown in the proof of Theorem 4.1. Therefore, if h is convex and semicontinuous on $\mathcal{Z}_1$, taking u, v for $\hat{\phi}$, P and $K(u, v)$ for $r(\hat{\phi}, P)$ respectively, we can see that the conditions in Lemma 5.1 are satisfied, since $r(\hat{\phi}, P)$ is semicontinuous in $\hat{\phi}$ for each fixed P by Lebesgue-Fatou's lemma. Thus, applying this relation and Lemma

5.2, we can see easily that assertions of the theorem follow in the present case.

If h is not convex but continuous, under Assumption 4′, for any $\varepsilon>0$ there exist $\delta(\varepsilon)>0$ and a $\delta(\varepsilon)$-covering $\{\mathcal{P}_{\varepsilon,q};\ q=1, 2,\cdots, m(\varepsilon)\}$ such that $\{P_a, P_b \in \mathcal{P}_{\varepsilon,q}\ (q=1, 2,\cdots, m(\varepsilon))\}$ implies

$$(5.10) \qquad |u(\hat{\phi}, P_a)-u(\hat{\phi}, P_b)|\leqq\varepsilon \qquad \text{for all } \hat{\phi}\in\hat{\Phi}(C)\ ,$$

where

$$(5.11) \qquad u(\hat{\phi}, P)=\int_S v(\hat{\phi}, P)\frac{dQ_P}{dQ_1}dQ_1\ .$$

Since for each fixed $P\in\mathcal{P}_1$, we can take out a certain $\mathcal{P}_{\varepsilon,q}$ including P, there exists $r_0(\varepsilon)$ such that $r\geqq r_0(\varepsilon)$ implies

$$(5.12) \qquad \sup_{1\leqq q\leqq m(\delta)} |u(\hat{\phi}_{(r)}, P_{\varepsilon,q})-u(\hat{\phi}_{(0)}, P_{\varepsilon,q})|\leqq\varepsilon\ ,$$

if $\hat{\phi}_{(r)}\to\hat{\phi}_{(0)}$ $(r\to\infty)$ with respect to the $P_1^*\times Q_1$-weak topology. We can see (5.12) holds since h is continuous on $\mathcal{Z}_1$. By (5.10) and (5.12), we have

$$(5.13) \qquad \begin{aligned} &|u(\hat{\phi}_{(r)}, P)-u(\hat{\phi}_{(0)}, P)| \\ &\leqq|u(\hat{\phi}_{(r)}, P)-u(\hat{\phi}_{(r)}, P_{\varepsilon,q})|+|u(\hat{\phi}_{(r)}, P_{\varepsilon,q}) \\ &\qquad -u(\hat{\phi}_{(0)}, P_{\varepsilon,q})|+|u(\hat{\phi}_{(0)}, P_{\varepsilon,q})-u(\hat{\phi}_{(0)}, P)| \\ &\leqq 3\varepsilon\ . \end{aligned}$$

Therefore, the inequality

$$(5.14) \qquad |r(\hat{\phi}_{(r)}, P)-r(\hat{\phi}_{(0)}, P)|\leqq 3\varepsilon$$

holds for $r\geqq r_{(0)}(\varepsilon)$ and for any $P\in\mathcal{P}_1$. Thus we see that $\sup_{P\in\mathcal{P}_1} r(\hat{\phi}, P)$ is a continuous function of $\hat{\phi}\in\hat{\Phi}(C)$. Therefore, since $\hat{\Phi}(C)$ is compact, there exists a minimax estimator $\tilde{\phi}\in\hat{\Phi}(C)$ which attains $\inf_{\hat{\phi}\in\hat{\Phi}(C)} \sup_{P\in\mathcal{P}_1} r(\hat{\phi}, P)$.

In this case, Lemma 5.1 also implies the first equality in (5.8). As for the second equality in (5.8), it suffices to apply Wald's theorem (see [10], Theorem 2.2). Thus the proof of the theorem is completed.

COROLLARY 5.1 (*Expression of a minimax estimator*). *Under the assumptions in Theorem* 5.1, *the minimax estimator* $\tilde{\phi}$ *can be represented as the limit of a certain sequence of Bayes estimators* $\{\phi_{\xi_i}\}$ *in the sense of risk function, i.e.*,

$$(5.15) \qquad \sup_{P\in\mathcal{P}_1} r(\tilde{\phi}, P)=\lim_{i\to\infty} r(\hat{\phi}_{\xi_i}, \xi_i)\ ,$$

where the sequence $\{\xi_i\}$ *of* a priori *probability measures should be suitably* *chosen and each ξ_i has a finite support on $\mathcal{P}_1$.*

PROOF. Let Ξ_1 be the family of probability measures of which supports are finite subsets in $\mathcal{P}_1$. Then by Theorem 5.1 we have

$$\sup_{P\in\mathcal{P}_1} r(\tilde{\phi}, P)=\sup_{\xi\in\Xi_1}\inf_{\hat{\phi}\in\hat{\Phi}(C)} r(\hat{\phi}, \xi) = \sup_{\xi\in\Xi_1} r(\hat{\phi}_\xi, \xi) . \tag{5.16}$$

Hence we can select ξ_i's from Ξ_1 such that (5.15) holds.

Remark 5.1. It should be noted that Assumption 4 is weaker than Assumption 4′ but the latter is more easily applicable.

Further, in proving the existence of $\tilde{\phi}$ in $\hat{\Phi}(C)$ in Theorem 5.1, it suffices to assume that for any $\varepsilon>0$ there exist $\delta(\varepsilon)>0$ and a $\delta(\varepsilon)$-covering composed of a countable number of elements even though $\mathcal{P}_1$ is not totally bounded. Total boundedness of $\mathcal{P}_1$ is needed only in proving the second equality in (5.8) by Wald's theorem.

6. ε-approximations for a Bayes estimator and a minimax estimator

In Sections 4 and 5 we have studied existence and expressions of Bayes and minimax estimators. However, if the structure of h is complicated, it is not so easy to find explicit forms of those estimators in practice. In such a case, suppose that we could select a finite system $\mathcal{P}_0=\{P_1,\cdots, P_m\}$ from $\mathcal{P}_1$ as a representative of $\mathcal{P}_1$ such that Bayes estimators and minimax estimators under the assumption that the true P is in $\mathcal{P}_0$ instead of $\mathcal{P}_1$ are approximations of those under the assumption that the true P is in $\mathcal{P}_1$. Then the problem becomes rather easy because it suffices to treat only a finite number of objects if there is no necessity for respecting exactness.

In this section we show an ε-approximation method for those estimators by considering an ε-covering on $\mathcal{P}_1$ under Assumption 4′ used in proving Theorem 5.1.

Now let us put

$$u(\hat{\phi}, P)=\int_{\mathcal{S}} v(\hat{\phi}, P)dQ_P . \tag{6.1}$$

Then, under Assumption 4′, for any positive ε there exist $\delta(\varepsilon)>0$ and $\delta(\varepsilon)$-covering $\{\mathcal{P}_{\varepsilon,q}: q=1, 2,\cdots, m(\varepsilon)\}$ such that $P_a, P_b\in\mathcal{P}_{\varepsilon,q}$ implies

$$|u(\hat{\phi}, P_a)-u(\hat{\phi}, P_b)|\leqq\varepsilon \qquad \text{for all } \hat{\phi}\in\hat{\Phi}(C) \quad (q=1, 2,\cdots, m(\varepsilon)) . \tag{6.2}$$

Taking $P_{\varepsilon,q}$ from each $\mathcal{P}_{\varepsilon,q}$ respectively as a representative, we fix a $\delta(\varepsilon)$-net $\mathcal{P}_{\varepsilon}=\{P_{\varepsilon,q}: q=1, 2,\cdots, m(\varepsilon)\}$ throughout this section.

For a given a priori probability measure ξ on $(\mathcal{P}_1, C)$, we define a discrete a priori probability measure ξ_ε by

$$\xi_\varepsilon(P_{\varepsilon,q})=\xi(\mathcal{P}_{\varepsilon,q}) \qquad (q=1, 2,\cdots, m(\varepsilon)) , \tag{6.3}$$

which has masses only on $\mathcal{P}_\varepsilon$. Then we have the following theorem.

THEOREM 6.1 (*ε-approximation to a Bayes estimator*). *If h is continuous on $\mathcal{Z}_1$, under Assumptions* 1′, 2 *and* 4′, *there exists a Bayes estimator $\phi_{\xi_\varepsilon} \in \hat{\Phi}(C)$ for ξ_ε defined by* (6.3) *such that*

$$|r(\hat{\phi}_{\xi_\varepsilon}, \xi_\varepsilon)-r(\hat{\phi}_\xi, \xi)|\leqq\varepsilon , \tag{6.4}$$

i.e. $\hat{\phi}_{\xi_\varepsilon}$ is an ε-approximation to the Bayes estimator $\hat{\phi}_\xi$ for ξ.

PROOF. At first we define a probability measure P_ε on $(\mathcal{X}, \mathfrak{A})$ and Q_ε on $(\mathcal{S}, \mathcal{B})$ by

$$dP_\varepsilon=\frac{1}{m(\varepsilon)+1}\left[\sum_{q=1}^{m(\varepsilon)} dP_{\varepsilon,q}+dP_0\right] , \tag{6.5}$$

$$dQ_\varepsilon=\frac{1}{m(\varepsilon)} \sum_{q=1}^{m(\varepsilon)} dQ_{\varepsilon,q} , \tag{6.6}$$

where P_0 is defined as in the proof of Theorem 3.1. Then, introducing the $P_\varepsilon\times Q_q$-weak topology into $\hat{\Phi}$, we can show the existence of $\hat{\phi}_{\xi_\varepsilon} \in \hat{\Phi}(C)$ which attains $\inf_{\hat{\phi}\in\hat{\Phi}(C)} r(\hat{\phi}, \xi_\varepsilon)$ just as in the proof of Theorem 4.1. It follows from (6.2) that for every $P_{\varepsilon,q}$ $(q=1,\cdots, m(\varepsilon))$, we have

$$|r(\hat{\phi}, P_{\varepsilon,q})-r(\hat{\phi}, P)|\leqq\varepsilon \qquad \text{for all } \hat{\phi}\in\hat{\Phi}(C) , \qquad \text{for all } P\in\mathcal{P}_{\varepsilon,q} . \tag{6.7}$$

Therefore, we have

$$\begin{aligned} &|r(\hat{\phi}, \xi_\varepsilon)-r(\hat{\phi}, \xi)| \\ &\leqq \sum_{q=1}^{m(\varepsilon)} \int_{\mathcal{P}_{\varepsilon,q}} |r(\hat{\phi}, P_{\varepsilon,q})-r(\hat{\phi}, P)|\, d\xi \\ &\leqq\varepsilon , \qquad\qquad \text{for all } \hat{\phi}\in\hat{\Phi}(C) . \end{aligned} \tag{6.8}$$

(6.8) implies

$$|r(\hat{\phi}_\xi, \xi_\varepsilon)-r(\hat{\phi}_\xi, \xi)|\leqq\varepsilon , \tag{6.9}$$

$$|r(\hat{\phi}_{\xi_\varepsilon}, \xi_\varepsilon)-r(\hat{\phi}_{\xi_\varepsilon}, \xi)|\leqq\varepsilon . \tag{6.10}$$

Thus we can get (6.4) from (6.9) and (6.10).

THEOREM 6.2 (*ε-approximation to a minimax estimator*). *If h is continuous on $\mathscr{Z}_1$, under Assumptions 1′ and 4′ there exists a minimax estimator $\tilde{\phi}_\varepsilon$ in $\hat{\Phi}(C)$ for $\mathscr{P}_\varepsilon$ which attains $\inf_{\hat{\phi} \in \hat{\Phi}(C)} \sup_{1 \leq q \leq m(\varepsilon)} r(\hat{\phi}, P_{\varepsilon,q})$, such that the following properties are satisfied;*

1°) *Let Ξ_ε be the family of ξ_ε's which have masses only on $\mathscr{P}_\varepsilon$, and $\hat{\phi}_{\xi_\varepsilon}$ a Bayes solution for ξ_ε. Then we have*

$$\sup_{1 \leq q \leq m(\varepsilon)} r(\tilde{\phi}_\varepsilon, P_{\varepsilon,q}) = \sup_{\xi_\varepsilon \in \Xi_\varepsilon} r(\hat{\phi}_{\xi_\varepsilon}, \xi_\varepsilon) \tag{6.11}$$

2°) *For a minimax estimator $\tilde{\phi}$ as to $\mathscr{P}_1$, we have*

$$\left| \sup_{1 \leq q \leq m(\varepsilon)} r(\tilde{\phi}_\varepsilon, P_{\varepsilon,q}) - \sup_{P \in \mathscr{P}_1} r(\tilde{\phi}, P) \right| \leq \varepsilon , \tag{6.12}$$

i.e. $\tilde{\phi}_\varepsilon$ is an ε-approximation to $\tilde{\phi}$

PROOF. It is easily seen that there exists such a $\tilde{\phi}_\varepsilon \in \hat{\Phi}(C)$, when we notice that $\sup_{1 \leq q \leq m(\varepsilon)} r(\hat{\phi}, P_{\varepsilon,q})$ is a continuous function of $\hat{\phi} \in \hat{\Phi}(C)$ with respect to the $P_\varepsilon \times Q_1$-weak topology which is defined in the proof of Theorem 6.1.

Next Theorem 5.1 implies

$$\sup_{1 \leq q \leq m(\varepsilon)} r(\tilde{\phi}_\varepsilon, P_{\varepsilon,q}) = \sup_{\xi_\varepsilon \in \Xi_\varepsilon} \inf_{\hat{\phi} \in \hat{\Phi}(C)} r(\hat{\phi}, \xi_\varepsilon) \tag{6.13}$$

$$= \sup_{\xi_\varepsilon \in \Xi_\varepsilon} r(\hat{\phi}_{\xi_\varepsilon}, \xi_\varepsilon) , \tag{6.13′}$$

namely (6.11) holds. We have in the same way

$$\sup_{P \in \mathscr{P}_1} (\tilde{\phi}, P) = \sup_{\xi \in \Xi_1} r(\hat{\phi}_\xi, \xi) \tag{6.14}$$

where Ξ_1 is the family of a priori probability measures whose supports are finite subsets in $\mathscr{P}_1$. On the other hand, Theorem 6.1 implies

$$\sup_{\xi_\varepsilon \in \Xi_\varepsilon} r(\hat{\phi}_{\xi_\varepsilon}, \xi_\varepsilon) \leq \sup_{\xi \in \Xi_1} r(\hat{\phi}_\xi, \xi) \tag{6.15}$$

$$\leq \sup_{\xi_\varepsilon \in \Xi_\varepsilon} r(\hat{\phi}_{\xi_\varepsilon}, \xi_\varepsilon) + \varepsilon .$$

By (6.13′), (6.14) and (6.15) we see that (6.12) holds.

Remark 6.1. We assumed throughout this section that $\mathscr{P}_1$ is totally bounded in order to find ε-approximations for $\hat{\phi}_\xi$ and $\tilde{\phi}$. However the theorems in this section could be immediately extended to the case where the original family of probability measures $\mathscr{P}_1$ has an ε-covering with a countable number of elements.

7. Weak consistency of a Bayes estimator and a minimax estimator

In the foregoing sections we have studied problems of the Bayes estimator $\hat{\phi}_\xi$ and the minimax estimator $\tilde{\phi}$ where the size n of prior information S is fixed. In this section, we study the asymptotic properties, i.e. weak consistency, of those estimators in the sense of the risk function as stated below. To this end, we add subscript n to the notations related to the prior information such as $S^{(n)}$, $\mathcal{S}^{(n)}$, $\hat{\phi}^{(n)}$, $\hat{\Phi}^{(n)}(C)$ and so on.

Suppose now an a priori distribution ξ is given on $(\mathcal{P}_1, \mathcal{C})$. Then $\hat{\phi}^{(n)} \in \hat{\Phi}^{(n)}(C)$ is called weakly consistent for ξ with respect to the Bayes risk $r(\hat{\phi}, \xi)$, if

$$r(\hat{\phi}^{(n)}, \xi) \to 0 \qquad (n \to \infty) . \tag{7.1}$$

On the other hand, $\hat{\phi}^{(n)} \in \hat{\Phi}^{(n)}(C)$ is called weakly consistent on $\mathcal{P}_1$ with respect to the risk $r(\hat{\phi}^{(n)}, P)$, if

$$r(\hat{\phi}^{(n)}, P) \to 0 \quad (n \to \infty) \qquad \text{for each fixed } P \in \mathcal{P}_1 . \tag{7.2}$$

Especially if the convergence in (7.2) is uniform in $P \in \mathcal{P}_1$, $\phi^{(n)}$ is called uniformly weakly consistent on $\mathcal{P}_1$ with respect to the risk. Then the purpose of this section lies in proving the weak consistency of the Bayes estimator $\hat{\phi}_\xi^{(n)}$ and the minimax estimator $\tilde{\phi}^{(n)}$.

At first let us introduce some notations for convenience' sake. Suppose now a pseudo-metric d is introduced into $\mathcal{P}_1$, such that for any $\varepsilon > 0$ there exists an ε-covering on $\mathcal{P}_1$, composed of at most a countable number of elements $\{\mathcal{P}_{\varepsilon,q}\colon q=1, 2\cdots\}$. Then selecting a representative $P_{\varepsilon,q}$ from each $\mathcal{P}_{\varepsilon,q}$, we fix an ε-net $\mathcal{P}_\varepsilon = \{P_{\varepsilon,q}\colon q=1, 2, \cdots\}$. Corresponding to $\mathcal{P}_\varepsilon$ let us consider a decomposition $\chi_\varepsilon^{(n)}(s) = (\chi_{\varepsilon,1}^{(n)}(s), \chi_{\varepsilon,2}^{(n)}(s), \cdots)$ of $\mathcal{S}^{(n)}$ whose components are $\mathcal{B}$-measurable real-valued functions, such that

$$\chi_{\varepsilon,q}^{(n)}(s) \geqq 0 , \quad \sum_q \chi_{\varepsilon,q}^{(n)}(s) = 1 \qquad (q=1, 2, \cdots) . \tag{7.3}$$

Further, let us put

$$\beta_{\varepsilon,q}^{(n)}(P) = \int_S \chi_{\varepsilon,q}^{(n)}(s) dQ_P^{(n)} , \tag{7.4}$$

$$\breve{\phi}_\varepsilon^{(n)}(x, s) = \sum_q \phi_{P_{\varepsilon,q}}(x) \chi_{\varepsilon,q}^{(n)}(s) , \tag{7.5}$$

$$\phi_\varepsilon^{(n)}(x) = \sum_q \phi_{P_{\varepsilon,q}}(x) \beta_{\varepsilon,q}^{(n)} . \tag{7.6}$$

As for $\mathcal{P}_1$, $Q_1^{(n)}$, $\mathcal{S}^{(n)}$, we assume the following assumption.

ASSUMPTION 5. A pseudo-metric d may be introduced into $\mathcal{P}_1$ so that the following conditions are satisfied:

1°) For any $\varepsilon>0$ there exists an ε-covering $\{\mathcal{P}_{\varepsilon,q}: q=1, 2, \cdots\}$ composed of at most a countable number of elements.

2°) $\{v(\phi, P); \phi \in \Phi(C)\}$ is a family of equi-continuous functions of $P \in \mathcal{P}_1$.

3°) For any $\varepsilon'>0$ there exist $\varepsilon(\varepsilon')>0$ and a certain ε-net $\{P_{\varepsilon,q}: P_{\varepsilon,q} \in \mathcal{P}_{\varepsilon,q}, q=1, 2, \cdots\}$ such that selecting a suitable $\chi_\varepsilon^{(n)}(s)$ we can determine $n_0(\varepsilon')$ so that $n \geqq n_0(\varepsilon')$ implies

$$(7.7) \qquad \sum_{d(P_{\varepsilon,q}, P_{\varepsilon,r})>\rho\varepsilon} \beta_{\varepsilon,q}^{(n)} < \varepsilon' \qquad \text{for all } P \in \mathcal{P}_1 ,$$

where ρ is a finite constant >1 which does not depend on ε', and $P_{\varepsilon,r}$ is a representative such that $\mathcal{P}_{\varepsilon,r} \ni P$.

LEMMA 7.1. *If h is concave and continuously differentiable on $\mathcal{Z}_1$, under Assumptions 1 and 2, there exists a C-measurable $a_{ij}(P)$ $(i=1, \cdots, k: j=1, \cdots, l)$ such that for each fixed $\phi \in \Phi(C)$.*

$$(7.8) \qquad J(\phi, P)=\sum_{i,j} a_{ij}(P)\psi_{ij}(\phi, P)$$

is C-measurable in $P \in \mathcal{P}_1$ and for each fixed P

$$(7.9) \qquad r(\hat{\phi}, P) \leqq r_J(\hat{\phi}, P) \qquad \text{for all } \hat{\phi} \in \hat{\Phi}(C)$$

holds, where

$$(7.10) \qquad r_J(\hat{\phi}, P)=\int_{\mathcal{S}} [J(\hat{\phi}, P)-J(\phi_P, P)]dQ_P^{(n)} .$$

By this lemma, we can show $\breve{\phi}_\varepsilon^{(n)}$ has the uniformly weak consistency defined above:

THEOREM 7.1 (*Weak consistency of $\breve{\phi}_\varepsilon^{(n)}$*). *Assume that h is continuous or continuously differentiable according to whether it is convex or concave. Further, assume Assumptions 1′ and 5 specifying 1°) in 5 such that according to whether the numbers of elements in ε-coverings on $\mathcal{P}_1$ are finite (there $\mathcal{P}_1$ is totally bounded) or countable, $\psi(\phi, P)$ is a bounded function of P for each fixed $\phi \in \Phi(C)$ or a bounded function of ϕ and P. Then for any $\varepsilon'>0$, taking $\varepsilon(\varepsilon')>0$ and $n_0(\varepsilon')$ at 3°) in Assumption 5, we can determine a constant K which does not depend on ε' and $P \in \mathcal{P}_1$ such that $n \geqq n_0(\varepsilon')$ implies*

$$(7.11) \qquad r(\breve{\phi}_{\varepsilon(\varepsilon')}^{n}, P) \leqq K\varepsilon' \qquad \textit{for all } P \in \mathcal{P}_1 .$$

PROOF. By Assumption 5. 2°) for $\varepsilon'>0$ we can take $\varepsilon>0$ such that $d(P_a, P_b) \leqq \rho\varepsilon$ implies

$$(7.12) \qquad | v(\phi, P)-v(\phi, P) | \leqq \varepsilon' \qquad \text{for all } \phi \in \Phi(C) \quad (q=1, 2, \cdots) ,$$

where ρ is given by Assumption 5. 3°).

At first we prove the theorem in the case where h is convex. If h is convex, we have by (7.4) and (7.5)

$$(7.13)\qquad \begin{aligned} r(\breve{\phi}_{\varepsilon}^{(n)}, P) &= \int_{\mathcal{S}} v(\breve{\phi}_{\varepsilon}^{(n)}, P)dQ_P^{(n)} - v(\phi_P, P) \\ &\leqq \sum_q \int_{\mathcal{S}} v(\phi_{P_{\varepsilon,q}}, P)\chi_{\varepsilon,q}^{(n)} dQ_P^{(n)} - v(\phi_P, P) \\ &= \sum_{d(P_{\varepsilon,q}, P_{\varepsilon,r}) \leqq \rho\varepsilon} \beta_{\varepsilon,q}^{(n)}[v(\phi_{P_{\varepsilon,q}}, P) - v(\phi_P, P)] \\ &\quad + \sum_{d(P_{\varepsilon,q}, P_{\varepsilon,r}) > \rho\varepsilon} \beta_{\varepsilon,q}^{(n)}[v(\phi_{P_{\varepsilon,q}}, P) - v(\phi_P, P)] . \end{aligned}$$

As for the first term, by Assumption 5. 2°), we have

$$(7.14)\qquad \begin{aligned} & v(\phi_{P_{\varepsilon,q}}, P) - v(\phi_P, P) \\ & \leqq | v(\phi_{P_{\varepsilon,q}}, P) - v(\phi_{P_{\varepsilon,q}}, P_{\varepsilon,q}) | + | v(\phi_{P_{\varepsilon,q}}, P_{\varepsilon,q}) - v(\phi_P, P) | \\ & \leqq 2\varepsilon' + 2\varepsilon' = 4\varepsilon' . \end{aligned}$$

As for the second term, by the condition with respect to $\psi(\phi, P)$ and continuity of h on $\mathcal{Z}_1$, we see that there exists K' which does not depend on ε' and $P \in \mathcal{P}_1$ such that

$$(7.15)\qquad \sup_q | v(\phi_{P_{\varepsilon,q}}, P) - v(\phi_P, P) | < K' .$$

Hence we obtain

$$(7.16)\qquad r(\breve{\phi}_{\varepsilon}^{(n)}, P) \leqq 4\varepsilon' + K'\varepsilon' \qquad \text{for all } P \in \mathcal{P}_1 ,$$

which also implies (7.11).

The weak consistency of $\hat{\phi}_{\xi}^{(n)}$, $\tilde{\phi}^{(n)}$ follows from Theorem 7.1.

THEOREM 7.2 (*Weak consistency of* $\hat{\phi}_{\xi}^{(n)}$ *and* $\tilde{\phi}^{(n)}$). *Under the assumptions in Theorem* 7.1 *and Assumption* 2, *the Bayes estimator* $\hat{\phi}^{(n)}$ *has weak consistency for a given* a priori *probability measure* ξ *and also a minimax estimator* $\tilde{\phi}^{(n)}$ *has the uniformly weak consistency on* $\mathcal{P}_1$.

PROOF. For any $\varepsilon'>0$, we take $\varepsilon(\varepsilon')>0$, $n_0(\varepsilon')$ and $\breve{\phi}_{\varepsilon}^{(n)}$ such that (7.11) in Theorem 7.1 holds. Then $n \geqq n_0(\varepsilon')$ implies

$$(7.17)\qquad r(\hat{\phi}_{\xi}^{(n)}, \xi) \leqq r(\breve{\phi}_{\varepsilon}^{(n)}, \xi) = \int_{\mathcal{P}_1} r(\breve{\phi}_{\varepsilon}^{(n)}, P)d\xi \leqq K\varepsilon' .$$

This shows that $\hat{\phi}_{\xi}^{(n)}$ is weakly consistent.

As for a minimax estimator $\tilde{\phi}^{(n)}$, we also have

$$(7.18)\qquad \sup_{P \in \mathcal{P}_1} r(\tilde{\phi}^{(n)}, P) \leqq \sup_{P \in \mathcal{P}_1} r(\breve{\phi}^{(n)}, P) \leqq K\varepsilon'$$

by Theorem 7.1. This shows that $\tilde{\phi}^{(n)}$ is uniformly weakly consistent.

Remark 7.1. As for existence of a Bayes estimator $\hat{\phi}_{\xi}^{(n)}$, if we assume the assumptions in Theorem 7.1, integrability of $v(\hat{\phi}, P)$ with respect to $P \in \mathcal{P}_1$ which is a condition in the Theorem 4.1 follows from the condition on $\psi(\phi, P)$. So, we see that Theorem 4.1 holds.

Further, we can show existence of a minimax estimator $\tilde{\phi}^{(n)}$, if we add the following assumption 4°) to Assumption 5.
4°) Let the size n be fixed. Then, for any $\varepsilon > 0$, there exists $\delta(\varepsilon) > 0$ such that

$$(7.19) \qquad d(P_a, P_b) \leqq \delta(\varepsilon) \text{ implies } \int_E (dQ_{P_a}^{(n)} - dQ_{P_b}^{(n)}) < \varepsilon \qquad \text{for all } E \in \mathcal{B} ,$$

because we have, by the above,

$$(7.20) \qquad \left| \int v(\hat{\phi}, P_a) dQ_{P_a}^{(n)} - \int v(\hat{\phi}, P_b) dQ_{P_b}^{(n)} \right|$$
$$\leqq \int \left| v(\hat{\phi}, P_a) - v(\hat{\phi}, P_b) \right| dQ_{P_a}^{(n)} + \left| \int v(\hat{\phi}, P_b) [dQ_{P_a}^{(n)} - dQ_{P_b}^{(n)}] \right|$$
$$\text{for all } \hat{\phi} \in \hat{\Phi}(C) .$$

The first term can be sufficiently small by equicontinuity of $v(\hat{\phi}, P)$ (Assumption 5, 2°)), and the second by 4°) since $v(\hat{\phi}(x, s), P_b)$ is a bounded function of $s \in \mathcal{S}^{(n)}$ for each fixed $\hat{\phi} \in \hat{\Phi}^{(n)}(C)$ if h is continuous on $\mathcal{Z}_1$. Hence Assumption 5 with 4°) implies Assumption 4′ where $\left\{ \int v(\hat{\phi}, P) dQ_P^{(n)} : \hat{\phi} \in \hat{\Phi}(C) \right\}$ should be equi-continuous in $P \in \mathcal{P}_1$. So, we have Theorem 5.1.

8. Specializations to various problems

In this section we shall show how our theory can be applied to various statistical problems to search for the optimum regions—the optimum selection region, the statistical prediction region and the optimum stratification. But we confine ourselves to showing the possibility of applying our theory to them. Detailed discussions about them will be published in other papers in the near future.

8.1. *Optimum selection region*

We shall state the problem according to the formulation proposed by Cochran [1], though there have been given several different formulations proposed by other researchers until now.

Let $X = (X_1, X_2, \cdots, X_p)$ be p-variate vector random variable and its first component X_1 in R^1 be unobservable at the present time but observable at a certain time in the future. Let the remaining $(p-1)$ variables $X_2, \cdots, X_p$ in R^{p-1} be observable at the present time and be considered as the auxiliary variables by which the optimum region should be selected in the following way. Let us consider a family $\mathcal{F}$ of all possible cylinder sets parallel to the $\mathcal{X}_1$-axis with its base in R^{p-1} and a given size α $(0<\alpha<1)$, and we are supposed to obtain an optimum one among $\mathcal{F}$ in which the conditional expectation of X_1 should be maximized.

We can easily apply our theory to this problem as follows: let

$$\mathcal{X}=R^p, \qquad k=m=1, \qquad l=2,$$

$$f(x)=f_1(x)=1, \qquad g(x)=g_1(x)=x_1,$$

$$h(z)=h(z_{11}, z_{12})=-z_{11} \qquad \text{and} \qquad \phi(x)=(\phi_1(x), \phi_2(x))$$

satisfying the condition that $\phi(x)=\phi(x')$ for any pair of $x=(x_1, x_2, \cdots, x_p)$ and $x'=(x_2, x_3, \cdots, x_p)$. Then the restriction C in the optimum selection region may be expressed as

$$\int_{R^p} \phi_1(x)dP(x)=\alpha . \tag{8.1}$$

Further, let $\psi(\phi, P)=(\psi_1(\phi, P), \psi_2(\phi, P))$, where

$$\psi_{ij}(\phi, P)=\int_{\mathcal{X}} g_i(x)\phi_j(x)dP(x) \qquad (j=1, 2) .$$

Then the conditional expectation of X_1 under a given selection function ϕ may be expressed as

$$\begin{aligned} \mathrm{E}\{X_1 \mid \phi\} &= \frac{1}{\alpha}\int_{\mathcal{X}} h[\psi_{11}(\phi, P), \psi_{12}(\phi, P)]\,dP(x) \\ &= \frac{-1}{\alpha}\int_{\mathcal{X}} x_1\phi_1(x)dP(x) , \end{aligned} \tag{8.2}$$

which should be minimized. Since $-1/\alpha$ is a constant, it is easily seen that our theory can be applied to this problem.

8.2. *Statistical prediction region*

Let $X=(X_1, \cdots, X_p)$ be a p-dimensional vector random variable consisting of two vectors $\xi=(X_1, \cdots, X_q)$ and $\eta=(X_{q+1}, \cdots, X_p)$ $(q<p)$, where ξ is observable at the present time but η is not, it will be observable only at a certain time point in the future. We are supposed to obtain an optimum prediction region $D(\xi)$ in R^{p-q} based upon the observation on ξ such that the expected probability content of $D(\xi)$ should be maximized under the restriction C that the expected volume of $D(\xi)$ be equal to a given constant C.

Our theory may be applied to this problem in the following way: $D(\xi)$ may be expressed by $\phi(\xi, \eta)=\phi(x)$ and let

$$\mathcal{X}=R^p, \quad k=m=1, \quad l=2,$$
$$f(x)=f_1(x)\equiv 1, \quad g(x)=g_1(x)\equiv 1,$$
$$\text{and} \quad h(z)=h(z_{11}, z_{12})=z_{11},$$

and $d\nu(x)=d\nu(\xi, \eta)=dP'(\xi)\cdot d\nu_0(\eta)$ where P' is the marginal probability distribution of ξ and ν_0 is Lebesgue measure in R^{p-q}. Then the restriction C may be expressed as

$$\int_{\mathcal{X}} \phi_1(x)d\nu(x)=C, \quad (C>0) \tag{8.3}$$

and the expected probability content corresponding to a prediction function ϕ_1

$$\nu(\phi, P)=\int_{\mathcal{X}} \phi_1(x)dP(x) \tag{8.4}$$

should be maximized under C.

Remark 8.1. 1) We may modify the original problem in the following way: under the restriction that the expected probability content of $D(\xi)$ be equal to a given constant β $(0<\beta<1)$, the expected volume of $D(\xi)$ should be minimized.

2) We may modify the original problem in another way: under the restriction that the volume of $D(\xi)$ in R^{p-q} be kept as constant for each ξ, the expected probability content of $D(\xi)$ should be maximized.

In this case some modifications must be done in our theory (see Ishii [4]).

8.3. *Optimum stratification*

Let P be a probability measure in $\mathcal{P}_1$ corresponding to a distribution function $F(x)$ (in $\mathcal{F}_1$) of univariate random variable X, and we are supposed to obtain the optimum stratifications ϕ_p which minimizes the variance $V(\bar{X}, \phi, P)$ of the stratified estimator $\bar{X}$ for the population mean μ among a set Φ of all possible stratifications, where $\phi(x)=(\phi_1(x), \cdots, \phi_l(x))$,

$$\bar{X}=\sum_{i=1}^{l} w_i\bar{X}_i, \quad \bar{X}_i=\frac{1}{n_i}\sum_{i=1}^{l} X_{ij},$$
$$w_i=\int_{\mathcal{X}} \phi_i(x)dP(x), \tag{8.5}$$
$$\text{and} \quad \mathrm{V}(\bar{X}\,|\,\phi)=\frac{1}{n}\sum_{i=1}^{l} w_i\sigma_i^2$$

in the proportionate allocation case ($n_i = w_i n$).

Our theory may be applied to this problem in the following way:

$$k=2\,, \qquad m=0 \qquad \text{(i.e. no restriction)},$$

$$l=l\,, \qquad g(x)=(g_1(x), g_2(x))\,,$$

$$g_1(x)\equiv 1\,, \qquad g_2(x)=x\,,$$

and

$$h(z)=-\sum_{j=1}^{l}\frac{z_{2j}^2}{z_{1j}}\,.$$

Then it is easily seen that the relation

$$\text{(8.6)} \qquad V(\bar{X}\,|\,\phi, P)=\int_{R^1} x^2 dP(x)+h[\psi(\phi, P)]$$

holds. So, we are supposed to study on the optimum stratification ϕ_p for P which minimizes the function $h[\psi(\phi, P)]$. (See Dalenius [2], Taga [8], and Ishii and Taga [9].)

9. Appendix

In this section we give proofs of all lemmas used in proving the theorems in the previous sections.

9.1. *Expression of a decision function* $\check{\phi}$

In Section 2 a decision function $\check{\phi}$ from $(\mathcal{S}, \mathcal{B})$ into Φ is defined as a separably-valued and weak $\mathcal{B}$-measurable mapping. This $\check{\phi}$ can be expressed as $\hat{\phi}(x, s)$ ((2.6)) and we have expanded the whole theory using this relation. We now show the reasoning in the following proposition.

PROPOSITION 9.1. *Let $\mathcal{B}(\mathcal{X}, \mathfrak{A})$ be the space of all bounded $\mathcal{B}$-measurable functions. We introduce a topology into the product space $[\mathcal{B}(\mathcal{X}, \mathfrak{A})]^l$ by a norm such that for each $\nu=(\nu_1, \cdots, \nu_l)\in[\mathcal{B}(\mathcal{X}, \mathfrak{A})]^l$*

$$\text{(9.1)} \qquad \|\nu\|=\left[\sum_{j=1}^{l}\|\nu_j\|^2\right]^{1/2},$$

where

$$\text{(9.2)} \qquad \|\nu_j\|=\sup_{x\in\mathcal{X}}|\nu_j(x)|\,.$$

Φ is a subspace of $[\mathcal{B}(\mathcal{X}, \mathfrak{A})]^l$. Next, let $\check{\Phi}$ be the space of all separably-valued and weak $\mathcal{B}$-measurable mappings $\check{\phi}$'s from $\mathcal{S}$ into Φ with respect to the norm stated above.

Then for each $\check{\phi} \in \check{\Phi}$ *there exists a* $\hat{\phi} \in \hat{\Phi}$ *such that*

1°)

$$\hat{\phi}(\cdot, s)=\check{\phi}(s) \qquad \text{for } Q_p\text{-almost all } s \in \mathcal{S} \ (Q_p \in \mathcal{Q}_1). \tag{9.3}$$

2°) $\int_{\mathcal{S}} \hat{\phi}(x, s)dQ_p(s)$ *is equal to* $\int_{\mathcal{S}} \check{\phi}(s)dQ_p(s)$ *as a function of* $x \in \mathcal{X}$.

Conversely, for each $\hat{\phi} \in \hat{\Phi}$, *if we put* $\check{\phi}(s)=\hat{\phi}(\cdot, s)$ *for* Q_p*-almost all* $s \in \mathcal{S}$ $(Q_p \in \mathcal{Q}_1)$, $\check{\phi}$ *becomes a weak* $\mathcal{B}$*-measurable mapping* $\mathcal{S}$ *into* Φ.

PROOF. Now $\check{\phi}(s) \in \check{\Phi}$ is Bochner integrable with respect to $Q_p \in \mathcal{Q}_1$ because of the weak $\mathcal{B}$-measurability, separability of $\check{\phi}$ and uniform boundedness of Φ (see K. Yoshida [11], V. 5). Therefore, we can find $\check{\phi} \in \check{\Phi}$ with the property 1°), 2°) applying the theorem in Dunford-Shwartz' [3] (Lemma III, 16). The inverse also follows from the above.

9.2. *Proofs of Lemmas* 3.1, 4.1 *and* 7.1

We show the following two propositions without proofs which may be seen easily.

PROPOSITION 9.2. *Let* D *be a convex set in the* p*-dimensional real space* R^p, *and let* $f(x)$ *be a concave real function on* D. *Then for each fixed* $x_0=(x_{01}, \cdots, x_{0p}) \in D$, *there exists* $a(x_0)=(a_1(x_0), \cdots, a_p(x_0))$ *such that*

$$f(x)-f(x_0) \leqq \sum_{i=1}^{p} a_i(x_0)(x_i-x_{0i}) \qquad \text{for all } x=(x_1, \cdots, x_p) \in D. \tag{9.4}$$

PROPOSITION 9.3. *Under the assumption in Proposition* 9.2, *if there exists* $x_0 \in D$ *which attains* $\inf_{x \in D} f(x)$, x_0 *also attains the infimum of*

$$g(x)=\sum_{i=1}^{p} a_i(x_0)x_i \qquad \text{for } x=(x_1, \cdots, x_p) \in D. \tag{9.5}$$

Conversely if another $x' \in D$ *attains* $\inf_{x \in D} g(x)$, x' *also attains* $\inf_{x \in D} f(x)$.

PROOF OF LEMMA 3.1. Let $\mathcal{Z}_p$ be

$$\mathcal{Z}_p=\{\psi(\phi, P); \phi \in \Phi(C)\} \qquad \text{for each fixed } P \in \mathcal{P}_1. \tag{9.6}$$

Then since $\mathcal{Z}_p$ is the image of $\Phi(C)$ by the continuous function $\psi(\cdot, P)$, $\mathcal{Z}_p$ is also compact in $\mathcal{Z}_1$. Therefore, there exists $z_p \in \mathcal{Z}_p$ such that z_p attains the infimum of the lower semi-continuous function $h(z)$. Then ϕ_p is an inverse image of z_p by ψ, so ϕ_p attains $\inf_{\phi \in \Phi(C)} a(\phi, P)$ for each fixed $P \in \mathcal{P}_1$. Hence, if we take a_{ij} in Proposition 9.3 as for this z_p, we can see that Lemma 3.1 follows from Proposition 9.2.

PROOF OF LEMMA 7.1. For the partial derivatives $(\partial h/\partial z_{ij})$'s,

$$I(z)=\sum_{i,j}\frac{\partial h}{\partial z_{ij}}(z) \qquad \text{for } z \in \mathcal{Z}_1 \tag{9.7}$$

is also continuous on $\mathcal{Z}_1$. Now for each fixed $P \in \mathcal{P}_1$, let $\mathcal{Z}_p^*$ be the set of all z_p's which attain $\inf_{z \in \mathcal{Z}_p} h(z)$ where $\mathcal{Z}_p$ is defined in the proof of Lemma 3.1. Then since $\mathcal{Z}_p^*$ is also compact, there exists z_p^* which attains $\inf_{z \in \mathcal{Z}_p^*} I(z)$. Thus taking $(\partial h/\partial z_{ij})(z_p^*)$ as $a_{ij}(P)$, we can define a function

$$J(\phi, P)=\sum_{i,j}\frac{\partial h}{\partial z_{ij}}[\psi(\phi_p^*, P)]\psi_{ij}(\phi, P)\ , \tag{9.8}$$

where ϕ_p^* is an inverse image of z_p^* by $\psi(\phi, P)$. $J(\phi, P)$ is C-measurable for each fixed $\phi \in \Phi(C)$ under Assumption 2.

Now we have

$$r_J(\hat{\phi}, P)=\int_{\mathcal{S}}[J(\hat{\phi}, P)-J(\phi_p^*, P)]dQ_p\ , \tag{9.9}$$

and for each fixed $P \in \mathcal{P}_1$ we obtain

$$r(\hat{\phi}, P)\leqq r_J(\hat{\phi}, P) \qquad \text{for all } \hat{\phi} \in \hat{\Phi}(C)\ , \tag{9.10}$$

by Proposition 9.2.

PROOF OF LEMMA 4.1. For each fixed $P \in \mathcal{P}_1$, let $\mathcal{Z}_p^\xi$ be the set of all $\int_{\mathcal{S}}\psi(\hat{\phi}_\xi, P)dQ_p$'s which are images of Bayes estimator $\hat{\phi}_\xi$'s by the continuous function $\int_{\mathcal{S}}\psi(\cdot, P)dQ_p$. Since $\mathcal{Z}_p^*$ is also compact, there exists a $z_p^\xi \in \mathcal{Z}_p^\xi$ which attains the infimum $I(z)$ on $\mathcal{Z}_p^\xi$, where $I(z)$ is given by (9.10). So, taking $(\partial h/\partial z_{ij})(z_p^\xi)=a_{ij}^\xi(P)$, and

$$J_\xi(\hat{\phi}, P)=\sum_{i,j} a_{ij}^\xi(P)\psi_{ij}(\hat{\phi}, P)\ , \tag{9.11}$$

we see that $J_\xi(\hat{\phi}, P)$ is C-measurable for each fixed $\hat{\phi} \in \hat{\Phi}(C)$. Therefore, we can define

$$r_{J_\xi}(\hat{\phi}, \xi)=\int_{\mathcal{P}_1}\int_{\mathcal{S}}\sum_{i,j} a_{ij}^\xi(P)[\psi_{ij}(\hat{\phi}, P)-\psi_{ij}(\hat{\phi}_\xi, P)]dQ_p d\xi(P)\ . \tag{9.12}$$

Now we have, by Proposition 9.2,

$$\begin{aligned}&\int_{\mathcal{S}} h[\psi(\hat{\phi}, P)]dQ_p-\int_{\mathcal{S}} h[\psi(\hat{\phi}_\xi, P)]dQ_p \\ &\leqq \int_{\mathcal{S}}\sum_{i,j} a_{ij}^\xi(P)\psi_{ij}(\hat{\phi}, P)dQ_p-\int_{\mathcal{S}}\sum_{i,j} a_{ij}^\xi(P)\psi_{ij}(\hat{\phi}_\xi, P)dQ_p \\ &\qquad\qquad \text{for all } \hat{\phi} \in \hat{\Phi}(C)\ .\end{aligned} \tag{9.13}$$

This (9.13) implies

$$r(\hat{\phi}, \xi)-r(\hat{\phi}_\xi, \xi)\leqq r_{J_\xi}(\hat{\phi}, \xi) \qquad \text{for all } \hat{\phi}\in\hat{\Phi}(C) . \tag{9.14}$$

Therefore, $\hat{\phi}_\xi$ attains the infimum of $r_{J_\xi}(\hat{\phi}, \xi)$ on $\hat{\Phi}(C)$. Further, if another $\hat{\phi}'_\xi\in\Phi(C)$ attains this infimum, $\hat{\phi}'_\xi$ also becomes a Bayes estimator.

9.3. *Proofs of Lemmas* 5.1 *and* 5.2

PROOF OF LEMMA 5.1. Let $\tilde{\Xi}_0$ be the linear space generated by Ξ_0 over the real field. Then Ξ_0 is a convex subset in $\tilde{\Xi}_0$. $K(u, \xi)$ is a real-valued function on $\mathcal{U}_0\times\Xi_0$ which is linear in $\xi\in\Xi_0$ for each fixed $u\in\mathcal{U}_0$. Now let $\xi_v\in\Xi_0$ be a measure which has mass 1 on $v\in\mathcal{V}_0$. Then if $\xi\in\Xi_0$ has mass $\beta_1, \beta_2,\cdots,\beta_r$ on $v_1, v_2,\cdots, v_r\in\mathcal{V}_0$ respectively, ξ is expressed as

$$\xi=\sum_{i=1}^{r}\beta_i\xi v_i , \tag{9.15}$$

where

$$\sum_{i=1}^{r}\beta_i=1 , \qquad \beta_i\geqq 0 \quad (i=1,\cdots, r) . \tag{9.16}$$

Therefore, for each fixed $u\in\mathcal{U}_0$, we have

$$\begin{aligned}\sup_{\xi\in\Xi_0} K(u, \xi)&=\sup_{\substack{r;\beta_1,\cdots,\beta_r;\\ v_1,\cdots,v_r}} K\left(u, \sum_{i=1}^{r}\beta_i\xi v_i\right)\\ &\leqq \sup_{r;\beta_1,\cdots,\beta_r}\sum_{i=1}^{r}\beta_i \sup_{v_i\in\mathcal{V}_0} K(u, \xi v_i)\\ &= \sup_{v\in\mathcal{V}_0} K(u, \xi v)\\ &= \sup_{v\in\mathcal{V}_0} K(u, v) .\end{aligned} \tag{9.17}$$

Since it holds in general that

$$\sup_{\xi\in\Xi_0} K(u, \xi)\geqq \sup_{v\in\mathcal{V}_0} K(u, v) , \tag{9.18}$$

(9.17) implies

$$\sup_{v\in\mathcal{V}_0} K(u, v)= \sup_{\xi\in\Xi_0} K(u, \xi) \qquad \text{for each fixed } u\in\mathcal{U}_0 . \tag{9.19}$$

Hence, we see that (5.4) holds.

Next, if $K(u, v)$ has the conditions 1°) and 2°), $K(u, \xi)$ also has the same. So, we can apply Isii's theorem (see [6], Theorem 3) to our case. Hence we have

$$\inf_{u\in\mathcal{U}_0}\sup_{\xi\in\Xi_0} K(u, \xi)= \sup_{\xi\in\Xi_0}\inf_{u\in\mathcal{U}_0} K(u, v) . \tag{9.20}$$

Thus (5.5) follows from (9.20) and (5.4). Therefore, if $\inf_{u \in \mathcal{U}_0} \sup_{v \in \mathcal{V}_0} K(u, v) \neq \pm\infty$, both sides of (9.20) are unequal to $\pm\infty$. Hence there exists $u_0 \in \mathcal{U}_0$ which attains the infimum of the left-hand side of (9.20). However, since (9.19) and (5.4) imply

$$(9.21)\qquad \begin{aligned} \sup_{v \in \mathcal{V}_0} K(u_0, v) &= \sup_{\xi \in \Xi_0} K(u_0, \xi) \\ &= \inf_{u \in \mathcal{U}_0} \sup_{\xi \in \Xi_0} K(u, \xi) \\ &= \inf_{u \in \mathcal{U}_0} \sup_{v \in \mathcal{V}_0} K(u, v) , \end{aligned}$$

this u_0 also attains $\inf_{u \in \mathcal{U}_0} \sup_{v \in \mathcal{V}_0} K(u, v)$.

PROOF OF LEMMA 5.2. Since $\left\{ \inf_{u \in \mathcal{U}_0} K(u, \xi); \xi \in \Xi \right\}$ is a subset in the real space, $\Xi_d = \{\xi_i; \xi_i \in \Xi\ (i=1, 2, \cdots)\}$ can be selected so that

$$(9.22)\qquad \lim_{i\to\infty} \inf_{u \in \mathcal{U}_0} K(u, \xi_i) = \sup_{\xi \in \Xi} \inf_{u \in \mathcal{U}_0} K(u, \xi) .$$

Then the conditions 3°) and 4°) imply

$$(9.23)\qquad \begin{aligned} \sup_{v \in \mathcal{V}_0} K(u_0, v) &\leqq \inf_{u \in \mathcal{U}_0} \sup_{\xi \in \Xi} K(u, \xi) \\ &= \sup_{\xi \in \Xi} \inf_{u \in \mathcal{U}_0} K(u, \xi) \\ &= \lim_{i\to\infty} \inf_{u \in \mathcal{U}_0} K(u, \xi_i) . \end{aligned}$$

On the other hand, we have, for any $\xi_i \in \Xi_d$,

$$(9.24)\qquad K(u_0, \xi_i) \leqq \sup_{v \in \mathcal{V}_0} K(u_0, v) .$$

Therefore, (9.23) implies that for any $\xi_i \in \Xi_d$,

$$(9.25)\qquad K(u_0, \xi_i) \leqq \lim_{i\to\infty} \inf_{u \in \mathcal{U}_0} K(u, \xi_i) .$$

Hence we have

$$(9.26)\qquad \lim_{i\to\infty} \left[K(u_0, \xi_i) - \inf_{u \in \mathcal{U}_0} K(u, \xi_i) \right] = 0 .$$

This shows that u_0 is a Bayes solution in the wide sense for Ξ_d.

Acknowledgement

We should like to express our hearty thanks to Professor K. Isii and Mr. T. Kamae, Osaka University, for their kind discussions and advices.

THE INSTITUTE OF STATISTICAL MATHEMATICS

REFERENCES

[1] Cochran, W. G. (1951). Improvement by means of selection, *Proc. 2nd Berkeley Symposium on Math. Statist. and Prob.*
[2] Dalenius, T. (1950-51). The problem of optimum stratification, *Skand. Aktuarietidskr.*, **3**, 203-214 and **4**, 133-148.
[3] Dunford, N. and Schwartz, J. T. (1963). *Linear Operators* I, Interscience Publishers, New York.
[4] Ishii, G. (1968). Minimax invariant prediction regions, *Ann. Inst. Statist. Math.*, **20**, 33-53.
[5] Isii, K. (1964). Inequalities of the types of Chebyshev and Cramer-Rao and mathematical programming, *Ann. Inst. Statist. Math.*, **16**, 277-293.
[6] Isii, K. (1967). Lagrange multiplyer method and admissibility, *Surikagaku-Kokyuroku*, **27** (in Japanese).
[7] Lehmann, E. L. (1959). *Testing Statistical Hypotheses*, John Wiley.
[8] Taga, Y. (1967). On optimum stratification for the objective variable based on concomitant variables using prior information, *Ann. Inst. Statist. Math.*, **19**, 101-129.
[9] Isii, K. and Taga, Y. (1969). On optimal stratifications for multivariate distributions, *Skand. Aktuarietidskr.*, 24-38.
[10] Wald, A. (1950). *Statistical Decision Function*, John Wiley and Sons, Inc., New York.
[11] Yoshida, K. (1965). *Functional Analysis*, Springer.