# ON SEQUENTIAL DECISION PROBLEMS WITH DELAYED OBSERVATIONS

YUKIO SUZUKI

(Received Aug. 15, 1966)

## 0. Introduction

General sequential or non-sequential decision problems were formulated and deeply investigated by A. Wald [5]. But, he did not consider the time lag between taking a sample from a population under consideration and observing it. Any decision problem, however, will be formulated as a decision problem without the time lag of this kind, if no intermediate action by the decision maker is not considered between taking and observing a sample.

Recently, T. W. Anderson [1] pointed out that in some actual sequential decision problems the time lag between taking a sample and its observation cannot be neglected and thus delayed observations should be taken into account for the choice of an optimal decision procedure. He also mentioned the difficulty of treating such a sequential decision problem with delayed observations and, instead of attacking the problem theoretically, he evaluated the effect of neglecting the delayed observations.

The main purpose of the present paper is to characterize Bayes solutions of general and special sequential decision problems with delayed observations. In section 1, we shall formulate general truncated sequential decision problems with delayed observations. In section 2 general problems are specified by imposing restrictions on the cost of observations and probabilistic properties of observations. Further, in section 3 the parameter and decision spaces are restricted to reduce the problems to dichotomous sequential decision problems with delayed observations. In section 4 theorems and lemmas concerning the discussion in section 3 are established. Finally, in section 5, we shall deal with the problem which was originally treated by T. W. Anderson [2].

## 1. General truncated sequential decision problems with delayed observations

In this section the truncated sequential problem with delayed observations is given in a general form and a constructive characterization of the optimal decision procedure is presented.

Notations and definitions:

(i) Elementary sample space: $(\mathcal{Z}, \mathcal{L}, P'_\theta)$, $\theta \in \Theta$

$\mathcal{Z}$: a space of elements $z$; $\mathcal{L}$: a $\sigma$-field of subsets of $\mathcal{Z}$; $P'_\theta$: the probability measure defined on the measurable space $\mathcal{Z}(\mathcal{L})$ for each parameter value $\theta \in \Theta$; $\Theta$: the space of parameters $\theta$ which is a measurable space with a $\sigma$-field $\mathcal{F}$ of subsets of $\Theta$.

(ii) Sample space: $(\mathfrak{X}, \mathcal{B}, P_\theta)$, $\theta \in \Theta$

For given positive integers $N$ and $m$, let $\{g_i(z),\ i=1, 2, \cdots, N+m\}$ be a sequence of measurable functions defined on $\mathcal{Z}(\mathcal{L})$ such that $g_i(z)$ gives a measurable mapping from the measurable space $\mathcal{Z}(\mathcal{L})$ into a measurable space $\mathfrak{X}_i(\mathcal{B}_i)$ for $i=1, 2, \cdots, N+m$, where $\mathcal{B}_i$ is a $\sigma$-field of subsets of the space $\mathfrak{X}_i$. We now define the space $\mathfrak{X}$ as the product space $\mathfrak{X}_1 \times \cdots \times \mathfrak{X}_{N+m}$ and $\mathcal{B}$ as the product $\sigma$-field generated by $\{\mathcal{B}_i, i=1, \cdots, N+m\}$ which is denoted by $\mathcal{B}=\boldsymbol{B}(\mathcal{B}_1 \times \cdots \times \mathcal{B}_{N+m})$. Then, the vector-valued function $g(z)=(g_1(z), \cdots, g_{N+m}(z))$ is a measurable function from $\mathcal{Z}(\mathcal{L})$ into $\mathfrak{X}(\mathcal{B})$; the probability measure on $\mathfrak{X}(\mathcal{B})$ induced by this function is denoted by $P_\theta$ for each $\theta \in \Theta$.

ASSUMPTION 1.1. For each $\theta \in \Theta$, the probability measure $P_\theta$ is absolutely continuous with respect to a $\sigma$-finite measure $\mu$ on $\mathfrak{X}(\mathcal{B})$; $\frac{dP_\theta}{d\mu}=p_\theta$.

(iii) Cost function: $C(j, x)$

Denote by $J$ the set $\{0, 1, 2, \cdots, N+m\}$. The cost function $C(j, x)$ is defined on $J \times \mathfrak{X}$ and is a bounded non-negative real-valued measurable function of $x$ for each $j \in J$ such that, if $\mathfrak{X} \ni x, y$ and $x_i=y_i$ $i=1, \cdots, j$, then $C(j, x)=C(j, y)$. We also assume that $C(j, x) \geqq C(j', x)$ for $j>j'$.

(iv) The space of terminal actions: $A \ni a$

$A$ is the space of terminal actions $a$ and is a measurable space with a $\sigma$-field $\mathfrak{A}$ of its subsets.

(v) Loss function: $L(\theta, a)$

The loss function $L(\theta, a)$ is a bounded non-negative real-valued measurable function defined on the product space $\Theta \times A$ ($\boldsymbol{B}(\mathscr{F} \times \mathfrak{A})$). Now, we shall introduce an intrinsic metric in the space $A$ with the help of the loss function. The intrinsic distance between two elements $a_1$ and $a_2$ of $A$ is defined by

$$(1.1) \qquad R(a_1, a_2) = \operatorname*{Sup}_{\theta \in \Theta} |L(\theta, a_1) - L(\theta, a_2)| .$$

Throughout this paper we shall consider as the $\sigma$-field $\mathfrak{A}$ of $A$ the smallest $\sigma$-field which contains all open subsets of $A$ in the sense of $R(a_1, a_2)$, (see Wald [5]).

ASSUMPTION 1.2. $A$ is compact in the sense of $R(a_1, a_2)$.

Further we shall introduce a convergence definition in $\Theta$: We shall say that $P_{\theta_i}$ converges in the regular sense to $P_{\theta_0}$ as $i \to \infty$ if for any positive integer $r$ ($1 \leqq r \leqq N+m$) we have

$$(1.2) \qquad \lim_{i \to \infty} P_{\theta_i}(B_r) = P_{\theta_0}(B_r)$$

uniformly in $B_r \in \mathscr{B}^{(r)} = \boldsymbol{B}(\mathscr{B}_1 \times \cdots \times \mathscr{B}_r)$. Also, we can introduce a metric into $\Theta$ by defining the distance between $\theta_1$ and $\theta_2$ as $r(\theta_1, \theta_2) = \operatorname*{Sup}_{B_r} |P_{\theta_1}(B_r) - P_{\theta_2}(B_r)|$ where the supremum is taken over the same range of $B_r$ as for the definition of the regular convergence.

ASSUMPTION 1.3. $\Theta$ is separable in the sense of the metric $r(\theta_1, \theta_2)$.

(vi) Space of prior probability measures: $\Xi \ni \xi$

The space of prior probability measures $\Xi$ is the set of all probability measures $\xi$ defined on the measurable space $\Theta(\mathscr{F})$.

(vii) Space of terminal decision functions: $D \ni d$

The space $D$ is the set of all measurable functions $d$ which map the product space $J \times \mathfrak{X}$ into $A$ such that if $x, y \in \mathfrak{X}$ and $x_i = y_i$ $i = 1, 2, \cdots, j+m$, then $d(j+m, x) = d(j+m, y)$ for all $d \in D$, where $x = (x_1, x_2, \cdots, x_{N+m})$, $y = (y_1, y_2, \cdots, y_{N+m})$ and $x_i, y_i \in \mathfrak{X}_i$ $i = 1, 2, \cdots, N+m$.

(viii) Class of stopping rules of sampling: $\mathcal{S} \ni S$

The stopping rules with which we are concerned throughout the present paper are defined as follows: a stopping rule of sampling, $S$, is a partition of the sample space $\mathfrak{X}$ such that (a) $S$ is a class of $N+1$ disjoint measurable subsets of $\mathfrak{X}$ whose set-theoretical sum is $\mathfrak{X}$: Thus it will be denoted by $(S_0, S_1, \cdots, S_N)$; (b) if $\mathfrak{X} \ni x, y$, $x_i = y_i$ $i = 1, 2, \cdots, j$ for some $j$ ($1 \leqq j \leqq N$) and $S_j \ni x$, then $S_j \ni y$.

DEFINITION 1.1. The space of sequential decision functions with delayed observations of size $m$ is defined as the product space $\mathcal{S}\times D$.

*Remark 1.1.* For a given $(S, d)\in\mathcal{S}\times D$, we can determine a sequential decision procedure for the sequential decision problem with delayed observations of size $m$. Suppose that $S=(S_0, \cdots, S_N)$ and the observed value $x$ has been known to belong to $S_i$, which means that a decision is made by sequential observation of the first $i$ elements of the vector $x$. Then, according to the stopping rule $S$, it is required for the experimenter to stop taking $x_{i+m+1}$ and the succeeding $x$'s as samples and thus to make a terminal decision on the basis of observed values $x_1, x_2, \cdots, x_i$ and the delayed observed values $x_{i+1}, \cdots, x_{i+m}$, immediately after the delayed observed values are obtained. In this case the terminal decision function $d$ gives a terminal action $d(i+m, x)$ which by (vii) depends only on $x_1, x_2, \cdots, x_{i+m}$.

When the loss and cost functions are given, we can define the risk function of a sequential decision function $(S, d)\in\mathcal{S}\times D$, which is denoted by $\rho(\theta, S, d)$: if $S=(S_0, S_1, \cdots, S_N)$,

$$(1.3)\qquad \rho(\theta, S, d)=\sum_{j=0}^{N}\int_{S_j}\{C(j+m, x)+L(\theta, d(j+m, x))\}\, p_\theta(x)\, d\mu(x)\,.$$

ASSUMPTION 1.4. $p_\theta(x)$ is measurable on $\mathfrak{X}\times\Theta(\boldsymbol{B}(\mathcal{B}\times\mathcal{F}))$.

The Bayes risk of the sequential decision function $(S, d)$ with respect to a prior probability measure $\xi\in\Xi$ is defined by $\int_\Theta \rho(\theta, S, d)\, d\xi(\theta)$ and is designated by $\rho(\xi, S, d)$. Then, by (1.3) and the above assumption, we have

$$(1.4)\qquad \rho(\xi, S, d)=\sum_{j=0}^{N}\int_\Theta\int_{S_j}\{C(j+m, x)+L(\theta, d(j+m, x))\}p_\theta(x)\, d\mu(x)\, d\xi(\theta)\,.$$

LEMMA 1.1. *For any $j+m\in J$ and any $d\in D$*

$$(1.5)\qquad \int_{\Theta\times S_j} L(\theta, d(j+m, x))\, p_\theta(x)\, d\mu(x)\, d\xi(\theta)=\int_{S_j} E_{j\xi}L(\boldsymbol{\theta}, d(j+m, x))dP_\xi(x),$$

*where $E_{j\xi}L(\boldsymbol{\theta}, d(j+m, x))$ is the conditional expectation of $L(\boldsymbol{\theta}, d(j+m, x))$ given $\xi\in\Xi$ and $x\in\mathfrak{X}$ and $P_\xi$ is a probability measure on $\mathfrak{X}(\mathcal{B})$ defined by*

$$(1.6)\qquad P_\xi(B)=\int_\Theta P_\theta(B)\, d\xi(\theta)\qquad \text{for any}\quad B\in\mathcal{B}\,.$$

PROOF. Let us define $\boldsymbol{B}(\mathcal{F}\times\mathcal{B})$ to be the smallest $\sigma$-field of subsets of $\Theta\times\mathfrak{X}$ which contains all rectangular subsets $A\times B$ where $A\in\mathcal{F}$ and $B\in\mathcal{B}$. If we define, for any rectangular subset $A\times B$, (1.7) $$\nu(A\times B)=\int_A P_\theta(B)\,d\xi(\theta)$$

then it is easily seen that the set function $\nu$ thus defined can be uniquely extended to a set function $\nu^*$ on $\boldsymbol{B}(\mathcal{F}\times\mathcal{B})$ which satisfies $\nu^*(A\times B)=\nu(A\times B)$ for any rectangular subset $A\times B\in\mathcal{F}\times\mathcal{B}$, and is a probability measure on the measurable space $\Theta\times\mathfrak{X}(\boldsymbol{B}(\mathcal{F}\times\mathcal{B}))$. Let us now consider a measurable mapping $t_{j+m}(\theta, x)$ from $\Theta\times\mathfrak{X}$ to the product space $\mathfrak{X}_1\times\cdots\times\mathfrak{X}_{j+m}\equiv\mathfrak{X}^{(j+m)}$ with the $\sigma$-field $\boldsymbol{B}(\mathcal{B}_1\times\cdots\times\mathcal{B}_{j+m})\equiv\mathcal{B}^{(j+m)}$ which is defined by $t_{j+m}(\theta, x)=(x_1, \cdots, x_{j+m})\equiv x^{(j+m)}$ where $x=(x_1, \cdots, x_{N+m})$. The induced probability measure on $\mathfrak{X}^{(j+m)}$ by $t_{j+m}(\theta, x)$ is denoted by $\nu'$ and clearly for any $B'\in\mathcal{B}^{(j+m)}$ we have

(1.8)
$$\begin{aligned}\nu'(B')&=\int_{t^{-1}_{j+m}(B')} p_\theta(x)\,d\mu(x)\,d\xi(\theta)\\ &=\int_\Theta P_\theta^{(j+m)}(B')\,d\xi(\theta)\\ &=\int_{B'} dP_\xi^{(j+m)}(x^{(j+m)})\end{aligned}$$

where $P_\theta^{(j+m)}$ (or $P_\xi^{(j+m)}$) is the marginal probability measure on $\mathfrak{X}^{(j+m)}$ of $P_\theta$ (or $P_\xi$). For any $B'\in\mathcal{B}^{(j+m)}$ let us define

(1.9) $$\nu''(B')\equiv\int_{t^{-1}_{j+m}(B')} L(\theta, d(j+m, x))\,d\nu^*(\theta, x)\ .$$

Obviously $\nu''$ is a measure on $\mathfrak{X}^{(j+m)}$ and is absolutely continuous with respect to the probability measure $\nu'$. Therefore, by the Radon-Nikodym theorem there exists a measurable function, say $f(x^{(j+m)})$, which is uniquely determined except a set of probability measure zero $(P_\xi^{(j+m)})$ and satisfies

(1.10) $$\nu''(B')=\int_{B'} f(x^{(j+m)})\,dP_\xi^{(j+m)}\ .$$

The $f(x^{(j+m)})$ is defined as the conditional expectation of $L(\boldsymbol{\theta}, d(j+m, x))$ given $t_{j+m}=x^{(j+m)}$. Now, noticing that $L(\theta, d(j+m, x))$ depends on $\theta$ and only the first $j+m$ elements of $x$, we have

(1.11) $$f(x^{(j+m)})=E_{j\xi}L(\boldsymbol{\theta}, d(j+m, x))\qquad\text{a.e.}\quad(P_\xi^{(j+m)})\ .$$

If we take $\bar{S}_j$ as $B'$ where $\bar{S}_j$ is the projection of $S_j$ to $\mathfrak{X}^{(j+m)}$, we have, from (1.9), (1.10) and (1.11),

$$\int_{t_{j+m}^{-1}(S_j)} L(\theta, d(j+m, x))p_\theta(x)\,d\mu(x)\,d\xi(\theta) \tag{1.12}$$
$$=\int_{\bar{S}_j} E_{j\xi}[L(\boldsymbol{\theta}, d(j+m, x)]\,dP_\xi^{(j+m)}(x)$$
$$=\int_{S_j} E_{j\xi}[L(\boldsymbol{\theta}, d(j+m, x))]\,dP_\xi(x)$$

since $d(j+m, x)$ depends only on $x_1, \cdots, x_{j+m}$.

*Remark 1.2.* We can easily see that

$$E_{j\xi}[L(\boldsymbol{\theta}, d(j+m, x))]=\int_\Theta L(\theta, d(j+m, x))\,d\xi'(\theta\,|\,\xi, x_1, \cdots, x_{j+m}) \tag{1.13}$$

except a set of probability measure zero ($P_\xi^{(j+m)}$) where $x_1, \cdots, x_{j+m}$ are the first $j+m$ elements of $x$ and $\xi'(B\,|\,\xi, x_1, \cdots, x_{j+m})$ for $B \in \mathcal{F}$ is the conditional probability measure given $x_1, \cdots, x_{j+m}$ and $\xi$.

*Remark 1.3.* As is easily seen, the probability measure $P_\xi(B)$ ($B \in \mathcal{B}$) is absolutely continuous with respect to the measure $\mu$ on $\mathfrak{X}$ and so we define $p_\xi(x) \equiv \dfrac{dP_\xi}{d\mu} = \int_\Theta p_\theta(x)\,d\xi(\theta)$.

By virtue of the lemma 1.1, (1.4) is rewritten as

$$\rho(\xi, S, d)=\sum_{j=0}^{N}\int_{S_j}\{C(j+m, x)+E_{j\xi}[L(\boldsymbol{\theta}, d(j+m, x))]\}p_\xi(x)\,d\mu(x)\,. \tag{1.14}$$

Let us now define an intrinsic distance between terminal decision functions $d_1$ and $d_2$ as follows:

$$s(d_1, d_2)=\operatorname*{Sup}_{\theta\in\Theta}\ \operatorname*{Max}_{j=0,\cdots,N}\left|\int_{\mathfrak{X}}[L(\theta, d_1(j+m, x)) - L(\theta, d_2(j+m, x))]\,p_\theta(x)\,d\mu(x)\right|. \tag{1.15}$$

Thus the space of terminal decision functions $D$ becomes a metric space. It is easily seen that $E_{j\xi}[L(\boldsymbol{\theta}, d(j+m, x)]$ is continuous in $d \in D$ with respect to the topology induced by $s(d_1, d_2)$.

ASSUMPTION 1.5. The space of terminal decision functions $D$ is compact with respect to the topology induced by the distance defined by (1.15).

This assumption is not so restrictive (see [4]). Under this assumption there exists a terminal decision function $d^*$ such that $\min_{d\in D}\rho(\xi, S, d) =\rho(\xi, S, d^*)$ for any pair $(\xi, S) \in \Xi\times\mathcal{S}$.

DEFINITION 1.2. Let us define for any $\xi \in \Xi$ and any $x \in \mathfrak{X}$

$$\tau^*_{j+m,\xi}(x)=\operatorname*{Inf}_{d\in D} E_{j\xi}[L(\boldsymbol{\theta}, d(j+m, x))] . \tag{1.16}$$

Since the space of terminal actions, $A$, is compact by the assumption 1.2 and $A=\{d(j+m, x) \mid d\in D\}$ for any $j=0, 1, \cdots, N$ and any $x\in\mathfrak{X}$, there exists an $a_{j+m,x}$ such that

$$\tau^*_{j+m,\xi}(x)=\min_{a\in A} E_{j\xi}[L(\boldsymbol{\theta}, a)]=E_{j\xi}[L(\boldsymbol{\theta}, a_{j+m,x})] . \tag{1.17}$$

Here, we should note that the minimizing value $a_{j+m,x}$ is regarded as a function of the posterior probability measure $\xi'$ produced by the prior probability measure $\xi$ and the observed values $x_1, \cdots, x_{j+m}$, hence we can write $a_{j+m,x}=a(\xi')$. Since

$$\tau^*_{j+m,\xi}(x)=E_{\xi'}[L(\boldsymbol{\theta}, a(\xi'))] \tag{1.18}$$

$\xi'\in\Xi$, we can consider that the minimizing function $a(\xi')$ is a function defined on $\Xi$.

Let us now introduce a distance definition into $\Xi$. The distance between two elements of $\Xi$, say $\xi_i$ and $\xi_j$, is defined as

$$l(\xi_i, \xi_j)=\operatorname*{Sup}_{F\in\mathcal{F}}\left|\int_F d\xi_i(\theta)-\int_F d\xi_j(\theta)\right| . \tag{1.19}$$

The topology induced by this distance definition will be called the regular topology. Then we can define a measurable space $\Xi(\mathfrak{S})$ with the smallest $\sigma$-field $\mathfrak{S}$, which includes all open set in the sense of the regular topology.

ASSUMPTION 1.6. The minimizing function $a(\xi)$, $\xi\in\Xi$, defined by (1.18) is a measurable mapping from $\Xi(\mathfrak{S})$ onto $A(\mathfrak{A})$.

If we define, using the minimizing function $a(\xi)$,

$$d^{**}(j+m, x)=a(\xi'(\cdot \mid \xi; j+m, x)) \tag{1.20}$$

for $j=0, \cdots, N$ and $x\in\mathfrak{X}$, where $\xi'(\cdot \mid \xi; j+m, x)$ stands for the posterior probability measure on $\Theta(\mathcal{F})$ induced by the prior probability measure $\xi$ and the observed values $x_1, \cdots, x_{j+m}$, then $d^{**}$ is really a terminal decision function. For the $\xi'(\cdot \mid \xi; j+m, x)$ is measurable in $x$ by the assumption 1.4 and $a(\xi)$ is measurable by the assumption 1.6. Thus, it easily follows that

$$\rho(\xi, S, d^*)=\rho(\xi, S, d^{**})\equiv\rho^*(\xi, S) . \tag{1.21}$$

*Remark 1.4.* Although the assumption 1.6 seems restrictive, it will be usually satisfied at least in the case of actual decision problems. For example, if the minimizing value $a$ of $E_\xi[L(\boldsymbol{\theta}, a)]$ is uniquely determined for each $\xi\in\Xi$, then it is easily shown that the minimizing function $a(\xi)$ is a continuous function of $\xi$. Of course, the assumption 1.6 is satisfied in the more general situations.

The next stage is to prove the existence of an optimal Bayes sequential sampling scheme, say $S^*$, for a given $\xi \in \Xi$. The process of the proof is quite similar to the one for the sequential problems without delayed observations (see [3]).

DEFINITION 1.3. For any $\xi \in \Xi$ and integer $j$ ($j=0, 1, \cdots, N$) let us define a function $U_j(x)$ on $\mathfrak{X}$ by

$$(1.22)\qquad U_j(x) \equiv \int_{\mathfrak{X}_{j+1}\times\cdots\times\mathfrak{X}_{j+m}} [C(j+m, x^{(j+m)}) + \tau^*_{j\xi}(x^{(j+m)})]\, dP_\xi^{(j+m)}(x_{j+1}, \cdots, x_{j+m} \mid x_1, \cdots, x_j)$$

where $P_\xi^{(j+m)}(B \mid x_1, \cdots, x_j)$ is the conditional probability measure on $\mathfrak{X}_{j+1}\times\cdots\times\mathfrak{X}_{j+m}$ given $\xi$ and $x_1, \cdots, x_j$.

*Remark 1.5.* Since $C(j+m, x)$ and $\tau^*_{j\xi}(x)$ depend only on $x_1, x_2, \cdots, x_{j+m}$, the conditional expectation of $C(j+m, x)+\tau^*_{j\xi}(x)$, $U_j(x)$, depends only on $x_1, x_2, \cdots, x_j$. $U_j(x)$ is clearly measurable.

DEFINITION 1.4. A sequence of functions $\{\alpha_j(x)\}$ $j=0, 1, \cdots, N$ is defined in the following successive manner:

$$(1.23)\qquad \begin{aligned} &\alpha_N(x) = U_N(x) \qquad \text{and} \\ &\alpha_j(x) = \min\{U_j(x), E_{j\xi}(\alpha_{j+1}(X) \mid x)\} \qquad \text{for} \quad j=0, 1, 2, \cdots, N-1 \end{aligned}$$

where

$$(1.24)\qquad E_{j\xi}(\alpha_{j+1}(X) \mid x) \equiv \int_{\mathfrak{X}_{j+1}} \alpha_{j+1}(x_1, \cdots, x_j, x_{j+1})\, dP_\xi^{(j+1)}(x_{j+1} \mid x_1, \cdots, x_j)$$

($\equiv$the conditional expectation of $\alpha_{j+1}$ given $\xi$ and $x_1, \cdots, x_j$).

Let us now define a stopping rule by means of $\{U_j(x)\}$ and $\{\alpha_j(x)\}$. Defining $S_j^*$ to be the set of $x \in \mathfrak{X}$ such that $U_i(x) > \alpha_i(x)$ for $i=1, 2, \cdots, j-1$ and $U_j(x) = \alpha_j(x)$, we have a stopping rule $S^* = (S_0^*, S_1^*, \cdots, S_N^*)$. Considering that $\alpha_j(x)$ and $U_j(x)$ depend only on $x_1, \cdots, x_j$, we can easily see that $S^*$ satisfies the property of a stopping rule which was given by (viii).

THEOREM 1.1. *The stopping rule $S^*$ defined above is a Bayes solution for the prior distribution $\xi$ in the sense that*

$$(1.25)\qquad \rho^*(\xi, S^*) = \min_{S \in \mathcal{S}} \rho^*(\xi, S)$$

*where the function* $\rho^*(\xi, S)$ *is given in (1.21).*

PROOF. For any stopping rule $S=(S_0, S_1, \cdots, S_N) \in \mathcal{S}$, we define

$$T_{r+1} \equiv S_{r+1} \cup \cdots \cup S_N \qquad r=0, 1, \cdots, N-1 \tag{1.26}$$

$$g(r) \equiv \sum_{j=0}^{r-1} \int_{S_j} \alpha_j(x)\, dP_\xi(x) + \int_{T_r} \alpha_r(x)\, dP_\xi(x) \qquad r=1, 2, \cdots, N \tag{1.27}$$

or

$$\equiv \sum_{j=0}^{r} \int_{S_j} \alpha_j(x)\, dP_\xi(x) + \int_{T_{r+1}} \alpha_r(x)\, dP_\xi(x) \tag{1.27$'$}$$

in particular,

$$g(N) = \sum_{j=0}^{N} \int_{S_j} \alpha_j(x)\, dP_\xi(x) \quad \text{and} \quad g(0) = \int_{\mathfrak{X}} \alpha_0(x)\, dP_\xi(x)\ . \tag{1.28}$$

Since whether $x$ belongs to $T_{r+1}$ or not depends only on the first $r$ elements $x_1, \cdots, x_r$ of $x$ and $E_{r\xi}[\alpha_{r+1}(X) \mid x]$ also depends only on these values, we obtain

$$\int_{T_{r+1}} \alpha_{r+1}(x)\, dP_\xi(x) = \int_{T_{r+1}} E_{r\xi}[\alpha_{r+1}(X) \mid x]\, dP_\xi(x)\ . \tag{1.29}$$

The validity of (1.29) can be assured along the same line about the conditional expectation as in lemma 1.1. From (1.27) and (1.29) we have

$$g(r+1) = \sum_{j=1}^{r} \int_{S_j} \alpha_j(x)\, dP_\xi(x) + \int_{T_{r+1}} E_{r\xi}[\alpha_{r+1}(X) \mid x]\, dP_\xi(x)\ . \tag{1.30}$$

Since $\alpha_r(x) \leqq E_{r\xi}[\alpha_{r+1}(X) \mid x]$ if $x \in T_{r+1}$, we get

$$g(r+1) \geqq g(r) \qquad \text{for} \quad r=0, 1, \cdots, N-1 \tag{1.31}$$

by comparing (1.27)′ and (1.30). The function $g(r)$ is, thus, an increasing function of $r$. By the definitions of $U_j(x)$ and $\rho^*(\xi, S)$, we have

$$\rho^*(\xi, S) = \sum_{j=0}^{N} \int_{S_j} U_j(x)\, dP_\xi(x) \tag{1.32}$$

and, since $\alpha_j(x) \leqq U_j(x)$ for any $x \in \mathfrak{X}$, we obtain

$$\rho^*(\xi, S) \geqq \sum_{j=0}^{N} \int_{S_j} \alpha_j(x)\, dP_\xi(x) = g(N) \tag{1.33}$$

and consequently by (1.31)

$$\rho^*(\xi, S) \geqq g(N) \geqq g(0) = \int_{\mathfrak{X}} \alpha_0(x)\, dP_\xi(x) \equiv \rho^*(\xi)\ . \tag{1.34}$$

If, however, $S=S^*$ and $T_{r+1}\ni x$, it follows that

$$\alpha_r(x)=E_{r\xi}[\alpha_{r+1}(X)\,|\,x]\,, \tag{1.35}$$

hence, from (2.24)′ and (2.27)

$$g(r+1)=g(r) \qquad r=0, 1, \cdots, N-1\,. \tag{1.36}$$

Therefore, by (1.34) and (1.36) we have

$$\rho^*(\xi, S)\geqq\rho^*(\xi)=\rho^*(\xi, S^*) \qquad \text{for any} \quad S\in\mathcal{S}\,,$$

which is what we wanted to prove.

THEOREM 1.2. *The function $\rho^*(\xi)$, defined in (1.34), is a concave function on $\Xi$.*

PROOF. For any $\xi_1$ and $\xi_2$ and any $\alpha$ $(0\leqq\alpha\leqq1)$, let $\xi=\alpha\xi_1+(1-\alpha)\xi_2$. Then, for any $(S, d)\in\mathcal{S}\times D$, we have, by (1.4), $\rho(\xi, S, d)=\alpha\rho(\xi_1, S, d)+(1-\alpha)\rho(\xi_2, S, d)$, consequently,

$$\rho(\xi, S, d)\geqq\alpha \operatorname*{Inf}_{(S', d')\in\mathcal{S}\times D}\rho(\xi_1, S', d')+(1-\alpha)\operatorname*{Inf}_{(S'', d'')\in\mathcal{S}\times D}\rho(\xi_2, S'', d'')$$

and hence

$$\rho^*(\xi)=\operatorname*{Inf}_{(S,d)\in\mathcal{S}\times D}\rho(\xi, S, d)\geqq\alpha\rho^*(\xi_1)+(1-\alpha)\rho^*(\xi_2)\,.$$

*Remark 1.6.* It is interesting to note that, in spite of the existence of delayed observations, the results obtained in this section are quite similar to the ones for the corresponding sequential decision problem without delayed observations (cf. Blackwell & Girshick [3]).

In the formulation of the problem in this section, we did not assume the independence of the random variables $g_1(z), \cdots, g_{N+m}(z)$. We were, therefore, limited in taking advantage of the Bayes theorem. In the succeeding section, however, we shall use this theorem extensively under the assumption of independence of the random variables to be observed.

## 2. Constant cost and identically, independently distributed observations

In this section we shall impose two additional assumptions on the general truncated sequential decision problem formulated in the previous section.

ASSUMPTION 2.1. For any $\theta\in\Theta$ $\{g_i(Z)\}$ $i=1, 2, \cdots, N+m$ are identically and independently distributed with the common probability density

$p_\theta(x)$ with respect to a measure $\mu$ on $\mathfrak{X}$ where $\mathfrak{X}_1=\mathfrak{X}_2=\cdots=\mathfrak{X}_{N+m}$; $\mathcal{B}_1=\mathcal{B}_2=\cdots=\mathcal{B}_{N+m}$.

ASSUMPTION 2.2. $C(j+m, x)=(j+m)c$, where $j=0, 1, 2, \cdots, N$ and $c>0$.

For any given sequential decision functions with delayed observations of size $m$, $(S, d)\in\mathcal{S}\times D$, we have

$$(2.1)\quad \rho(\theta, S, d)=\sum_{j=0}^{N}\int_{S_j}[(j+m)c+L(\theta, d(j+m, x))]\prod_{i=1}^{N+m}p_\theta(x_i)\,d\mu(x_i)$$
$$=\sum_{j=0}^{N}(j+m)cP_\theta(S_j)+\sum_{j=0}^{N}\int_{S_j}L(\theta, d(j+m, x))\prod_{i=1}^{N+m}p_\theta(x_i)\,d\mu(x_i)$$

$$(2.2)\quad \rho(\xi, S, d)=\sum_{j=0}^{N}(j+m)cP_\xi(S_j)+\sum_{j=0}^{N}\int_{\Theta}\int_{S_j}L(\theta, d(j+m, x))$$
$$\times\prod_{i=1}^{N+m}p_\theta(x_i)\,d\mu(x_i)\,d\xi(\theta)\ .$$

Now by the lemma 1.1,

$$(2.3)\quad \int_{\Theta}\int_{S_j}L(\theta, d(j+m, x))\prod_{i=1}^{N+m}p_\theta(x_i)\,d\mu(x_i)\,d\xi(\theta)$$
$$=\int_{S_j}E_{j\xi}[L(\boldsymbol{\theta}, d(j+m.\ x))]\,p_\xi(x)\prod_{i=1}^{N+m}d\mu(x_i)$$

where

$$(2.4)\quad p_\xi(x)=\int_{\Theta}\prod_{i=1}^{N+m}p_\theta(x_i)\,d\xi(\theta)$$

$$(2.5)\quad E_{j\xi}[L(\boldsymbol{\theta}, d(j+m, x))]=\int_{\Theta}L(\theta, d(j+m, x))\,d\xi(\theta\,|\,x_1, \cdots, x_{j+m})$$

and

$$(2.6)\quad d\xi(\theta\,|\,x_1, \cdots, x_{j+m})=\begin{cases}\dfrac{\prod_{i=1}^{j+m}p_\theta(x_i)\,d\xi(\theta)}{\int_{\Theta}\prod_{i=1}^{j+m}p_\theta(x_i)\,d\xi(\theta)} & \text{if } \int_{\Theta}\prod_{i=1}^{j+m}p_\theta(x_i)\,d\xi(\theta)>0\,,\\ \text{arbitrary} & \text{otherwise.}\end{cases}$$

LEMMA 2.1. *Under the assumption 2.1, we have*

$$(2.7)\quad p_\xi^{(j+m)}(x_{j+1}, \cdots, x_{j+m}\,|\,x_1, \cdots, x_j)$$
$$=\int_{\Theta}\prod_{i=j+1}^{j+m}p_\theta(x_i)\,d\xi(\theta\,|\,x_1, \cdots, x_j)\ .$$

PROOF. Let $t(\theta, x^{(j+m)})$ be a measurable mapping from $\Theta\times\mathfrak{X}^{(j+m)}(\boldsymbol{B}(\mathcal{F}\times\mathcal{B}^{(j+m)}))$ to $\mathfrak{X}^{(j)}(\mathcal{B}^{(j)})$: $t(\theta, x^{(j+m)})=(x_1, \cdots, x_j)$, that is, the value of $t$ is equal to the vector which consists of the first $j$ element of $x^{(j+m)}$. Then, $t^{-1}(B)=\Theta\times B\times\mathfrak{X}_{j+1}\times\cdots\times\mathfrak{X}_{j+m}$ for $B\in\mathcal{B}^{(j)}$. Define

$$A=\Theta\times\mathfrak{X}_1\times\cdots\times\mathfrak{X}_j\times\bar{A}$$

where

$$\bar{A}\in\boldsymbol{B}(\mathcal{B}_{j+1}\times\cdots\times\mathcal{B}_{j+m})\,.$$

Since $A\cap t^{-1}(B)=\Theta\times B\times\bar{A}$, we have for any such $A$

$$\text{(2.8)}\quad \mu^*(A\cap t^{-1}(B))=\int_\Theta\int_{B\times\bar{A}}p_\theta^{(j+m)}(x_1,\cdots,x_{j+m})\,d\mu(x_1)\cdots d\mu(x_{j+m})\,d\xi(\theta)$$

$$=\int_\Theta\int_B\left[\int_{\bar{A}}\prod_{i=j+1}^{j+m}p_\theta(x_i)\,d\mu(x_i)\right]\prod_{i=1}^{j}p_\theta(x_i)\,d\mu(x_i)\,d\xi(\theta)$$

(by the assumption 2.1) .

Since

$$\text{(2.9)}\qquad d\xi(\theta\,|\,x_1,\cdots,x_j)=\prod_{i=1}^{j}p_\theta(x_i)\,d\xi(\theta)\Big/\int_\Theta\prod_{i=1}^{j}p_\theta(x_i)\,d\xi(\theta)\,,$$

(2.8) can be rewritten as

$$\mu^*(A\cap t^{-1}(B))=\int_B\int_\Theta\left[\int_{\bar{A}}\prod_{i=j+1}^{j+m}p_\theta(x_i)\,d\mu(x_i)\right]d\xi(\theta\,|\,x_1,\cdots,x_j)\,p_\xi^{(j)}(x_1,\cdots,x_j)$$
$$\times\,d\mu(x_1)\cdots d\mu(x_j)$$

(2.10)

$$\mu^*(A\cap t^{-1}(B))=\int_B\int_{\bar{A}}\left[\int_\Theta\prod_{i=j+1}^{j+m}p_\theta(x_i)\,d\xi(\theta\,|\,x_1,\cdots,x_j)\right]$$
$$\times\prod_{i=j+1}^{j+m}d\mu(x_i)\,p_\xi^{(j)}(x_1,\cdots,x_j)\prod_{i=1}^{j}d\mu(x_i)\,.$$

From (2.10) we conclude that the conditional probability of $A$ (or $\bar{A}$) given $x_1,\cdots,x_j$ is

$$\int_{\bar{A}}\left[\int_\Theta\prod_{i=j+1}^{j+m}p_\theta(x_i)\,d\xi(\theta\,|\,x_1,\cdots,x_j)\right]\prod_{i=j+1}^{j+m}d\mu(x_i)$$

and consequently the corresponding probability density function w.r.t. $\underbrace{\mu\times\cdots\times\mu}_{m}$ is

$$\int_\Theta\prod_{i=j+1}^{j+m}p_\theta(x_i)\,d\xi(\theta\,|\,x_1,\cdots,x_j)\,.$$

Thus the lemma is proved.

DEFINITION 2.1. The conditional probability measure given a prior probability measure $\xi\in\Xi$ and observed values $x_1,\cdots,x_j$ is defined by the Bayes theorem. It is denoted by $\boldsymbol{\xi}(\xi; x_1, \cdots, x_j)$ or $\xi_j(B)$ for $B \in \mathcal{F}$.

LEMMA 2.2. *Under the assumption 2.1*

$$\boldsymbol{\xi}(\xi; x_1, \cdots, x_{j+1}) = \boldsymbol{\xi}(\xi_j; x_{j+1})$$

*where* $\xi_j$ *is given by* $\boldsymbol{\xi}(\xi; x_1, \cdots, x_j)$.

PROOF. Let $\zeta_{j+1} \equiv \boldsymbol{\xi}(\xi_j; x_{j+1})$. For any $B \in \mathcal{F}$,

$$\int_B d\zeta_{j+1}(\theta) = \int_B p_\theta(x_{j+1})\, d\xi_j(\theta) \Big/ \int_\Theta p_\theta(x_{j+1})\, d\xi_j(\theta) . \tag{2.11}$$

Since

$$d\xi_j(\theta) = \prod_{i=1}^{j} p_\theta(x_i)\, d\xi(\theta) \Big/ \int_\Theta \prod_{i=1}^{j} p_\theta(x_i)\, d\xi(\theta) ,$$

we have

$$\begin{aligned}\zeta_{j+1}(B) &= \int_B \prod_{i=1}^{j+1} p_\theta(x_i)\, d\xi(\theta) \Big/ \int_\Theta \prod_{i=1}^{j+1} p_\theta(x_i)\, d\xi(\theta) \\ &= \xi_{j+1}(B) . \end{aligned}$$

Q.E.D.

As in the previous section, we define

$$\begin{aligned}\operatorname*{Inf}_{d \in D} E_{j\xi}[L(\boldsymbol{\theta}, d(j+m, x))] &= \tau^*_{j\xi}(x) \\ &= E_{j\xi}[L(\boldsymbol{\theta}, d^*(j+m, x))]\end{aligned} \tag{2.12}$$

where the existence of $d^*$ is insured by the compactness of $D$, the continuity of $E_{j\xi}[L(\boldsymbol{\theta}, d(j+m, x))]$ in $d$ and the assumption 1.6. Since $\tau^*_{j\xi}(x)$ depends on $x$ and $\xi'$ through the functional form $\boldsymbol{\xi}(\xi; x_1, \cdots, x_{j+m})$, we can write

$$\tau^*_{j\xi}(x) = \tau^*[\boldsymbol{\xi}(\xi; x_1, \cdots, x_{j+m})] . \tag{2.13}$$

The functions $U_j(x)$, defined in the previous section, take the simpler forms:

$$U_j(x) = (j+m)c + E\tau^*[\boldsymbol{\xi}(\xi; x_1, \cdots, x_j, X_{j+1}, \cdots, X_{j+m})] \tag{2.14}$$

or

$$= (j+m)c + E\tau^*[\boldsymbol{\xi}(\xi_j; X_{j+1}, \cdots, X_{j+m})] \qquad j = 0, 1, 2, \cdots, N$$

where

$$\xi_j = \boldsymbol{\xi}(\xi; x_1, \cdots, x_j) .$$

Therefore, $U_j(x)$ is regarded as a function of $\xi_j$ and hence it may be designated by $U_j(\xi_j)$. It should be noted that $U_j+(N-j)c=U_N$ $j=0, 1, \cdots, N-1$. The functions $\alpha_j(x)$ or $\alpha_j(\xi_j)$ $j=0, 1, 2, \cdots, N$ are defined by

$$(2.15)\quad \begin{aligned} &\alpha_N(x)=\alpha_N(\xi_N)=U_N(\xi_N) \qquad \text{for any} \quad \xi_N \in \Xi \\ &\alpha_j(x)=\alpha_j(\xi_j)=\min\{U_j(\xi_j), E\alpha_{j+1}[\boldsymbol{\xi}(\xi_j; X_{j+1})]\} . \end{aligned}$$

Now, if we define

$$(2.16)\quad \varphi_0(\xi)=E\tau^*[\boldsymbol{\xi}(\xi; X_{N+1}, \cdots, X_{N+m})]$$

and in general for $j=1, 2, \cdots, N$,

$$(2.17)\quad \varphi_j(\xi)=\min\{\varphi_0(\xi), c+E\varphi_{j-1}[\boldsymbol{\xi}(\xi; X_{N-j+1})]\} .$$

Then, we have

$$(2.18)\quad \alpha_j(\xi)=(N-j+m)c+\varphi_j(\xi) \qquad j=0, 1, \cdots, N.$$

As has been shown in the previous section, the optimal stopping rule $S^*=\{S_0^*, S_1^*, \cdots, S_N^*\}$ is specified by $\{U_j(x)\}$ and $\{\alpha_j(x)\}$. Owing to assumptions 2.1 and 2.2, however, we can characterize the optimal stopping rule in the space of prior probability measures $\Xi$. The optimal stopping rule is completely determined by a sequence of subsets of $\Xi$, $\{\Xi_j^*\}$ $j=0, 1, \cdots, N$ such that $\Xi_N^* \subset \Xi_{N-1}^* \subset \cdots \subset \Xi_0^*$ and each $\Xi_j^*$ is defined as

$$(2.19)\quad \Xi_j^*=\{\xi \in \Xi \text{ such that } \varphi_j(\xi)=\varphi_0(\xi)\} .$$

By $\{\Xi_j^*\}$ $j=0, \cdots, N$, the optimal decision procedure proceeds as follows:

[0] (i) If $\xi_0 \in \Xi_N^*$, the observations are made on random variables $X_1, X_2, \cdots, X_m$ and the optimal terminal decision is made on the basis of these observed values. It is defined by $d_0^*(x_1, \cdots, x_m)$ such that for each $(x_1, \cdots, x_m)$

$$(2.20)\quad \int_\Theta L(\theta, d_0^*(x_1, \cdots, x_m))\, d\xi_m(\theta)=\tau^*[\boldsymbol{\xi}(\xi_0; x_1, \cdots, x_m)]$$

where $\xi_m=\boldsymbol{\xi}(\xi_0; x_1, \cdots, x_m)$. (2.20) is also written as

$$(2.21)\quad \int_\Theta L(\theta, d_0^*(\xi_m))\, d\xi_m(\theta)=\tau^*(\xi_m) .$$

(ii) If $\xi_0 \notin \Xi_N^*$, the random variable $X_{m+1}$ is taken as a sample to be added to the samples $X_1, \cdots, X_m$.

*Remark 2.1.* The existence of $d_0^*(x_1, \cdots, x_m)$ or $d_0^*(\xi_m)$ and its measurability are insured by the assumptions introduced in the previous section.

[1] If the case [0] (ii) has occurred, we first compute $\xi_1 = \boldsymbol{\xi}(\xi_0; x_1)$ on the basis of the prior probability measure $\xi_0$ and the observed value $x_1$ of $X_1$.

(i) If $\xi_1 \in \Xi^*_{N-1}$, then the optimal terminal decision is made on the basis of $\xi_1$ and the observed values $x_2, \cdots, x_{m+1}$ of $X_2, \cdots, X_{m+1}$ which have been taken as samples. The optimal terminal decision $d_1^*(\xi_1; x_2, \cdots, x_{m+1})$ or $d_1^*(\xi_{m+1})$ is given by

$$\int_\Theta L(\theta, d_1^*(\xi_{m+1}))\, d\xi_{m+1} = \tau^*(\xi_{m+1}) . \tag{2.22}$$

(ii) If $\xi_1 \notin \Xi^*_{N-1}$, the random variable $X_{m+2}$ is taken as a sample to be added to the samples $X_1, \cdots, X_{m+1}$.

In general, the procedure can be written down as follows: Let $i$ be a positive integer $(1 \leqq i < N)$,

[$i$] If the case [$i-1$] (ii) has occurred, we first compute $\xi_i = \boldsymbol{\xi}(\xi_{i-1}; x_i)$ on the basis of $\xi_{i-1}$ and the observed value $x_i$ of $X_i$.

(i) If $\xi_i \in \Xi^*_{N-i}$ then the optimal terminal decision is made on the basis of $\xi_i$ and the automatically following observed values, say $x_{i+1}, \cdots, x_{i+m}$ of $X_{i+1}, \cdots, X_{i+m}$. The optimal terminal decision $d_i^*(\xi_i; x_{i+1}, \cdots, x_{i+m})$ or $d_i^*(\xi_{i+m})$ is given by

$$\int_\Theta L(\theta, d_i^*(\xi_{i+m}))\, d\xi_{i+m}(\theta) = \tau^*(\xi_{i+m}) \tag{2.23}$$

where

$$\xi_{i+m} = \boldsymbol{\xi}(\xi_i; x_{i+1}, \cdots, x_{i+m}) .$$

(ii) If $\xi_i \notin \Xi^*_{N-i}$, then the random variable $X_{i+m+1}$ is taken as a sample to be added to the samples $X_1, \cdots, X_{i+m}$ which have already been taken.

This procedure is repeated as long as the posterior probability measure $\xi_i$ does not belong to $\Xi^*_{N-i}$, $i<N$. However, if we have reached to $i=N$, following the above procedure, we have to make the optimal terminal decision by the decision rule $d_N^*$ such that

$$\int_\Theta L(\theta; d_N^*(\xi_{N+m}))\, d\xi_{N+m}(\theta) = \tau^*(\xi_{N+m})$$

where

$$\xi_{N+m} = \boldsymbol{\xi}(\xi_N; x_{N+1}, \cdots, x_{N+m}) .$$

The optimality of the sequential decision procedure defined above is easily proved as a special case of the theorems 1.1 which was given in the more general form.

THEOREM 2.1. *Under the assumptions 2.1 and 2.2, the sequential decision procedure defined above is optimal for the sequential decision problem with delayed observations which was formulated in the previous section.*

THEOREM 2.2. *The function* $\tau^*(\xi)$ *defined by (2.13) is a concave function of* $\xi \in \Xi$.

The proof is quite similar to the one for the theorem 1.2, so it is omitted.

THEOREM 2.3. *Every* $\Xi_j^*$ $j=0, 1, \cdots, N$ *is closed, but not necessarily convex where the closedness is based on the regular topology introduced into the space* $\Xi$ *in the previous section.*

PROOF. The closedness easily follows from the continuity of $\varphi_j(\xi)$ and $E\varphi_j[\boldsymbol{\xi}(\xi; X_j)]$ $j=0, 1, \cdots, N$ as functions of $\xi$. The statement that $\Xi_j^*$ $j=0, 1, \cdots, N$ is not necessarily convex will be clarified in the sections 3 and 4. This is a typical character due to the existence of delayed observations.

## 3. Truncated and non-truncated dichotomous decision problems

In this section we shall treat a dichotomous decision problem which is obtained from the problem treated in the previous section by imposing further restrictions on the parameter space, action space and loss function. The restrictions are as follows: The parameter space $\Theta$ consists of only two elements $\theta_1$ and $\theta_2$; the action space $A$ consists of only two elements $a_1$ and $a_2$; finally the loss function is defined by $L(\theta_1, a_2)=w_{12}$, $L(\theta_2, a_1)=w_{21}$ and $L(\theta_1, a_1)=L(\theta_2, a_2)=0$ where $w_{12}$ and $w_{21}$ are positive.

*Remark 3.1.* Due to the finiteness of the spaces $\Theta$ and $A$, all assumptions given in section 1 are clearly satisfied.

The problem is to obtain the optimal sequential decision procedure for a dichotomous decision problem with delayed observations of size $m$. The difference between the usual truncated or non-truncated dichotomous sequential problem and ours is the existence of delayed observations in the latter: In our model it is assumed that the $m$ additional observed values are obtained if we wait for them after we have stopped sampling according to a stopping rule. We also assume that we always utilize these additional observed values for our terminal decision.

Since $\Theta=\{\theta_1, \theta_2\}$ in our problem, any prior probability measure $\xi$ over $\Theta$ is determined by $\xi(\{\theta_1\})$, say $\eta$, and therefore it is parametrized by the parameter $\eta$, $0 \leqq \eta \leqq 1$.

Let us first consider the problem truncated at $N+m$. In order to analyze the problem by "backward induction", consider first the case where the maximum number of observations with observed values $x_1, x_2, \cdots, x_N, \cdots, x_{N+m}$ have been made. If the prior probability of $\theta_1$ is $\eta_0$, the posterior probability of $\theta_1$, $\eta_{N+m}=\boldsymbol{\eta}(\eta_0; x_1, \cdots, x_{N+m})$, is given by

$$(3.1) \qquad \eta_{N+m}=\eta_0 \prod_{i=1}^{N+m} p_{\theta_1}(x_i) \Big/ \left\{\eta_0 \prod_{i=1}^{N+m} p_{\theta_1}(x_i)+(1-\eta_0) \prod_{i=1}^{N+m} p_{\theta_2}(x_i)\right\} .$$

In this case the conditional expected loss for choosing the terminal action $a_1$ plus the cost of observation is

$$(3.2) \qquad (N+m)c+w_{21}(1-\eta_{N+m})$$

and, similarly, the conditional expected loss for choosing the other terminal action $a_2$ is

$$(3.3) \qquad (N+m)c+w_{12}\,\eta_{N+m} .$$

Thus, comparing both conditional expected losses, we obtain the following optimal (Bayes) terminal decision procedure: Choose the action $a_1$ if $w_{21}(1-\eta_{N+m})\leqq w_{12}\eta_{N+m}$, otherwise choose the action $a_2$. Then, the conditional Bayes risk of this procedure is

$$(3.4) \qquad \min\{(N+m)c+w_{21}(1-\eta_{N+m}),\ (N+m)c+w_{12}\eta_{N+m}\} .$$

Let us now consider the case where random variables $X_1, \cdots, X_N$ have already been observed with observed values $x_1, x_2, \cdots, x_N$ respectively, but $X_{N+1}, \cdots, X_{N+m}$ have not yet been observed. In this case we shall say that we are at the $N$th stage. Since, at the $N$th stage, $\eta_{N+m}=\boldsymbol{\eta}(\eta_0; x_1, \cdots, x_N, X_{N+1}, \cdots, X_{N+m})$, the Bayes risk (3.4) is a random variable. Further, since $\eta_N=\boldsymbol{\eta}(\eta_0; x_1, \cdots, x_N)$, we can write $\eta_{N+m}=\boldsymbol{\eta}(\eta_N; X_{N+1}, \cdots, X_{N+m})$. The expected value of the Bayes risk evaluated at the $N$th stage is a function of $\eta_N$ and is denoted by $U_N(\eta_N)$:

$$(3.5) \qquad U_N(\eta_N)=E\min\left\{\begin{array}{l}(N+m)c+w_{21}[1-\boldsymbol{\eta}(\eta_N;\, X_{N+1}, \cdots, X_{N+m})] \\ (N+m)c+w_{12}\boldsymbol{\eta}(\eta_N;\, X_{N+1}, \cdots, X_{N+m})\end{array}\right\}$$
$$=(N+m)c+\varphi_0(\eta_N)$$

where

$$(3.6) \qquad \varphi_0(\eta_N)=E\min\{w_{21}[1-\boldsymbol{\eta}(\eta_N;\, X_{N+1}, \cdots, X_{N+m})],\ w_{12}\boldsymbol{\eta}(\eta_N;\, X_{N+1}, \cdots, X_{N+m})\} .$$

In order to proceed backward, let us suppose that we are at the $N-1$st stage, that is, we have observed $X_1, \cdots, X_{N-1}$ with the observed values, say $x_1, \cdots, x_{N-1}$, but the succeeding $m$ random variables $X_N, \cdots, X_{N+m-1}$ have not yet been observed. Then, there are two choices: (i) the one is to stop taking a sample any more and wait until the observation of the additional sample $X_N, \cdots, X_{N+m-1}$ is completed and follow the optimal Bayes terminal decision procedure based on all observed values; (ii) the other is to take one more sample $X_{N+m}$ and to follow the optimal decision procedure which was already described. In the case (i), the conditional Bayes risk given $X_1=x_1, \cdots, X_{N-1}=x_{N-1}$ is denoted by $U_{N-1}(\eta_{N-1})$:

$$(3.7)\quad U_{N-1}(\eta_{N-1})=(N-1+m)c+E\min\left\{\begin{array}{l} w_{21}[1-\boldsymbol{\eta}(\eta_{N-1};\, X_N, \cdots, X_{N-1+m})] \\ w_{12}\boldsymbol{\eta}(\eta_{N-1};\, X_N, \cdots, X_{N-1+m}) \end{array}\right\}$$
$$=(N-1+m)c+\varphi_0(\eta_{N-1})$$

where $\eta_{N-1}=\boldsymbol{\eta}(\eta_0;\, x_1, x_2, \cdots, x_{N-1})$. In the case (ii), the conditional Bayes risk given $X_1=x_1, \cdots, X_{N-1}=x_{N-1}$ is expressed by

$$(3.8)\qquad (N+m)c+E\varphi_0[\boldsymbol{\eta}(\eta_{N-1};\, X_N)]\ .$$

Therefore, the best way to decide whether to stop taking a sample or continue it at the $N-1$st stage should be solely based on (3.7) and (3.8) as follows: If (3.7) is not greater than (3.8), that is,

$$(3.9)\qquad \varphi_0(\eta_{N-1})\leqq c+E\varphi_0[\boldsymbol{\eta}(\eta_{N-1};\, X_N)]$$

we should stop sampling immediately; on the other hand, if (3.7) is greater than (3.8), that is

$$(3.10)\qquad \varphi_0(\eta_{N-1})>c+E\varphi_0[\boldsymbol{\eta}(\eta_{N-1};\, X_N)]$$

we should continue to take a sample. Therefore, we can write the conditional Bayes risk at the $N-1$st stage, $\alpha_{N-1}(\eta_{N-1})$, given $X_1=x_1, \cdots, X_{N-1}=x_{N-1}$ as follows:

$$(3.11)\qquad \alpha_{N-1}(\eta_{N-1})=(N-1+m)c+\varphi_1(\eta_{N-1})$$

where we define

$$(3.12)\qquad \varphi_1(\eta_{N-1})=\min\{\varphi_0(\eta_{N-1}),\ c+E\varphi_0[\boldsymbol{\eta}(\eta_{N-1};\, X_N)]\}\ .$$

We, thus, have expressed the conditional Bayes risk at the $N-1$st stage by using the conditional Bayes risk at the $N$th stage, which is easily seen from (3.5), (3.11) and (3.12). Further, conditions (3.9) and (3.10) which determine the stopping rule at the $N-1$st stage are rewritten respectively as

$$(3.13)\qquad \varphi_1(\eta_{N-1})=\varphi_0(\eta_{N-1})$$

and

$$(3.14)\qquad \varphi_1(\eta_{N-1})\neq\varphi_0(\eta_{N-1})\ .$$

Let us now define functions $\varphi_j$ in the same manner as in section 2:

$$\varphi_i(\eta_{N-i}) = \min\{\varphi_0(\eta_{N-i}),\ c + E\varphi_{i-1}[\boldsymbol{\eta}(\eta_{N-i};\ X_{N-i+1})]\} \tag{3.15}$$

for $i=2, 3, \cdots, N$, where $\eta_{N-i}=\boldsymbol{\eta}(\eta_0; x_1, \cdots, x_{N-i})$. In this way all $\varphi$'s are defined by proceeding backward from the $N$th stage to the starting stage of our sequential decision problem. In accordance with the successive determination of $\varphi_i$, the optimal stopping rule and optimal terminal decision procedure are determined successively as follows:

$[N-i]$ Stopping rule: Suppose that observation has been continued with observed values $x_1, \cdots, x_{N-i}$: that is, we are at the $N-i$th stage. We first compute $\eta_{N-i}=\boldsymbol{\eta}(\eta_0; x_1, \cdots, x_{N-i})=\boldsymbol{\eta}(\eta_{N-i-1}; x_{N-i})$. The stopping rule at this stage is: (i) Stop taking a sample if $\varphi_i(\eta_{N-i})=\varphi_0(\eta_{N-i})$ and (ii) Take $X_{N-i+m+1}$ as a sample if $\varphi_i(\eta_{N-i})\neq\varphi_0(\eta_{N-i})$.

$[N-i]$ Terminal decision procedure: When $[N-i]$ (i) occurred, we have to wait until the additional observed values, say, $x_{N-i+1}, \cdots, x_{N-i+m}$ on $X_{N-i+1}, \cdots, X_{N-i+m}$ are obtained. Thus, the posterior probability $\eta_{N-i+m}$ is computed by $\eta_{N-i+m}=\boldsymbol{\eta}(\eta_{N-i}; x_{N-i+1}, \cdots, x_{N-i+m})$ and then the terminal decision is made according to the following rule: Choose the action $a_1$ if $w_{21}(1-\eta_{N-i+m})\leqq w_{12}\eta_{N-i+m}$ and choose the action $a_2$ otherwise.

So far we have given formally the optimal decision procedure for the dichotomous case of the truncated sequential problem with delayed observations. One of the main purposes of this section, however, is to give a concrete characterization of the optimal sequential decision procedure. We will see later that the optimal stopping rules for some non-truncated sequential decision problems with delayed observations are completely characterized by the sequential probability ratio test just as the one for the sequential problem without delayed observation is characterized. For this purpose the following three propositions will be used. The proofs of these three will be given in the succeeding section.

PROPOSITION I. The function $\varphi_0(\eta)$ defined by (3.6) is a concave function defined over the closed interval $[0, 1]$ and $\varphi_0(0)=\varphi_0(1)=0$ (for the proof, see the theorem 4.1).

PROPOSITION II. For $i$ $1\leqq i\leqq N$, $E\varphi_{i-1}[\boldsymbol{\eta}(\eta; X)]$ is a concave function of $\eta$ $0\leqq\eta\leqq1$, where $X$ is a random variable with the density function $\eta p_{\theta_1}(x)+(1-\eta)p_{\theta_2}(x)$. Further, for any $i$, $1\leqq i\leqq N$, define the set of $\eta$, $H_i=\{\eta\mid\varphi_0(\eta)\leqq c+E\varphi_{i-1}[\boldsymbol{\eta}(\eta; X)],\ 0\leqq\eta\leqq1\}$. Then, it is shown that $H_i$ consists of mutually disjoint subsets such that

(a) $$H_i=\underline{H}_i+\overline{H}_i$$

$$\underline{H}_i=\sum_{j=1}^{J_i}\underline{H}_{ij}\qquad \overline{H}_i=\sum_{k=1}^{K_i}\overline{H}_{ik}$$

where $\underline{H}_i \cap \overline{H}_i = \phi$ and $\{\underline{H}_{ij}\}$ or $\{\overline{H}_{ik}\}$ is a set of mutually disjoint subsets $\underline{H}_{ij}$ or $\overline{H}_{ik}$ of $[0, 1]$

(b) each $\underline{H}_{ij}$ or $\overline{H}_{ik}$ is connected and closed,

(c) for any positive integers $j$ and $j'$ with $1 \leqq j < j' \leqq J_i$, we have $\eta < \eta'$ for any $\eta \in \underline{H}_{ij}$ and any $\eta' \in \underline{H}_{ij'}$

(d) in the similar way, for any positive integers $k$, $k'$ $(1 \leqq k < k' \leqq K_i)$, we have $\eta' < \eta$ for any $\eta' \in \overline{H}_{ik'}$ and any $\eta \in \overline{H}_{ik}$,

(e) $\underline{H}_{i1} \ni 0$ and $\overline{H}_{i1} \ni 1$,

(f) for any positive integer $i$ and $i'$ with $0 \leqq i < i' \leqq N$,

$$\underline{H}_{i'} \subseteqq \underline{H}_i \quad \text{and} \quad \overline{H}_{i'} \subseteqq \overline{H}_i . \tag{3.16}$$

Now, $\underline{H}_{i1}$ and $\overline{H}_{i1}$ can be represented by closed intervals $[0, \underline{\eta}_i]$ and $[\overline{\eta}_i, 1]$, respectively. Then we can see that the sequences $\{\underline{\eta}_i\}$ and $\{\overline{\eta}_i\}$ $i=1, 2, \cdots, N$ are monotone decreasing and increasing, respectively. If we are concerned with the non-truncated case of the above problem, we will have infinite sequences $\{\underline{\eta}_i\}$ and $\{\overline{\eta}_i\}$ $i=1, 2, \cdots$. From the monotonicity and boundedness of these, we can see that there exist two limits:

$$\lim_{i\to\infty} \underline{\eta}_i = \underline{\eta} \quad \text{and} \quad \lim_{i\to\infty} \overline{\eta}_i = \overline{\eta} . \tag{3.17}$$

*Remark 3.2.* If we are concerned with a dichotomous sequential decision problem without any delayed observation, that is, $m=0$, then we always have $J_i = K_i = 1$ for $i=0, 1, \cdots, N$, while in this case the function $\varphi_0(\eta)$ is defined simply by $\min\{w_{21}(1-\eta), w_{12}\eta\}$ instead of the form given in (3.6).

The optimal stopping rule for our truncated sequential decision problem with delayed observations is completely characterized by $\{H_i\}$ $i=1, 2, \cdots, N$. Let us assume that the prior probability of $\theta_1$ is $\eta_0$. The optimal stopping rule is described as follows:

[0, i] If $\eta_0 \in H_N$, no sample is taken except $X_1, \cdots, X_m$;
[0, ii] if $\eta_0 \notin H_N$, $X_{m+1}$ is taken as a sample to be added to $X_1, \cdots, X_m$.
In case [0, i] occurs, we wait until the observations on $X_1, \cdots, X_m$ are completed and then compute $\boldsymbol{\eta}(\eta_0; x_1, \cdots, x_m) = \eta_m$ for the observed values $x$'s of $X$'s. The decision $a_1$ is chosen if $w_{21}(1-\eta_m) \leqq w_{12}\eta_m$ and the decision $a_2$ otherwise. On the other hand, if the case [0, ii] occurred and the observed value, say $x_1$, of $X_1$ has obtained, $\boldsymbol{\eta}(\eta_0; x_1) = \eta_1$ is computed. Again, two cases occur:

[1, i] If $\eta_1 \in H_{N-1}$, no additional sample is taken.
[1, ii] if $\eta_1 \notin H_{N-1}$, $X_{m+2}$ is taken as an additional sample to the sample $X_2, \cdots, X_{m+1}$ which have already been chosen. In the case [1, i] we

compute $\eta_{m+1}=\boldsymbol{\eta}(\eta_0; x_1, \cdots, x_{m+1})$, using the observed values $x_1, \cdots, x_{m+1}$ of $X_1, \cdots, X_{m+1}$ and choose the terminal decision $a_1$ if $w_{21}(1-\eta_{m+1})\leqq w_{12}\eta_{m+1}$ and the terminal decision $a_2$ otherwise. On the other hand, in case [1, ii] occurs, $\eta_2=\boldsymbol{\eta}(\eta_1; x_2)$ is determined after the observation on $X_2$ is made, with the observed value $x_2$. This value of $\eta_2$ is used to decide whether an additional sample $X_{m+3}$ should be taken or not. In general we can write the procedure as follows: When we have already observed $X_1, \cdots, X_i$ with observed values $x_1, \cdots, x_i$ $(i=0, \cdots, N-1)$, we first compute $\eta_i=\boldsymbol{\eta}(\eta_0; x_1, \cdots, x_i)$ or $\boldsymbol{\eta}(\eta_{i-1}; x_i)$ and get the following rule:

[$i$, i] If $\eta_i \in H_{N-i}$, no additional sample is taken,
[$i$, ii] if $\eta_i \notin H_{N-i}$, $X_{i+m+1}$ is taken as an additional sample to the sample already chosen. In case [$i$, i] occurs, the terminal decision $a_1$ or $a_2$ is chosen according to $w_{21}(1-\eta_{i+m})\leqq w_{12}\eta_{i+m}$ or $w_{21}(1-\eta_{i+m})>w_{12}\eta_{i+m}$, where $\eta_{i+m}=\boldsymbol{\eta}(\eta_0; x_1, \cdots, x_{i+m})$ or $\boldsymbol{\eta}(\eta_{i+m-1}; x_{i+m})$ for observed values $x_i$'s of $X_i$'s. On the contrary, if the case [$i$, ii] occurs, $\eta_{i+1}=\boldsymbol{\eta}(\eta_i; x_{i+1})$ is computed with the observed value $x_{i+1}$ of $X_{i+1}$ and we can proceed with this $\eta_{i+1}$ to the next stage where $i$ is replaced by $i+1$ in the above procedure. Finally, if we have observed $X_1, \cdots, X_N$, we wait until the additional sample $X_{N+1}, \cdots, X_{N+m}$ is observed. Computing $\eta_{N+m}=\boldsymbol{\eta}(\eta_N; x_{N+1}, \cdots, x_{N+m})$, we can choose the optimal terminal decision by the procedure: Choose $a_1$ if $w_{21}(1-\eta_{N+m})\leqq w_{12}\eta_{N+m}$ and $a_2$ otherwise.

Thus we have shown that the optimal decision procedure for the truncated sequential problem with delayed observations is completely characterized by the sequence of subsets of [0, 1], $\{H_i\}$, $i=1, 2, \cdots, N$.

Let us now turn to the non-truncated case. In this case the optimal decision procedure takes a simpler form. Let us define $\underline{H}=\bigcap_{i=1}^{\infty}\underline{H}_i$ and $\overline{H}=\bigcap_{i=1}^{\infty}\overline{H}_i$. Then, the optimal (Bayes) decision procedure for the non-truncated sequential problem is described as follows: When we have already observed $X_1, \cdots, X_i$ with observed values $x_1, \cdots, x_i$ $(i=0, 1, 2, \cdots)$, we first compute $\eta_i=\boldsymbol{\eta}(\eta_0; x_1, \cdots, x_i)$ or $\boldsymbol{\eta}(\eta_{i-1}; x_i)$ and get the following rule:

[$i$, i] If $\eta_i \in \underline{H}\cup\overline{H}$, no additional sample is taken,
[$i$, ii] if $\eta_i \notin \underline{H}\cup\overline{H}$, the random variable $X_{i+m+1}$ is taken as an additional sample to the samples $X_{i+1}, \cdots, X_{i+m}$ which have already been chosen, but not yet observed. The way to choose a suitable terminal decision in the case [$i$, i] and the way to move to the $i+1$st stage in the case [$i$, ii] are the same as for the truncated case except that we always use $\underline{H}\cup\overline{H}$ instead of $\{H_i\}$, $i=1, 2, \cdots, N$. Of course no truncation is considered in the present case.

PROPOSITION III. If there exist positive numbers $e_1$ and $e_2$ which satisfy $e_1 \leqq p_{\theta_1}(x)/p_{\theta_2}(x) \leqq e_2$ for any $x$ such that $p_{\theta_2}(x)>0$, then there exists a positive number $c^*$ such that, for any positive $c \leqq c^*$, the optimal stopping rule of the non-truncated sequential decision problems with delayed observations with the cost $c$ per each observation is completely characterized by the sequential probability ratio test.

Under the condition of the proposition III, we will show in the following section that $\underline{H}=[0, \underline{\eta}]$ and $\overline{H}=[\overline{\eta}, 1]$ where $\underline{\eta}$ and $\overline{\eta}$ are defined by (3.19). Then, the proposition III easily follows if we rewrite the conditions: $0 \leqq \eta_i \leqq \underline{\eta}$, $\overline{\eta} \leqq \eta_i \leqq 1$, $\underline{\eta}<\eta_i<\overline{\eta}$, $w_{21}(1-\eta_{i+m}) \leqq w_{12}\eta_{i+m}$ and $w_{21}(1-\eta_{i+m})>w_{12}\eta_{i+m}$, $i=0, 1, 2, \cdots$. In fact, defining $l_i(x_1, \cdots, x_i)=\prod_{j=1}^{i} p_{\theta_1}(x_j) \Big/ \prod_{j=1}^{i} p_{\theta_2}(x_j)$ and considering the definition of $\eta_i$, we have the above conditions:

$$l_i(x_1, \cdots, x_i) \leqq \frac{\underline{\eta}(1-\eta_0)}{(1-\underline{\eta})\eta_0}, \qquad l_i(x_1, \cdots, x_i) \geqq \frac{\overline{\eta}(1-\eta_0)}{(1-\overline{\eta})\eta_0}$$

$$\frac{\underline{\eta}(1-\eta_0)}{(1-\underline{\eta})\eta_0} < l_i(x_1, \cdots, x_i) < \frac{\overline{\eta}(1-\eta_0)}{(1-\overline{\eta})\eta_0}, \qquad l_{i+m}(x_1, \cdots, x_{i+m}) \geqq \frac{w_{21}(1-\eta_0)}{w_{12}\eta_0}$$

and

$$l_{i+m}(x_1, \cdots, x_{i+m}) < \frac{w_{21}(1-\eta_0)}{w_{12}\eta_0}, \qquad \text{respectively.}$$

In conclusion, we should emphasize that, at least when the condition of the proposition III is satisfied, the Bayes stopping rule as well as the Bayes terminal decision procedure for the non-truncated dichotomous sequential decision problem with delayed observations are completely characterized by the sequential probability ratio just as the Bayes decision procedure for the non-truncated sequential dichotomous decision problem without delayed observation is characterized by the sequential probability ratio test.

## 4. Theorems and lemmas

In this section lemmas and theorems are proved to establish Propositions I, II and III which were stated in section 3. In what follows, the sample space or the product sample space will be simply denoted by $R$ or $R^m$.

THEOREM 4.1. *The function $\varphi_0(\eta)$ defined by (3.6) is a concave function over the closed interval $[0, 1]$ and $\varphi_0(0)=\varphi_0(1)=0$.*

PROOF. By the definition of $\varphi_0(\eta)$, for $\eta \in [0, 1]$

$$(4.1) \qquad \varphi_0(\eta)=E \min \{w_{21}[1-\boldsymbol{\eta}(\eta;X_1,\cdots,X_m)],\ w_{12}\boldsymbol{\eta}(\eta;X_1,\cdots,X_m)\}$$

where

$$(4.2) \qquad \boldsymbol{\eta}(\eta;X_1,\cdots,X_m)=\frac{\eta p_{\theta_1}(X_1,\cdots,X_m)}{\eta p_{\theta_1}(X_1,\cdots,X_m)+(1-\eta)p_{\theta_2}(X_1,\cdots,X_m)}$$

and

$$p_{\theta_i}(X_1,\cdots,X_m)=\prod_{j=1}^{m} p_{\theta_i}(X_j) \qquad i=1,2\ .$$

Since the joint probability density function of $X_i$'s for a prior probability of $\theta_1$, say $\eta$, is $\eta p_{\theta_1}(x_1,\cdots,x_m)+(1-\eta)p_{\theta_2}(x_1,\cdots,x_m)$, we have

$$(4.3) \qquad \begin{aligned}\varphi_0(\eta)&=\int_{R^m} \min \{w_{21}[1-\boldsymbol{\eta}(\eta;x_1,\cdots,x_m)],\ w_{21}\boldsymbol{\eta}(\eta;x_1,\cdots,x_m)\}\\ &\qquad\times[\eta p_{\theta_1}(x_1,\cdots,x_m)+(1-\eta)p_{\theta_2}(x_1,\cdots,x_m)]\prod_{i=1}^{m}d\mu(x_i)\\ &=\int_{R^m}\min\{w_{21}(1-\eta)p_{\theta_2}(x_1,\cdots,x_m),\ w_{12}\eta p_{\theta_1}(x_1,\cdots,x_m)\}\prod_{i=1}^{m}d\mu(x_i)\end{aligned}$$

For any $\eta_1, \eta_2$ and $\lambda$ such that $0\leqq\eta_1,\ \eta_2,\ \lambda\leqq 1$, we have

$$\begin{aligned}\varphi_0(\lambda\eta_1+(1-\lambda)\eta_2)&=\int_{R^m}\min\{w_{21}[1-\lambda\eta_1-(1-\lambda)\eta_2]p_{\theta_2}(x_1,\cdots,x_m),\\ &\qquad w_{12}[\lambda\eta_1+(1-\lambda)\eta_2]p_{\theta_1}(x_1,\cdots,x_m)\}\prod_{i=1}^{m}d\mu(x_i)\\ &=\int_{R^m}\min\{w_{21}[\lambda(1-\eta_1)+(1-\lambda)(1-\eta_2)]p_{\theta_2}(x_1,\cdots,x_m),\\ &\qquad w_{12}[\lambda\eta_1+(1-\lambda)\eta_2]p_{\theta_1}(x_1,\cdots,x_m)\}\prod_{i=1}^{m}d\mu(x_i)\\ &\geqq\int_{R^m}[\min\{w_{21}\lambda(1-\eta_1)p_{\theta_2}(x_1,\cdots,x_m),\ w_{12}\lambda\eta_1 p_{\theta_1}(x_1,\cdots,x_m)\}\\ &\qquad+\min\{w_{21}(1-\lambda)(1-\eta_2)p_{\theta_2}(x_1,\cdots,x_m),\\ &\qquad w_{12}(1-\lambda)\eta_2 p_{\theta_2}(x_1,\cdots,x_m)\}]\prod_{i=1}^{m}d\mu(x_i)\\ &=\lambda\varphi_0(\eta_1)+(1-\lambda)\varphi_0(\eta_2)\ .\end{aligned}$$

*Remark 4.1.* For the proof of the above theorem, the independence of $X_1,\cdots,X_m$ is not necessary. This theorem gives the postulate I in section 3.

LEMMA 4.1. *If $\varphi(\eta)$ is an arbitrary concave function defined on $[0, 1]$, then we have, for any positive integer $k$,*

$$(4.4) \qquad \varphi(\eta)\geqq E\varphi[\boldsymbol{\eta}(\eta;X_1,\cdots,X_k)]$$

*where the expectation is taken under the assumption that the k-dimensional random variable* $(X_1, \cdots, X_k)$ *has a probability density function* $\eta p_{\theta_1}(x_1, \cdots, x_k)+(1-\eta)p_{\theta_2}(x_1, \cdots, x_k)$. *The strict inequality holds when the function* $\varphi(\eta)$ *is strictly concave and the probability distribution of* $(X_1, \cdots, X_k)$ *does not concentrate on one point.*

PROOF. From the concavity of $\varphi(\eta)$, it is shown that, for any $\eta_0 \in [0, 1]$, there exists a straight line through the point $(\eta_0, \varphi(\eta_0))$ such as the curve $\{(\eta, \varphi(\eta)), 0 \leqq \eta \leqq 1\}$ lies beneath the straight line. Suppose that its equation is $y-\varphi(\eta_0)=a(\eta-\eta_0)$. Then, we have

$$(4.5) \qquad y \geqq \varphi(\eta) \quad \text{or} \quad \varphi(\eta_0)+a(\eta-\eta_0) \geqq \varphi(\eta) .$$

If we define a random variable $Y$ by the equation

$$(4.6) \qquad Y=\varphi(\eta_0)-a\eta_0+a\boldsymbol{\eta}(\eta_0; X_1, \cdots, X_k) ,$$

then by (4.5) we have, with probability one,

$$Y \geqq \varphi[\boldsymbol{\eta}(\eta_0; X_1, \cdots, X_k)]$$

or

$$(4.7) \qquad \varphi(\eta_0)-a\eta_0+a\boldsymbol{\eta}(\eta_0; X_1, \cdots, X_k) \geqq \varphi[\boldsymbol{\eta}(\eta_0; X_1, \cdots, X_k)] .$$

Therefore, taking the expectation of both sides of (4.7), we obtain

$$(4.8) \qquad \varphi(\eta_0)-a\eta_0+aE\boldsymbol{\eta}(\eta_0; X_1, \cdots, X_k) \geqq E\varphi[\boldsymbol{\eta}(\eta_0; X_1, \cdots, X_k)]$$

and since $E\boldsymbol{\eta}(\eta_0; X_1, \cdots, X_k)=\eta_0$ we finally have the inequality (4.4). The strict concavity insures the strict inequality in (4.5) and (4.7). Thus, the last statement of the lemma is valid. Before going to the theorem 4.2, let us state its content in a restrictive form, which is the following lemma.

Lemma 4.2. *If* $\varphi(\eta)$ *is a concave function defined on* $[0, 1]$ *with the continuous second derivative* $\varphi''(\eta)$ *on* $[0, 1]$(*), *then the function* $E\varphi[\boldsymbol{\eta}(\eta; X_1, \cdots, X_k)]$ *for any positive integer k is a concave function in* $\eta$ *defined on* $[0, 1]$, *where the expectation is taken under the assumption that the k-dimensional random variable* $(X_1, \cdots, X_k)$ *has the joint probability density function* $\eta p_{\theta_1}(x_1, \cdots, x_k)+(1-\eta)p_{\theta_2}(x_1, \cdots, x_k)$ *for each* $\eta$.

PROOF. Let us designate $E\varphi[\boldsymbol{\eta}(\eta; X_1, \cdots, X_k)]$ by $\Phi(\eta)$. Since we can write

$$(4.9) \qquad \Phi(\eta)=\int_{R^k} \varphi[\boldsymbol{\eta}(\eta; x_1, \cdots, x_k)][\eta p_{\theta_1}(x_1, \cdots, x_k) + (1-\eta)p_{\theta_2}(x_1, \cdots, x_k)] \prod_{i=1}^{k} d\mu(x_i)$$

---

(*) $\varphi''(\eta)$ for $\eta=0$ and 1 should be interpreted as $[D^+(D^+\varphi(\eta))]_{\eta=0}$ and $[D^-(D^-\varphi(\eta))]_{\eta=1}$.

it is seen that the derivative of $\Phi(\eta)$ exists and is obtained by the differentiation of (4.9) under the integral sign, which is allowed in the present case. Thus we have

$$\Phi'(\eta)=\int_{R^k}\Big\{\varphi'[\boldsymbol{\eta}(\eta;x_1,\cdots,x_k)]\left[\frac{d}{d\eta}\boldsymbol{\eta}(\eta;x_1\cdots,x_k)\right][\eta p_{\theta_1}(x_1,\cdots,x_k)+(1-\eta)p_{\theta_2}(x_1,\cdots,x_k)]+\varphi[\boldsymbol{\eta}(\eta;x_1,\cdots,x_k)]\times[p_{\theta_1}(x_1,\cdots,x_k)-p_{\theta_2}(x_1,\cdots,x_k)]\Big\}\prod_{i=1}^{k}d\mu(x_i)\,.$$

Now, from the definition of $\boldsymbol{\eta}[\cdot]$, we have

$$\text{(4.10)}\quad \frac{d}{d\eta}\boldsymbol{\eta}[\eta;x_1,\cdots,x_k]=\frac{p_{\theta_1}(x_1,\cdots,x_k)p_{\theta_2}(x_1,\cdots,x_k)}{[\eta p_{\theta_1}(x_1,\cdots,x_k)+(1-\eta)p_{\theta_2}(x_1,\cdots,x_k)]^2}\,.$$

Therefore, we obtain

$$\text{(4.11)}\quad \Phi'(\eta)=\int_{R^k}\Big\{\varphi'[\boldsymbol{\eta}(\eta;x_1,\cdots,x_k)]\frac{p_{\theta_1}(x_1,\cdots,x_k)p_{\theta_2}(x_1,\cdots,x_k)}{\eta p_{\theta_1}(x_1,\cdots,x_k)+(1-\eta)p_{\theta_2}(x_1,\cdots,x_k)}+\varphi[\boldsymbol{\eta}(\eta;x_1,\cdots,x_k)][p_{\theta_1}(x_1,\cdots,x_k)-p_{\theta_2}(x_1,\cdots,x_k)]\Big\}\prod_{i=1}^{k}d\mu(x_i)\,.$$

Again, by the assumption that $\varphi''(\eta)$ is continuous on [0, 1], we can obtain $\Phi''(\eta)$ by the integration of the derivative of the integrand of (4.11). That is

$$\text{(4.12)}\quad \Phi''(\eta)=\int_{R^k}\varphi''[\boldsymbol{\eta}(\eta;x_1,\cdots,x_k)]\times\frac{(p_{\theta_1}(x_1,\cdots,x_k)p_{\theta_2}(x_1,\cdots,x_k))^2}{[\eta p_{\theta_1}(x_1,\cdots,x_k)+(1-\eta)p_{\theta_2}(x_1,\cdots,x_k)]^3}\prod_{i=1}^{k}d\mu(x_i)\,.$$

Since $\varphi''(\eta)\leqq 0$ for $\eta\in(0,1)$, it follows that $\Phi''(\eta)\leqq 0$ for $\eta\in(0,1)$. Thus the lemma has been proved.

*Remark 4.2.* If $\varphi''(\eta)<0$ for $\eta\in(0,1)$ and if the $k$-dimensional $\mu$-measure of the set $\{(x_1,\cdots,x_k)\,|\,p_{\theta_1}(x_1,\cdots,x_k)\,p_{\theta_2}(x_1,\cdots,x_k)>0\}$ is not zero, then we have $\Phi''(\eta)<0$ for $\eta\in(0,1)$.

The concavity of $E\varphi[\boldsymbol{\eta}(\eta;X_1,\cdots,X_k)]$ can also be established when the concavity of $\varphi(\eta)$ alone is assumed. For this purpose it seems convenient to use the following lemma.

LEMMA 4.3. *A necessary and sufficient condition for a function $f(x)$ to be a concave function defined on [0, 1] is :*

(1) *$f(x)$ is continuous on [0, 1].*

(2) *$D^-f(x)$*[(*)] *and* *$D^+f(x)$*[(*)] *exist for* $x \in (0, 1)$. *Further,* $D^-f(x)$ *exists or* $-\infty$ *at* $x=1$ *and* $D^+f(x)$ *exists or* $+\infty$ *at* $x=0$.

(3) $D^-f(x) \geqq D^+f(x)$ *for any* $x \in (0, 1)$.

*and*

(4) $D^+f(x_1) \geqq D^-f(x_2)$ *for any pair* $x_1, x_2$ *such that* $x_1 < x_2$ *and* $x_1, x_2 \in [0, 1]$.

PROOF.

(i) The proof of necessity

For any $x \in (0, 1)$, consider a monotone increasing sequence $\{y_i\}$ such that $y_i \in [0, 1]$ $i=1, 2, \cdots$ and $\lim_{i\to\infty} y_i = x$, and also a monotone decreasing sequence $\{z_j\}$ such that $z_j \in [0, 1]$ $j=1, 2, \cdots$ and $\lim_{j\to\infty} z_j = x$. Since $f(x)$ is concave, we have

$$f(x) \geqq \frac{z_j - x}{z_j - y_i} f(y_i) + \frac{x - y_i}{z_j - y_i} f(z_j) \tag{4.13}$$

or

$$\frac{f(x) - f(y_i)}{x - y_i} \geqq \frac{f(z_j) - f(y_i)}{z_j - y_i} \geqq \frac{f(z_j) - f(x)}{z_j - x} \tag{4.14}$$

and the left and right terms of (4.14) are monotone decreasing as $y_i \to x$ and monotone increasing as $z_j \to x$, respectively. Thus, the existence of $D^-f(x)$ and $D^+f(x)$, and the continuity of $f(x)$, that is, (1) and the first past of (2) immediately follow. Also, the inequality (3) is an immediate result. The proof of (4) runs as follows. For any pair $x_1, x_2$ ($x_1 < x_2$, $x_1, x_2 \in [0, 1]$), let us choose $y$ and $z$ such that $x_1 < y$, $z < x_2$. Then, from the concavity of $f(x)$ we have

$$\frac{f(y) - f(x_1)}{y - x_1} \geqq \frac{f(x_2) - f(x_1)}{x_2 - x_1} \geqq \frac{f(x_2) - f(z)}{x_2 - z}. \tag{4.15}$$

Now, since the left term is monotone increasing as $y \to x_1 + 0$ and the right term is monotone decreasing as $z \to x_2 - 0$, we get

$$D^+f(x_1) \geqq \frac{f(x_2) - f(x_1)}{x_2 - x_1} \geqq D^-f(x_2) \tag{4.16}$$

which was to be proved.

---

(*) $D^-f(x)$ and $D^+f(x)$ are defined by

$$D^-f(x) = \lim_{h\to -0} \frac{f(x+h) - f(x)}{h} \qquad D^+f(x) = \lim_{h\to +0} \frac{f(x+h) - f(x)}{h}.$$

(ii) The proof of sufficiency

From the conditions (3) and (4), we obtain

$$(4.17)\qquad D^+f(x_1)\geqq D^-f(y_1)\geqq D^+f(y_1)\geqq D^-f(y_2)\geqq D^+f(y_2)\geqq D^-f(x_2)$$

where $0\leqq x_1<y_1<y_2<x_2\leqq 1$. Therefore, both $D^-f(y)$ and $D^+f(y)$ are monotone decreasing functions defined over $(0, 1]$ and $[0, 1)$, respectively. Hence, they have at most a countable number of discontinuity points. It is easily shown that $D^-f(y)$ and $D^+f(y)$ are left-continuous and right-continuous, respectively. From this fact and the inequalities (4.17), we can conclude that $D^-f(y)=D^+f(y)$ except at most a countable number of points in $[0, 1]$. Therefore, for any pair $x_1, x_2$ with $0<x_1<x_2<1$, we have

$$(4.18)\qquad f(x_2)=f(x_1)+\int_{x_1}^{x_2} D^{\pm}f(x)\,dx \qquad \text{(Lebesgue integral)}\,.$$

On the other hand, as is easily seen, it holds that

$$(4.19)\qquad (x_2-x_1)D^-f(x_2)\leqq\int_{x_1}^{x_2} D^{\pm}f(x)\,dx\leqq(x_2-x_1)D^+f(x_1)\,.$$

From (4.18) and (4.19), we have

$$(4.20)\qquad D^-f(x_2)\leqq\frac{f(x_2)-f(x_1)}{x_2-x_1}\leqq D^+f(x_1)\,.$$

By (4.17) and (4.20), for any $x$ such that $0\leqq x_1<x<x_2\leqq 1$, we get

$$(4.21)\qquad \frac{1}{x_2-x}\int_{x}^{x_2} D^{\pm}f(x)\,dx\leqq D^+f(x)\leqq\frac{f(x)-f(x_1)}{x-x_1}$$

since $D^+f(x)$ and $D^-f(x)$ are monotone decreasing. Finally, from (4.18) and (4.21), we obtain

$$(4.22)\qquad \frac{f(x_2)-f(x)}{x_2-x}\leqq\frac{f(x)-f(x_1)}{x-x_1}\qquad \text{for}\quad 0\leqq x_1<x<x_2\leqq 1\,,$$

which implies the concavity of $f(x)$. The extension of lemma 4.2 is now given in the following theorem and its proof is performed by applying lemma 4.3.

THEOREM 4.2. *If $\varphi(\eta)$ is a concave function defined on $[0, 1]$ and $D^+\varphi(0)$ and $D^-\varphi(1)$ are finite, then, for any positive integer $k$, $E\varphi[\boldsymbol{\eta}(\eta; X)]$ is also a concave function defined on $[0, 1]$, where $X=(X_1\cdots, X_k)$ is considered to be a $k$-dimensional random variable with a probability density function $\eta p_{\theta_1}(x)+(1-\eta)p_{\theta_2}(x)$; $p_{\theta_i}(x)$ is the joint probability density function of $X_1, \cdots, X_k$ with respect to a $\sigma$-finite measure $\mu$ over $R^k$*

*when $\theta_i$ is the true parameter, $i=1, 2$. Further, $D^+E\varphi[\boldsymbol{\eta}(\eta; X)]$ and $D^-E\varphi[\boldsymbol{\eta}(\eta; X)]$ take the forms given by (4.27) and (4.28), respectively.*

PROOF. Let us denote $E\varphi[\boldsymbol{\eta}(\eta; X)]$ by $\Phi(\eta)$. $\Phi(\eta)$ is expressed as

$$\Phi(\eta)=\int_{R^k}\varphi[\boldsymbol{\eta}(\eta; x)][\eta p_{\theta_1}(x)+(1-\eta)p_{\theta_2}(x)]\,d\mu(x)\,. \tag{4.23}$$

To assert that $\Phi(\eta)$ is concave, we shall prove that the conditions (1), (2), (3) and (4) of lemma 4.3 are satisfied by $\Phi(\eta)$. The condition (1) is obviously satisfied since functions $\varphi(\eta)$ and $\boldsymbol{\eta}(\eta; x)$ are continuous in $\eta$ and the integral of (4.23) is uniformly convergent, as is easily seen. Let us prove the condition (2). If $0\leqq\eta$, $\eta+\Delta\leqq 1$, we have

$$\begin{aligned}\frac{\Phi(\eta+\Delta)-\Phi(\eta)}{\Delta}=\int_{R^k}\Big\{&\frac{\varphi[\boldsymbol{\eta}(\eta+\Delta; x)]-\varphi[\boldsymbol{\eta}(\eta; x)]}{\Delta}\\ &\times[\eta p_{\theta_1}(x)+(1-\eta)p_{\theta_2}(x)]+\varphi[\boldsymbol{\eta}(\eta+\Delta; x)](p_{\theta_1}(x)-p_{\theta_2}(x))\Big\}\,d\mu(x)\,.\end{aligned} \tag{4.24}$$

Now,

$$\begin{aligned}&\boldsymbol{\eta}(\eta+\Delta; x)-\boldsymbol{\eta}(\eta; x)\\ &\quad=\frac{p_{\theta_1}(x)p_{\theta_2}(x)\Delta}{[(\eta+\Delta)p_{\theta_1}(x)+(1-\eta-\Delta)p_{\theta_2}(x)]\,[\eta p_{\theta_1}(x)+(1-\eta)p_{\theta_2}(x)]}\,.\end{aligned} \tag{4.25}$$

Therefore, for any $x$ such that $p_{\theta_1}(x)p_{\theta_2}(x)>0$, $\Delta\gtrless 0$ implies

$$\boldsymbol{\eta}(\eta+\Delta; x)-\boldsymbol{\eta}(\eta; x)\gtrless 0 \tag{4.26}$$

and the derivative of $\boldsymbol{\eta}(\eta; x)$ in $\eta$ is given by (4.10). Since, by the assumption that $D^+\varphi(0)$ and $D^-\varphi(1)$ are finite, the integral and $\lim_{\Delta\to+0}$ or $\lim_{\Delta\to-0}$ can be exchanged for (4.24), we have

$$\begin{aligned}D^+\Phi(\eta)=\int_{R^k}\Big\{&D^+\varphi[\boldsymbol{\eta}(\eta; x)]\frac{p_{\theta_1}(x)p_{\theta_2}(x)}{[\eta p_{\theta_1}(x)+(1-\eta)p_{\theta_2}(x)]}\\ &+\varphi[\boldsymbol{\eta}(\eta; x)](p_{\theta_1}(x)-p_{\theta_2}(x))\Big\}\,d\mu(x)\end{aligned} \tag{4.27}$$

for $\eta\in[0, 1)$ and for $\eta\in(0, 1]$

$$\begin{aligned}D^-\Phi(\eta)=\int_{R^k}\Big\{&D^-\varphi[\boldsymbol{\eta}(\eta; x)]\frac{p_{\theta_1}(x)p_{\theta_2}(x)}{[\eta p_{\theta_1}(x)+(1-\eta)p_{\theta_2}(x)]}\\ &+\varphi[\boldsymbol{\eta}(\eta; x)](p_{\theta_1}(x)-p_{\theta_2}(x))\Big\}\,d\mu(x)\,.\end{aligned} \tag{4.28}$$

Thus, the condition (2) has been proved. Since $\varphi(\eta)$ is concave, it holds that for any $x\in R^k$, $D^+\varphi[\boldsymbol{\eta}(\eta; x)]\leqq D^-\varphi[\boldsymbol{\eta}(\eta; x)]$. Thus, from (4.27) and

(4.28) we obtain condition (3), that is, $D^+\Phi(\eta)\leqq D^-\Phi(\eta)$ for $\eta\in(0,1)$. Finally we wish to prove that the condition (4) of lemma 4.3 is satisfied. Suppose $\eta_1<\eta_2$ and $\eta_1, \eta_2\in[0,1]$. Then, by (4.27) and (4.28) we have

$$\begin{aligned}(4.29)\qquad & \mathrm{D}^+\Phi(\eta_1)-D^-\Phi(\eta_2)\\ &=\int_{R^k}\Big\{\Big[\frac{D^+\varphi[\boldsymbol{\eta}(\eta_1;x)]}{\eta_1 p_{\theta_1}(x)+(1-\eta_1)p_{\theta_2}(x)}-\frac{D^-\varphi[\boldsymbol{\eta}(\eta_2;x)]}{\eta_2 p_{\theta_1}(x)+(1-\eta_2)p_{\theta_2}(x)}\Big]p_{\theta_1}(x)p_{\theta_2}(x)\\ &\quad-\Big[\varphi[\boldsymbol{\eta}(\eta_2;x)]-\varphi[\boldsymbol{\eta}(\eta_1;x)]\Big](p_{\theta_1}(x)-p_{\theta_2}(x))\Big\}\,d\mu(x)\ .\end{aligned}$$

For an arbitrary $x\in R^k$ such that $p_{\theta_1}(x)p_{\theta_2}(x)>0$, we have $\boldsymbol{\eta}(\eta_2;x)-\boldsymbol{\eta}(\eta_1;x)>0$ by (4.26) since $\eta_2>\eta_1$. For such $x$, the total number of discontinuity points of $D^+\varphi(\eta)$ in the interval $(\boldsymbol{\eta}(\eta_1;x), \boldsymbol{\eta}(\eta_2;x))$ is at most countablly infinite since the function $D^+\varphi(\eta)$ is monotone decreasing on $[0,1]$. Consider first the case where the discontinuity points are finite and denote them by $\{\zeta_i\}$ $i=1,2,\cdots,n(\eta_1,\eta_2,x)$ where $n(\eta_1,\eta_2,x)$ is the total number of discontinuity points of $D^+\varphi(\eta)$ in the interval $(\boldsymbol{\eta}(\eta_1;x), \boldsymbol{\eta}(\eta_2;x))$ and $\zeta_i<\zeta_{i+1}$, $i=1,\cdots,n(\eta_1,\eta_2,x)-1$. Also let us denote $n(\eta_1,\eta_2,x)$ simply by $n^*$ and for convenience define $\zeta_0=\boldsymbol{\eta}(\eta_1;x)$ and $\zeta_{n^*+1}=\boldsymbol{\eta}(\eta_2;x)$. Since $D^+\varphi(\eta)$ is continuous in each open subinterval $(\zeta_i,\zeta_{i+1})$, we have, for some $\theta_i\in(0,1)$,

$$(4.30)\qquad \varphi(\zeta_{i+1})-\varphi(\zeta_i)=D^+[\zeta_i+\theta_i(\zeta_{i+1}-\zeta_i)](\zeta_{i+1}-\zeta_i)\ .$$

Thus

$$\begin{aligned}(4.31)\qquad \varphi[\boldsymbol{\eta}(\eta_2;x)]-\varphi[\boldsymbol{\eta}(\eta_1;x)]&=\varphi(\zeta_{n^*+1})-\varphi(\zeta_0)=\sum_{i=0}^{n^*}(\varphi(\zeta_{i+1})-\varphi(\zeta_i))\\ &=\sum_{i=0}^{n^*}D^+[\zeta_i+\theta_i(\zeta_{i+1}-\zeta_i)](\zeta_{i+1}-\zeta_i)\ .\end{aligned}$$

Let us now define a sequence $\{\eta_{(i)}\}$, corresponding to $\{\zeta_i\}$, by the equations

$$(4.32)\qquad \zeta_i=\frac{\eta_{(i)}p_{\theta_1}(x)}{\eta_{(i)}p_{\theta_1}(x)+(1-\eta_{(i)})p_{\theta_2}(x)}\qquad i=0,1,\cdots,n^*+1\ .$$

Clearly, $\eta_{(0)}=\eta_1$ and $\eta_{(n^*+1)}=\eta_2$; in addition

$$(4.33)\qquad \eta_{(0)}<\eta_{(1)}<\cdots\eta_{(n^*+1)}\ .$$

We, then, have

$$\begin{aligned}(4.34)\qquad &\frac{D^+\varphi[\boldsymbol{\eta}(\eta_1;x)]}{\eta_1 p_{\theta_1}(x)+(1-\eta_1)p_{\theta_2}(x)}-\frac{D^-\varphi[\boldsymbol{\eta}(\eta_2;x)]}{\eta_2 p_{\theta_1}(x)+(1-\eta_2)p_{\theta_2}(x)}\\ &=\sum_{i=0}^{n^*}\Big\{\frac{D^+\varphi[\boldsymbol{\eta}(\eta_{(i)};x)]}{\eta_{(i)}p_{\theta_1}(x)+(1-\eta_{(i)})p_{\theta_2}(x)}-\frac{D^-\varphi[\boldsymbol{\eta}(\eta_{(i+1)};x)]}{\eta_{(i+1)}p_{\theta_1}(x)+(1-\eta_{(i+1)})p_{\theta_2}(x)}\Big\}\end{aligned}$$

$$+\sum_{i=1}^{n^*}\frac{D^-\varphi[\boldsymbol{\eta}(\eta_{(i)}\,;x)]-D^+\varphi[\boldsymbol{\eta}(\eta_{(i)}\,;x)]}{\eta_{(i)}p_{\theta_1}(x)+(1-\eta_{(i)})p_{\theta_2}(x)}$$
$$\geqq\sum_{i=0}^{n^*}\left\{\frac{D^+\varphi[\boldsymbol{\eta}(\eta_{(i)}\,;x)]}{\eta_{(i)}p_{\theta_1}(x)+(1-\eta_{(i)})p_{\theta_2}(x)}-\frac{D^-\varphi[\boldsymbol{\eta}(\eta_{(i+1)}\,;x)]}{\eta_{(i+1)}p_{\theta_1}(x)+(1-\eta_{(i+1)})p_{\theta_2}(x)}\right\},$$

where the proper inequality holds if and only if $n^*\geqq 1$, since by the definition of $\eta_{(i)}$ we have

$$D^-\varphi[\boldsymbol{\eta}(\eta_{(i)}\,;x)]-D^+\varphi[\boldsymbol{\eta}(\eta_{(i)}\,;x)]>0\,.$$

Using (4.31), (4.32) and (4.34), we obtain the following inequality concerning the integrand of (4.29):

(4.35) The integrand of (4.29)

$$\geqq p_{\theta_1}(x)p_{\theta_2}(x)\sum_{i=0}^{n^*}\left\{\frac{D^+\varphi[\boldsymbol{\eta}(\eta_{(i)}\,;x)]}{\eta_{(i)}p_{\theta_1}(x)+(1-\eta_{(i)})p_{\theta_2}(x)}\right.$$
$$-\frac{D^-\varphi[\boldsymbol{\eta}(\eta_{(i+1)}\,;x)]}{\eta_{(i+1)}p_{\theta_1}(x)+(1-\eta_{(i+1)})p_{\theta_2}(x)}$$
$$\left.-\frac{D^+\varphi[\boldsymbol{\eta}(\eta_{(i)};x)+\theta_i(\boldsymbol{\eta}(\eta_{(i+1)};x)-\boldsymbol{\eta}(\eta_{(i)};x))](\eta_{(i+1)}-\eta_{(i)})(p_{\theta_1}(x)-p_{\theta_2}(x))}{[\eta_{(i)}p_{\theta_1}(x)+(1-\eta_{(i)})p_{\theta_2}(x)]\,[\eta_{(i+1)}p_{\theta_1}(x)+(1-\eta_{(i+1)})p_{\theta_2}(x)]}\right\}$$

(4.36)

$$=p_{\theta_1}(x)p_{\theta_2}(x)\sum_{i=0}^{n^*}\left\{\frac{D^+\varphi[\boldsymbol{\eta}(\eta_{(i)};x)]-D^-\varphi[\boldsymbol{\eta}(\eta_{(i)};x)+\theta_i(\boldsymbol{\eta}(\eta_{(i+1)};x)-\boldsymbol{\eta}(\eta_{(i)};x))]}{\eta_{(i)}p_{\theta_1}(x)+(1-\eta_{(i)})p_{\theta_2}(x)}\right.$$
$$\left.+\frac{D^+\varphi[\boldsymbol{\eta}(\eta_{(i)};x)+\theta_i(\boldsymbol{\eta}(\eta_{(i+1)};x)-\eta(\eta_{(i)};x))]-D^-\varphi[\boldsymbol{\eta}(\eta_{(i+1)};x)]}{\eta_{(i+1)}p_{\theta_1}(x)+(1-\eta_{(i+1)})p_{\theta_2}(x)}\right\}$$

since it holds that

$$\frac{1}{\eta_{(i+1)}p_{\theta_1}(x)+(1-\eta_{(i+1)})p_{\theta_2}(x)}-\frac{1}{\eta_{(i)}p_{\theta_1}(x)+(1-\eta_{(i)})p_{\theta_2}(x)}$$
$$=\frac{-(\eta_{(i+1)}-\eta_{(i)})(p_{\theta_1}(x)-p_{\theta_2}(x))}{[\eta_{(i)}p_{\theta_1}(x)+(1-\eta_{(i)})p_{\theta_2}(x)]\,[\eta_{(i+1)}p_{\theta_1}(x)+(1-\eta_{(i+1)})p_{\theta_2}(x)]}\,.$$

Noting that $\boldsymbol{\eta}(\eta_{(i)};x)<\boldsymbol{\eta}(\eta_{(i)};x)+\theta_i(\boldsymbol{\eta}(\eta_{(i+1)};x)-\boldsymbol{\eta}(\eta_{(i)};x))<\boldsymbol{\eta}(\eta_{(i+1)};x)$ and applying the condition (4) of lemma 4.3, we get

$$D^+\varphi[\boldsymbol{\eta}(\eta_{(i)};x)]\geqq D^-\varphi[\boldsymbol{\eta}(\eta_{(i)};x)+\theta_i(\boldsymbol{\eta}(\eta_{(i+1)};x)-\boldsymbol{\eta}(\eta_{(i)};x))]$$
$$\geqq D^+\varphi[\boldsymbol{\eta}(\eta_{(i)};x)+\theta_i(\boldsymbol{\eta}(\eta_{(i+1)};x)-\boldsymbol{\eta}(\eta_{(i)};x))]$$
$$\geqq D^-\varphi[\boldsymbol{\eta}(\eta_{(i+1)};x)]\,.$$

These inequalities imply that the right-hand side of (4.36) is non-negative; therefore the integrand of (4.29) is non-negative. Consequently

$D^+\Phi(\eta_1)-D^-\Phi(\eta_2)$ is non-negative. Thus, the condition (4) of lemma 4.3 has been proved for the case where $n(\eta_1, \eta_2, x)$ is finite. When $n(\eta_1, \eta_2, x)$ is infinite, we can prove the same conclusion by a similar procedure, although we then have to consider possible accumulation points of $\{\zeta_i\}$ or $\{\eta_{(i)}\}$, which makes the description of the proof a little more complicated. The proof for this case, however, is omitted in this paper.

*Remark 4.3.* Proposition I of section 3 is a special case of theorem 4.2. In fact, in the proposition $\varphi(\eta)$ is a concave function given by $\min\{w_{21}(1-\eta), w_{12}\eta\}$ $0\leqq\eta\leqq 1$; furthermore, $X_{N+1}, \cdots, X_{N+m}$ are assumed to be independent and identically distributed, which is more restrictive than the assumptions of theorem 4.2.

COROLLARY 4.1. *Under the conditions of the theorem 4.2, we have*

$$D^+\Phi(0)=D^+\varphi(0) \quad \text{and} \quad D^-\Phi(1)=D^-\varphi(1)\,. \tag{4.38}$$

The following corollary is also easily derived from theorem 4.2 and lemma 4.1.

COROLLARY 4.2. *Suppose that $X_1, X_2, \cdots, X_n$ be independent and identically distributed with probability density function $p_{\theta_1}(x)$ or $p_{\theta_2}(x)$ and $\varphi(\eta)$ be a concave function, with finite $D^+\varphi(0)$ and $D^-\varphi(1)$, defined on $[0, 1]$. Then, $E\varphi[\boldsymbol{\eta}(\eta; X_1, \cdots, X_k)]$ $k=1, 2, \cdots, n$ are all concave and $E\varphi[\boldsymbol{\eta}(\eta; X_1, \cdots, X_k)] \geqq E\varphi[\boldsymbol{\eta}(\eta; X_1, \cdots, X_k, X_{k+1})]$ for $k=0, 1, \cdots, n-1$, where we define $\boldsymbol{\eta}(\eta)=\eta$.*

PROOF. Under the condition that $X_1=x_1, \cdots, X_k=x_k$, we have

$$\varphi[\boldsymbol{\eta}(\eta; x_1, \cdots, x_k)]=\varphi(\eta_k)\geqq E\varphi[\boldsymbol{\eta}(\eta_k; X_{k+1})] \qquad \text{(lemma 4.1)}$$

where $\eta_k=\boldsymbol{\eta}(\eta; x_1, \cdots, x_k)$. Then, taking expectations of both sides with respect to $X_1, \cdots, X_k$, we obtain

$$E\varphi[\boldsymbol{\eta}(\eta; X_1, \cdots, X_k)]\geqq E\varphi[\boldsymbol{\eta}(\eta; X_1, \cdots, X_k, X_{k+1})]\,.$$

*Remark 4.4.* Proposition II of section 3 is obtained from the above corollary. For the derivation of the proposition, geometrical or graphical considerations are quite useful.

Let us now consider $\varphi(\eta)=\min\{w_{21}(1-\eta), w_{12}\eta\}$ $0\leqq\eta\leqq 1$ and the corresponding function $\Phi(\eta)$ which is defined by (4.23). Since

$$D^+\varphi(\eta)=\begin{cases} w_{12} & \text{if} \quad 0\leqq\eta<\dfrac{w_{21}}{w_{21}+w_{12}} \\[2ex] -w_{21} & \text{if} \quad \dfrac{w_{21}}{w_{21}+w_{12}}\leqq\eta<1 \end{cases}$$

$$D^-\varphi(\eta)=\begin{cases} w_{12} & \text{if } 0<\eta\leqq\dfrac{w_{21}}{w_{21}+w_{12}} \\ -w_{21} & \text{if } \dfrac{w_{21}}{w_{21}+w_{12}}<\eta\leqq 1 \end{cases} \tag{4.39}$$

$$D^+\varphi(0)=\lim_{\eta\to+0}D^-\varphi(\eta) \quad\text{and}\quad D^-\rho(1)=\lim_{\eta\to1-0}D^+\varphi(\eta)\ ,$$

we have, by (4.27) and (4.28),

$$\begin{aligned} D^+\Phi(\eta)=&\int_{B^{(1)}(\eta)} w_{12}\left[\frac{p_{\theta_1}(x)p_{\theta_2}(x)}{\eta p_{\theta_1}(x)+(1-\eta)p_{\theta_2}(x)}+\boldsymbol{\eta}(\eta;x)(p_{\theta_1}(x)-p_{\theta_2}(x))\right]d\mu(x)\\ &+\int_{B^{(2)}(\eta)\cup B^{(3)}(\eta)} w_{21}\left[\frac{-p_{\theta_1}(x)p_{\theta_2}(x)}{\eta p_{\theta_1}(x)+(1-\eta)p_{\theta_2}(x)}\right.\\ &\left.+(1-\boldsymbol{\eta}(\eta;x))\,(p_{\theta_1}(x)-p_{\theta_2}(x))\right]d\mu(x) \end{aligned} \tag{4.40}$$

and

$$\begin{aligned} D^-\Phi(\eta)=&\int_{B^{(1)}(\eta)\cup B^{(2)}(\eta)} w_{12}\left[\frac{p_{\theta_1}(x)p_{\theta_2}(x)}{\eta p_{\theta_1}(x)+(1-\eta)p_{\theta_2}(x)}\right.\\ &\left.+\boldsymbol{\eta}(\eta;x)(p_{\theta_1}(x)-p_{\theta_2}(x))\right]d\mu(x)\\ &+\int_{B^{(3)}(\eta)} w_{21}\left[\frac{-p_{\theta_1}(x)p_{\theta_2}(x)}{\eta p_{\theta_1}(x)+(1-\eta)p_{\theta_2}(x)}\right.\\ &\left.+(1-\boldsymbol{\eta}(\eta;x))\,(p_{\theta_1}(x)-p_{\theta_2}(x))\right]d\mu(x)\ , \end{aligned} \tag{4.41}$$

where

$$\begin{aligned} B^{(1)}(\eta)&=\left\{x \mid \boldsymbol{\eta}(\eta;x)<\frac{w_{21}}{w_{21}+w_{12}}\right\}\ ,\\ B^{(2)}(\eta)&=\left\{x \mid \boldsymbol{\eta}(\eta;x)=\frac{w_{21}}{w_{21}+w_{12}}\right\}\ ,\\ B^{(3)}(\eta)&=\left\{x \mid \boldsymbol{\eta}(\eta;x)>\frac{w_{21}}{w_{21}+w_{12}}\right\}\ . \end{aligned} \tag{4.42}$$

Replacing $\boldsymbol{\eta}(\eta;x)$ with its defined form, we obtain the following simple forms:

$$D^+\Phi(\eta)=w_{12}\int_{B^{(1)}(\eta)}p_{\theta_1}(x)\,d\mu(x)-w_{21}\int_{B^{(2)}(\eta)\cup B^{(3)}(\eta)}p_{\theta_2}(x)\,d\mu(x) \tag{4.43}$$

and

$$D^-\Phi(\eta)=w_{12}\int_{B^{(1)}(\eta)\cup B^{(2)}(\eta)}p_{\theta_1}(x)\,d\mu(x)-w_{21}\int_{B^{(3)}(\eta)}p_{\theta_2}(x)\,d\mu(x)\ . \tag{4.44}$$

Thus

$$D^-\Phi(\eta)-D^+\Phi(\eta)=\int_{B^{(2)}(\eta)}(w_{12}\,p_{\theta_1}(x)+w_{21}\,p_{\theta_2}(x))\,d\mu(x)\ . \tag{4.45}$$

When $x\in B^{(2)}(\eta)$, that is, $\boldsymbol{\eta}(\eta;x)=\dfrac{w_{21}}{w_{21}+w_{12}}$, we have $p_{\theta_2}(x)=\dfrac{w_{12}}{w_{21}(1-\eta)}p_{\theta_1}(x)$ and therefore for $\eta\in(0,1)$

$$D^-\Phi(\eta)-D^+\Phi(\eta)=\frac{w_{12}}{1-\eta}\int_{B^{(2)}(\eta)}p_{\theta_1}(x)\,d\mu(x) \tag{4.46}$$

or

$$=\frac{w_{21}}{\eta}\int_{B^{(2)}(\eta)}p_{\theta_2}(x)\,d\mu(x)\ .$$

Thus we get

LEMMA 4.4. *Suppose* $w_{12}$ *and* $w_{21}$ *are positive. Then, a necessary and sufficient condition that* $D^-\Phi(\eta)-D^+\Phi(\eta)>0$ is $\int_{B^{(2)}(\eta)}p_{\theta_1}(x)\,d\mu(x)\neq 0$ *or equivalently* $\int_{B^{(2)}(\eta)}p_{\theta_2}(x)\,d\mu(x)\neq 0.$

*Remark 4.5.* The above discussions will be used in the section 5.

The following discussions are concerned with the proof of the postulate III.

LEMMA 4.5. *If there exist positive numbers* $e_0$ *and* $e_1$ *such that*

$$e_0=\operatorname{Inf}\frac{p_{\theta_1}(x)}{p_{\theta_2}(x)}\qquad \operatorname{Sup}\frac{p_{\theta_1}(x)}{p_{\theta_2}(x)}=e_1\ , \tag{4.47}$$

*where the infimum and supremum are taken for* $x$ *such that* $p_{\theta_2}(x)>0$, *then there exist positive numbers* $b_0$ *and* $b_1$ *such that* $0<b_0<b_1<1$ *and*

$$E\varphi[\boldsymbol{\eta}(\eta;X_1,\cdots,X_m)]=\begin{cases} w_{12}\eta\ , & \text{for}\quad \eta\in[0,b_0]\\ w_{21}(1-\eta)\ , & \text{for}\quad \eta\in[b_1,1]\end{cases} \tag{4.48}$$

*where* $\varphi(\eta)=\min\{w_{12}\eta,\,w_{21}(1-\eta)\}$ *for* $0\leqq\eta\leqq 1.$

PROOF. Define $b_0$ and $b_1$ by

$$\begin{aligned} b_0&=\max\left\{\eta\,\middle|\,\operatorname{Sup}\boldsymbol{\eta}(\eta;x_1,\cdots,x_m)\leqq\frac{w_{21}}{w_{12}+w_{21}}\right\}\\ b_1&=\min\left\{\eta\,\middle|\,\operatorname{Inf}\boldsymbol{\eta}(\eta;x_1,\cdots,x_m)\geqq\frac{w_{21}}{w_{12}+w_{21}}\right\}\ ,\end{aligned} \tag{4.49}$$

the supremum and infimum being taken for $(x_1, \cdots, x_m)$ such that $\prod_{j=1}^{m} p_{\theta_i}(x_j)>0$, $i=1, 2$. Then, since by (4.47)

$$\frac{\eta e_0^m}{\eta e_1^m+(1-\eta)}<\boldsymbol{\eta}(\eta; x_1, \cdots, x_m)<\frac{\eta e_1^m}{\eta e_0^m+(1-\eta)}$$

for any $\eta \in [0, 1]$, we can easily obtain

$$0<\frac{w_{21}}{e_1^m(w_{12}+w_{21})+w_{21}(1-e_0^m)}<b_0<b_1<\frac{w_{21}}{e_0^m(w_{12}+w_{21})+w_{21}(1-e_1^m)}<1 .$$

From (4.49), we can see that, as long as $\eta \in [0, b_0]$ or $\eta \in [b_1, 1]$, the range of the random variable $\boldsymbol{\eta}(\eta; X_1, \cdots, X_m)$ is contained within the interval $\left[0, \frac{w_{21}}{w_{12}+w_{21}}\right]$ or $\left[\frac{w_{21}}{w_{12}+w_{21}}, 1\right]$, respectively. Now, since $\varphi(\eta) = w_{12}\eta$ for $\eta \in \left[0, \frac{w_{21}}{w_{12}+w_{21}}\right]$ and $\varphi(\eta)=w_{21}(1-\eta)$ for $\eta \in \left[\frac{w_{21}}{w_{12}+w_{21}}, 1\right]$, we have the results. Using lemma 4.5 and corollary 4.2, we can get the following theorem to establish the postulate III of the previous section.

THEOREM 4.3. *Suppose that the condition (4.47) is satisfied. Then, there exist $c^*$ and $n^*$ such that for any $c \in (0, c^*]$ the Bayes sequential stopping rule for the sequential dichotomous decision problem with delayed observations, constant cost $c$ and independent observations takes a simple form: that is, in terms of the notations defined in the section 3,*

(4.50) $$\underline{H}=[0, \underline{\eta}_n] \quad \text{and} \quad \overline{H}_n=[\overline{\eta}_n, 1] \quad \text{for} \quad n \geqq n^* .$$

*Further, there exist $\underline{\eta}$ and $\overline{\eta}$ such that $0<\underline{\eta}<\overline{\eta}<1$ and*

(4.51) $$\underline{H}=\bigcap_{n=0}^{\infty} H_n=[0, \underline{\eta}] , \quad \overline{H}=\bigcap_{n=0}^{\infty} \overline{H}_n=[\overline{\eta}, 1] .$$

*Thus, the non-truncated Bayes stopping rule is completely characterized by a sequential probability ratio test.*

PROOF. To show explicitly the dependence of the functions $\varphi_i(\eta)$ $i=1, 2, \cdots$ on the cost of observation $c$, let us write $\varphi_i(\eta, c)$ instead of $\varphi_i(\eta)$ which was defined by (3.6) and (3.15). As we have already seen (corollary 4.2), for any $c \in (0, \infty)$

(4.52) $$\varphi_i(\eta, c)\geqq\varphi_{i+1}(\eta, c)\geqq c \quad \text{for} \quad \eta \in [0, 1],\ i=1, 2, \cdots .$$

Thus there exists a limiting function, say $\varphi_\infty(\eta, c)$, such as

(4.53) $$\lim_{n\to\infty} \varphi_n(\eta, c)=\varphi_\infty(\eta, c) .$$

Then, clearly it is a concave function of $\eta$ for any fixed $c$. We can also see from the definition of $\varphi_i$'s that

$$\varphi_i(\eta, c) \geqq \varphi_i(\eta, c') \quad \text{for} \quad c > c' . \tag{4.54}$$

Therefore, we can conclude that $\varphi_\infty(\eta, c)$ is monotone increasing in $c$ and further we can show that

$$\lim_{c \to +0} \varphi_\infty(\eta, c) = 0 . \tag{4.55}$$

Using the $b_0$ and $b_1$ of lemma 4.5, let us define $c^*$ as follows:

$$c^* = \max \left\{ c \,\middle|\, \begin{array}{l} \varphi_\infty(b_0, c) \leqq w_{12} b_0 \\ \varphi_\infty(\eta, c) < \varphi_0(\eta) \qquad \text{for} \quad \eta \in (b_0, b_1) \\ \text{and} \quad \varphi_\infty(b_1, c) \leqq w_{21}(1 - b_1) \end{array} \right\} . \tag{4.56}$$

The existence of positive $c^*$ is assured by (4.54) and lemma 4.5. For any $c$ such that $0 < c \leqq c^*$, $n^*$ is defined by

$$n^*(c) = \min \left\{ i \,\middle|\, \begin{array}{l} \varphi_i(b_0, c) \leqq w_{12} b_0 \\ \varphi_i(\eta, c) < \varphi_0(\eta) \qquad \text{for} \quad \eta \in (b_0, b_1) \\ \text{and} \quad \varphi_i(b_1, c) \leqq w_{21}(1 - b_1) \end{array} \right\} . \tag{4.57}$$

Hence, $n^*$ is defined as a function of $c$ whose existence is verified from (4.53) and $0 < c \leqq c^*$. It is also seen that $n^*(c)$ is monotone increasing. Thus, $n \geqq n^*(c^*)$ always implies $n \geqq n^*(c)$ for $c \in [0, c^*]$. For $c^*$ and $n^*(c)$ thus defined, we can show that, for any $c$ $(0 < c \leqq c^*)$ and $n$ $(n \geqq n^*(c))$, there exist exactly two roots, say $\underline{\eta}_n$ and $\bar{\eta}_n$, of the equation $\varphi_0(\eta) = \varphi_n(\eta, c)$ such that $\underline{\eta}_n < \bar{\eta}_n$ and

$$\begin{array}{lll} \varphi_0(\eta) < \varphi_n(\eta, c) & \text{for} & \eta < \underline{\eta}_n \quad \text{or} \quad \eta > \bar{\eta}_n \\ \varphi_0(\eta) > \varphi_n(\eta, c) & \text{for} & \underline{\eta}_n < \eta < \bar{\eta}_n . \end{array} \tag{4.58}$$

This follows from the partial linearity of $\varphi_0(\eta)$ given by (4.48) and the concavity of $\varphi_0$ and $\varphi_n$ as well as the definitions of $c^*$ and $n^*$. Thus, if we regard $c$ as the constant cost of observation, then, by (4.58) and the definition of $\underline{H}_n$ and $\bar{H}_n$ (postulate II), we have (4.50) for $n \geqq n^*(c)$. Furthermore, since $\{\underline{\eta}_n\}$ and $\{\bar{\eta}_n\}$ are monotone decreasing and increasing, respectively, there exist $\lim_{n\to\infty} \underline{\eta}_n = \underline{\eta}$ and $\lim_{n\to\infty} \bar{\eta}_n = \bar{\eta}$; consequently it follows that $\underline{H} = [0, \underline{\eta}]$ and $\bar{H} = [\bar{\eta}, 1]$. The last part of the theorem was already stated in the previous section. Thus we have proved the theorem.

## 5. Related problems

First we will give a brief review of the paper written by Anderson [2]. He raised a problem concerning choice of observations. It is stated as follows. In establishing statistical means to decide between two hypotheses $H_0$ and $H_1$, the experimenter may have the choice of observing a variable $X$ alone or of observing two variables $X$ and $Y$. While observation of two variables is more informative than observation of the one variable, it is also more expensive. The question is whether it is worthwhile for the experimenter to pay the greater cost necessary for the two variables. He must make a decision whether to observe $Y$ before the observation on $X$ is made, where $X$ and $Y$ are assumed to be independent. We will call the above problem Problem I. In particular T. W. Anderson was interested in the problem whether the Bayes procedures of the above problem take a simple form or not. We call this problem Problem II. He ensured that Bayes procedures take a simple form if $X$ and $Y$ are Gaussian with known variances. However, he also raised an example for which Bayes procedures do not take a simple form. His conclusion is, therefore, that the nature of the solution to Problem I depends on the distributions of $X$ and $Y$. Let us first formulate Problem I.

Problem: A statistical hypothesis $H_1$ is tested against another statistical hypothesis $H_2$.

Notations

- $f_i(x)$ : probability density function of $X$ when $H_i$ is true.
- $g_i(y)$ : probability density function of $Y$ when $H_i$ is true.
- $W_i$ : loss for rejecting $H_i$ when $H_i$ is true.
- $C$ : cost of observing $Y$ (or the difference between the cost of observing $X$ and $Y$ and the cost of observing $X$ alone).
- $\eta$ : a prior probability of $H_1$ ($1-\eta$ is the corresponding prior probability of $H_2$).

Defining $\varphi(\eta)=\min\{(1-\eta)W_2, \eta W_1\}$, we obtain the Bayes risk, when $X$ alone is observed, in the form:

$$\Phi(\eta)=E\varphi[\boldsymbol{\eta}(\eta; X)] .$$

On the other hand, the Bayes risk, when $X$ and $Y$ are observed, is $\Psi(\eta)+C$, where $\Psi(\eta)=E\varphi[\boldsymbol{\eta}(\eta; X, Y)]$. Since $X$ and $Y$ are independent,

$$\begin{aligned} \underset{Y}{E}\,\underset{X}{E}\varphi[\boldsymbol{\eta}(\eta; X, Y)] &= \underset{X}{E}\,\underset{Y}{E}\varphi[\boldsymbol{\eta}(\eta; X, Y)] \\ &= E\Phi[\boldsymbol{\eta}(\eta; X)] . \end{aligned}$$

By lemma 4.1 and theorem 4.2, we have that $\Phi(\eta)$ and $\Psi(\eta)$ are concave and $\Phi(\eta)\geqq\Psi(\eta)$ $0\leqq\eta\leqq 1$; further we have $\Phi(0)=\Phi(1)=\Psi(0)=\Psi(1)=0$, $D^+\Phi(0)=D^+\Psi(0)=W_1$ and $D^-\Phi(1)=D^-\Psi(1)=-W_2$. Problem II can be extended to the following problem—Problem III: Under what conditions $\Phi(\eta)-\Psi(\eta)+C$ changes its sign at most three times for any non-negative $C$.

Noting that, for any positive $C$, $\Phi(0)<\Psi(0)+C$ and $\Phi(1)<\Psi(1)+C$, we can see that, if $\Phi(\eta)-\Psi(\eta)-C$ changes its sign exactly three times, the set $\{\eta;\Phi(\eta)\geqq\Psi(\eta)+C\}$ is an interval, which makes the Bayes decision procedure for Problem I simple.

Since $\Phi(\eta)-\Psi(\eta)$ is non-negative and $\Phi(0)-\Psi(0)=\Phi(1)-\Psi(1)=0$, the unimodality of $\Phi(\eta)-\Psi(\eta)$ is equivalent to the statement that $\Phi(\eta)-\Psi(\eta)-C$ changes the sign at most three times for any positive $C$. Thus, the Problem III reduces to the problem what conditions on the distributions of $X$ and $Y$ ensure the unimodality of $\Phi(\eta)-\Psi(\eta)$. Let us define several notations:

$G(z\mid H_i)$: distribution function of the random variable $\dfrac{f_1(X)}{f_2(X)}$ when $H_i$ is true $(i=1,2)$.

$H(z\mid H_i)$: distribution function of the random variable $\dfrac{f_1(X)\,g_1(Y)}{f_2(X)\,g_2(Y)}$ when $H_i$ is true $(i=1,2)$.

Define

$$\frac{W_1}{W_1+W_2}G(z\mid H_1)+\frac{W_2}{W_1+W_2}G(z\mid H_2)=G(z\mid W_1,W_2)$$

$$\frac{W_1}{W_1+W_2}H(z\mid H_1)+\frac{W_2}{W_1+W_2}H(z\mid H_2)=H(z\mid W_1,W_2)\,.$$

THEOREM 5.1. *A necessary and sufficient condition that $\Phi(\eta)-\Psi(\eta)$ is unimodal is that $G(z\mid W_1,W_2)-H(z\mid W_1,W_2)$ changes its sign exactly once.*

PROOF. By theorem 4.2, we have

$$D^+\Phi(\eta)=\int\Big\{D^+\varphi[\boldsymbol{\eta}(\eta;x)]\frac{f_1(x)f_2(x)}{\eta f_1(x)+(1-\eta)f_2(x)}$$
$$+\varphi[\boldsymbol{\eta}(\eta;x)](f_1(x)-f_2(x))\Big\}\,d\mu(x)\,.$$

Since $\varphi(\eta)=\min\{(1-\eta)W_2,\eta W_1\}$, we have

$$D^+\varphi(\eta)=\begin{cases} W_1 & \text{for} \quad \eta<\dfrac{W_1}{W_1+W_2} \\ -W_2 & \text{for} \quad \eta\geqq\dfrac{W_1}{W_1+W_2}\,. \end{cases}$$

Thus, as is given by (4.40),

$$\begin{aligned} D^+\Phi(\eta)= & W_1\int_{B^{(1)}}\frac{f_1(x)f_2(x)}{\eta f_1(x)+(1-\eta)f_2(x)}\,d\mu(x) \\ & -W_2\int_{B^{(2)}\cup B^{(3)}}\frac{f_1(x)f_2(x)}{\eta f_1(x)+(1-\eta)f_2(x)}\,d\mu(x) \\ & +W_1\int_{B^{(1)}}\boldsymbol{\eta}(\eta\,;x)(f_1(x)-f_2(x))\,d\mu(x) \\ & +W_2\int_{B^{(2)}\cup B^{(3)}}[1-\boldsymbol{\eta}(\eta\,;x)](f_1(x)-f_2(x))\,d\mu(x) \end{aligned}$$

where

$$B^{(1)}=\left\{x\mid\boldsymbol{\eta}(\eta;x)<\frac{W_1}{W_1+W_2}\right\}, \qquad B^{(2)}=\left\{x\mid\boldsymbol{\eta}(\eta;x)=\frac{W_1}{W_1+W_2}\right\}$$

$$B^{(3)}=\left\{x\mid\boldsymbol{\eta}(\eta;x)>\frac{W_1}{W_1+W_2}\right\}.$$

We can easily check the following:

$$\frac{f_1(x)f_2(x)}{\eta f_1(x)+(1-\eta)f_2(x)}+\boldsymbol{\eta}(\eta\,;x)(f_1(x)-f_2(x))=f_1(x)$$

$$\frac{f_1(x)f_2(x)}{\eta f_1(x)+(1-\eta)f_2(x)}-(1-\boldsymbol{\eta}(\eta\,;x))(f_1(x)-f_2(x))=f_2(x)\,.$$

Thus we have, as is given by (4.43),

$$\begin{aligned} D^+\Phi(\eta) &= W_1\int_{B^{(1)}}f_1(x)\,d\mu(x)-W_2\int_{B^{(2)}\cup B^{(3)}}f_2(x)\,d\mu(x) \\ &= W_1G\left(\frac{W_2(1-\eta)}{W_1\eta}\,\middle|\,H_1\right)-W_2\left[1-G\left(\frac{W_2(1-\eta)}{W_1\eta}\,\middle|\,H_2\right)\right]. \end{aligned}$$

In a similar way, we have

$$D^+\Psi(\eta)=W_1H\left(\frac{W_2(1-\eta)}{W_1\eta}\,\middle|\,H_1\right)-W_2\left[1-H\left(\frac{W_2(1-\eta)}{W_1\eta}\,\middle|\,H_2\right)\right].$$

Now

$$\begin{aligned} D^+\{\Phi(\eta)-\Psi(\eta)\} &= D^+\Phi(\eta)-D^+\Psi(\eta) \\ &= W_1G\left(\frac{W_2(1-\eta)}{W_1\eta}\,\Big|\,H_1\right)+W_2G\left(\frac{W_2(1-\eta)}{W_1\eta}\,\Big|\,H_2\right) \\ &\quad -\left[W_1H\left(\frac{W_2(1-\eta)}{W_1\eta}\,\Big|\,H_1\right)+W_2H\left(\frac{W_2(1-\eta)}{W_1\eta}\,\Big|\,H_2\right)\right]. \end{aligned}$$

As is easily seen, the unimodality of $\Phi(\eta)-\Psi(\eta)$ is equivalent to the requirement that $D^+\{\Phi(\eta)-\Psi(\eta)\}$ changes its sign exactly once from positive to negative when $\eta$ runs from 0 to 1, which implies our theorem.

*Remark 5.1.* In the above proof we utilized the fact that $\Phi(\eta)$ and $\Psi(\eta)$ are concave on [0, 1] and that $\Phi(\eta)-\Psi(\eta)$ is non-negative for $\eta \in [0, 1]$.

We can easily check that the necessary and sufficient condition is satisfied for the first example by Anderson in which $X$ and $Y$ have normal distributions with known variances. Also we can easily assure that this condition is not satisfied for the example treated in section 3 of Anderson's paper.

## Acknowledgements

The author is heartily thankful to Professor Z. Suetuna, Director of the Institute of Statistical Mathematics, for his warm encouragement. Also, he wishes to thank Drs. K. Matusita and H. Akaike for invaluable encouragements and suggestions. His sincere thanks are extended to Mr. H. Hudimoto for valuable discussions and to Mrs. K. Hudimoto, Misses Y. Nogami and T. Naito for their assistance in preparing the manuscript.

THE INSTITUTE OF STATISTICAL MATHEMATICS

## REFERENCES

[1] T. W. Anderson, "Sequential analysis with delayed observations," *Jour. Amer. Statist. Assoc.*, 59 (1964), 1006-1015.
[2] T. W. Anderson, "On Bayes procedures for a problem with choice of observations," *Ann. Math. Statist.*, 35 (1964), 1128-1135.
[3] D. Blackwell and M. A. Girshick, *Theory of Games and Statistical Decisions*, John Wiley & Sons, New York, 1954.
[4] K. Matusita, "On the theory of statistical decision functions; Correction," *Ann. Inst. Statist. Math.*, 3 (1951) 17-35; 4 (1952), 51-53.
[5] A. Wald, *Statistical Decision Functions*, John Wiley & Sons, New York, 1950.

# CORRECTIONS TO

# "ON SEQUENTIAL DECISION PROBLEMS WITH DELAYED OBSERVATIONS"

YUKIO SUZUKI

The author is indebted to Professor K. Miyasawa of Tokyo University for drawing his attention to the incorrect statements in theorem 2.3 and its proof of the above titled paper (Ann. Inst. Statist. Math., 18 (1966), 229–267). In the theorem "but not necessarily convex" should be deleted and in its proof "The statement that $\Xi_j^*$ $j=0, \cdots, N$ is not necessarily convex will be clarified in the sections 3 and 4. This is a typical character due to the existence of delayed observations." should be deleted. Then the theorem and its proof are correct.

The author now considers that it is suitable to add the following remark after the rectified theorem and its proof:

*Remark 2.2.* A typical character of $\Xi^*$ $j=0, \cdots, N$ due to the existence of delayed observations will be clarified by Proposition II and its justification in the sections 3 and 4.

Other corrections:

(1) $\tau_{j+m,\xi}^*(x)$ in expressions (1.16), (1.17) and (1.18) should read $\tau_{j,\xi}^*(x)$.

(2) $\underline{H}$ in expression (4.50) should read $\underline{H}_n$.

(3) $H_n$ in expression (4.51) should read $\underline{H}_n$.