

ASYMPTOTICALLY MOST INFORMATIVE PROCEDURE IN THE CASE OF EXPONENTIAL FAMILIES

BY KAZUTOMO KAWAMURA

§ 1. Introduction.

Recently we showed the following fact in our paper [2]. We considered in [2] two binomial trials E_1, E_2 having unknown means p_1, p_2 respectively. And we have introduced the *notion of costs* such that we must pay costs c_1, c_2 to the observation of a result given by the trials E_1, E_2 respectively. In each step we are admitted to select one of the two trials E_1, E_2 . Be continued the selections by some way we denoted the sequence of trials till n -th step as $E^{(1)}, \dots, E^{(n)}$ and the sequence of costs till n -th step as $C^{(1)}, \dots, C^{(n)}$. Of course we may select at i -th step $E^{(i)}$ from the two trials E_1, E_2 depending previous $i-1$ data X_1, \dots, X_{i-1} given by $E^{(1)}, \dots, E^{(i-1)}$. A procedure \mathfrak{Q} was given in [2] such that the sum of information given by two dimensional likelihood ratio relative to the sum of costs till n -th step to discriminate $p_1 > p_2$ or $p_1 < p_2$ is asymptotically maximized. In [2] we assumed the unknown true two dimensional parameter (p_1, p_2) did not exist on the boundary $p_1 = p_2$. In our another paper [3] we considered analogous model having two kinds of trials E_1, E_2 which are obeyed normal distributions with unknown means m_1, m_2 and known same variance σ^2 and costs c_1, c_2 respectively. Then analogous procedure \mathfrak{Q} is asymptotically optimal in the same sense described above. In [3] we noted that our procedure \mathfrak{Q} reduced to a policy which does not depending on previous n data X_1, \dots, X_n but only on sample sizes n_1 of E_1, n_2 of E_2 till n -th step. We have omitted the proof of the problem in [3] because we can easily get analogous proof.

In this paper we generalize these problems to k trials E_1, \dots, E_k having exponential distributions with one dimensional unknown parameter $\theta_1, \dots, \theta_k$ respectively. That is, an observation X of E_j has a probability density function of exponential type in Kullback's sense [4] with one dimensional unknown parameter $\theta_j (j=1, \dots, k)$ respectively. And we introduced the boundary $\pi: \mu \cdot \theta = p (\theta = (\theta_1, \dots, \theta_k))$ as a hyperplane in k dimensional euclidean space where $\mu = (\mu_1, \dots, \mu_k)$ is any fixed k dimensional unit vector having all non-zero components and p is any fixed non-negative number and $\mu \cdot \theta$ is the inner product of two vectors μ and θ . Moreover we use the *notion of costs* introduced by Kunisawa [6], as we used the notion in [2], [3], then we can get some information of θ_j by paying of cost $c_j (j=1, \dots, k)$ respectively. Then we shall show analogously that under the generalized procedure \mathfrak{Q}^* given in the following Section 3 the sum of information relative to the sum of costs payed till n -th step to discriminate $\mu \cdot \theta$ larger than p or not is asymptotically

Received June 30, 1966.

maximized. Additionally we show in this paper that under the original procedure \mathfrak{P} given in [2], [3] the ratio is also asymptotically maximized in the sense of the generalized procedure \mathfrak{P}^* . Moreover the problem given in [3] will be shown in special example of case $k=2$ in Section 5. Finally note that in this paper we need not to assume that our unknown true parameter θ is not an element of our hyper-plane π .

§ 2. Notations, definitions and some lemmas.

Definition of the exponential family introduced by S. Kullback. Suppose that $f(x, \theta_0)$ and $f(x, \theta_1)$ are generalized densities of a dominated set of probability measures on the measurable space (\mathfrak{X}, Φ) so that

$$(2.1) \quad \nu_i(E) = \int_E f(x, \theta_i) dx \quad E \in \Phi, (i=0, 1).$$

For a given $f(x, \theta_0)$ we seek the member of the dominated set of probability measures that is "nearest" to or most closely resembles the probability measure ν_0 in the sense of smallest directed divergence

$$(2.2) \quad I(\theta_1, \theta_0) = \int_R f(x, \theta_1) \log \frac{f(x, \theta_1)}{f(x, \theta_0)} dx.$$

as a restriction of $f(x, \theta_1)$ we shall require $f(x, \theta_1)$ minimizing $I(\theta_1, \theta_0)$ subject to

$$\int_R T(x) f(x, \theta_1) dx = \theta_1$$

where θ_1 is any fixed constant and $Y=T(x)$ a measurable statistic. Then the minimum value of $I(\theta_1, \theta_0)$ is given if and only if

$$(2.3) \quad f(x, \theta_1) = e^{(\theta_1)T(x)} f(x, \theta_0) / M(\tau(\theta_1))$$

where

$$M(\tau(\theta_1)) = \int_R f(x, \theta_0) e^{(\theta_1)T(x)} dx.$$

We remark that

$$f(x, \theta) = f(x, \theta_0) e^{(\theta)T(x)} / M(\tau(\theta))$$

is said to generate an exponential family of distributions, the family of exponential type determined by $f(x, \theta_0)$, as θ ranges over its values satisfying $M(\tau(\theta)) < \infty$.

$$\theta = \frac{d}{d\tau} \log M(\tau(\theta)) \quad \text{and} \quad \nu_0(x: T(x)=\theta) \neq 1,$$

then $\tau(\theta)$ is a strictly increasing function of θ . Using this fact if $f(x, \theta_1)$ and $f(x, \theta_2)$ are the members of common exponential family generated by $f(x, \theta_0)$ then for any fixed θ in the interval $[\theta_1, \theta_2]$ ($\theta_1 < \theta_2$)

$$\begin{aligned}
 M(\tau(\theta)) &= \int_R f(x, \theta_0) e^{\tau(\theta)T(x)} dx \\
 &\leq \int_{T(x) \geq 0} f(x, \theta_0) e^{\tau(\theta_2)T(x)} dx \\
 &\quad + \int_{T(x) < 0} f(x, \theta_0) e^{\tau(\theta_1)T(x)} dx < \infty
 \end{aligned}$$

so that the exponential family is defined on connected interval or full line $R=(-\infty, \infty)$. Next we consider n independent observations x_1, \dots, x_n from true density function $f(x, \theta)$ in the exponential family generated by $f(x, \theta_0)$. Then logarithm of the likelihood function $\prod_{i=1}^n f(x_i, \theta)$ is expressed as follows.

$$\log \prod_{i=1}^n f(x_i, \theta) = \log \prod_{i=1}^n f(x_i, \theta_0) + \tau(\theta) \sum_{i=1}^n T(x_i) - n \log M(\tau(\theta))$$

The maximum value of $\prod_{i=1}^n f(x_i, \theta)$ is given if and only if $\theta = \hat{\theta}_n$ as follows

$$\begin{aligned}
 \frac{d}{d\theta} \log \prod_{i=1}^n f(x_i, \theta) &= 0, \\
 \frac{d\tau}{d\theta} \cdot \left\{ \sum_{i=1}^n T(x_i) - n \frac{d}{d\tau} \log M(\tau(\theta)) \right\} &= 0
 \end{aligned}$$

where $d\tau/d\theta > 0$ is satisfied.

$$(2.4) \quad \hat{\theta}_n = \frac{d}{d\tau} \log M(\tau(\theta)) = \frac{\sum_{i=1}^n T(x_i)}{n}.$$

In the following line we suppose that the true parameter θ is finite, then by the strong law of large numbers $\hat{\theta}_n \rightarrow \theta$ as $n \rightarrow \infty$ is satisfied with probability 1.

Definition of parameter space Θ . Our k exponential families with one dimensional parameter $\theta_j (j=1, \dots, k)$ are defined on one dimensional open intervals Θ_j of θ_j respectively. Then our unknown parameter $\theta = (\theta_1, \dots, \theta_k)$ is an element of k dimensional parameter space

$$\Theta = \Theta_1 \otimes \dots \otimes \Theta_k$$

which is a product space of k open intervals $\Theta_1, \dots, \Theta_k$. Next we divide the space Θ by our hyperplane $\mu \cdot \theta = p$ as follows. Of course we assume the hyperplane acrosses our parameter space Θ .

$$\begin{aligned}
 H_1 &= \{\theta : \mu \cdot \theta > p, \theta \in \Theta\}, & H_2 &= \{\theta : \mu \cdot \theta < p, \theta \in \Theta\}, \\
 \pi &= \{\theta : \mu \cdot \theta = p, \theta \in \Theta\}.
 \end{aligned}$$

Then $\Theta = H_1 + H_2 + \pi$ is satisfied. Next let $E^{(i)}$ be i -th trial which is one of k elements E_1, \dots, E_k and define X_i to be i -th random variable which is given by trial $E^{(i)}$ randomizedly.

Calculation of likelihood function of θ . Given the first n trials in some way, we define the number of E_j in $E^{(1)}, \dots, E^{(n)}$ as n_j . In the following line we

suppose $n_j \geq 1$ ($j=1, \dots, k$) then all n_j observations X_i from the trial E_j in the first n trials $E^{(1)}, \dots, E^{(n)}$ is considered as n_j independent observations from the trial E_j . Therefore the likelihood function of the parameter θ , is expressed by

$$\prod_{\{i: E^{(i)}=E_j\}} f(x_i, \theta, E_j)$$

where $\{i: E^{(i)}=E_j\}$ is a set of n_j elements of i in $1, \dots, n$ satisfying $E^{(i)}=E_j$ for any fixed $j(j=1, \dots, k)$. Then for given $E^{(1)}, \dots, E^{(n)}$ the likelihood function of θ is expressed by the product as following

$$(2.5) \quad \prod_{i=1}^n f(x_i, \theta, E^{(i)}) = \prod_{\{i: E^{(i)}=E_1\}} f(x_i, \theta, E_1) \cdots \prod_{\{i: E^{(i)}=E_k\}} f(x_i, \theta, E_k)$$

Of course, for any fixed E_j , the probability density function $f(x, \theta, E_j)$ of the trial E_j for every $\theta=(\theta_1, \dots, \theta_k)$ in Θ is considered as a function of θ , only and independent of θ_i ($i \neq j$).

Definition of $\hat{\theta}_n$ and unique existence of the value. We denote as $\hat{\theta}_n$ the maximum likelihood estimate of θ . The j -th component of $\hat{\theta}_n$ will maximize the j -th likelihood function

$$\prod_{\{i: E^{(i)}=E_j\}} f(x_i, \theta, E_j)$$

with respect to θ_j . We denote the value of θ_j which maximizes the j -th likelihood function as $\hat{\theta}_{n_j}$. Then $\hat{\theta}_{n_j}$ given by the n_j trials of E_j is uniquely expressed as followings from the discussion of exponential family

$$(2.6) \quad \hat{\theta}_{n_j} = \frac{\sum_{\{i: E^{(i)}=E_j\}} T_j(x_i)}{n_j}$$

where $T_j(x)$ is a statistic of trial E_j satisfying

$$\int_{\mathcal{R}} T_j(x) f(x, \theta, E_j) dx = \theta_j \quad (j=1, \dots, k)$$

as in (2.4). Hence our $\hat{\theta}_n$ is expressed as following, if $n_j \geq 1$ ($j=1, \dots, k$) is satisfied,

$$(2.7) \quad \hat{\theta}_n = \{\hat{\theta}_{n_1}, \dots, \hat{\theta}_{n_k}\}.$$

And the uniqueness of $\hat{\theta}_n$ is reduced to that of $\hat{\theta}_{n_j}$. So that we get the following lemma as to be proved.

LEMMA 1. *Under any sequence $E^{(1)}, \dots, E^{(n)}$ if $n_j \geq 1$ ($j=1, \dots, k$) is satisfied then there exists $\hat{\theta}_n$ uniquely on our parameter space Θ .*

Definition of $\tilde{\theta}_n$. Next we shall denote by $\tilde{\theta}_n$ the maximum likelihood estimate of θ on the subspace $a(\hat{\theta}_n)$ over the first n trials where $a(\theta)$ is defined as follows:

$$(2.8) \quad \begin{aligned} a(\theta) &= \Theta - H_i & \text{if } \theta \in H_i & \quad (i=1, 2), \\ &= \Theta & \text{if } \theta \in \pi. \end{aligned}$$

Definition of sum of information. Now we define the sum of information to discriminate $\mu \cdot \theta > p$ or not using our $\hat{\theta}_n$ and $\check{\theta}_n$ as follows

$$(2.9) \quad \begin{aligned} S_n(\hat{\theta}_n, \check{\theta}_n) &= \sum_{i=1}^n \log \frac{f(x_i, \hat{\theta}_n, E^{(i)})}{f(x_i, \check{\theta}_n, E^{(i)})} \\ &= \log \frac{\max_{\theta \in \Theta} \prod_{i=1}^n f(x_i, \theta, E^{(i)})}{\max_{\varphi \in \alpha(\hat{\theta}_n)} \prod_{i=1}^n f(x_i, \varphi, E^{(i)})}. \end{aligned}$$

Definition of mean discrimination. As a measure of discrimination between two probability density functions $f(x, \theta, E_j)$, $f(x, \varphi, E_j)$ we can use by Kullback [4]

$$(2.10) \quad I(\theta, \varphi, E_j) = \int_{\mathcal{R}} \left[\log \frac{f(x, \theta, E_j)}{f(x, \varphi, E_j)} \right] f(x, \theta, E_j) dx$$

where $\theta = (\theta_1, \dots, \theta_k)$, $\varphi = (\varphi_1, \dots, \varphi_k)$ and $j = 1, \dots, k$.¹⁾

Existence and uniqueness of $\check{\theta}_n$. To find $\check{\theta}_n$, we must minimize $S_n(\hat{\theta}_n, \varphi)$ with respect to φ in $\alpha(\hat{\theta}_n)$ from the definition of $\check{\theta}_n$. Since $f(x, \theta, E_j)$ belongs to the exponential family defined in E_j we have

$$\log \frac{\prod_{\{i: E^{(i)}=E_j\}} f(x_i, \hat{\theta}_n, E_j)}{\prod_{\{i: E^{(i)}=E_j\}} f(x_i, \varphi, E_j)} = n_j I(\hat{\theta}_n, \varphi, E_j)$$

so that

$$S_n(\hat{\theta}_n, \varphi) = \sum_{j=1}^k n_j I(\hat{\theta}_n, \varphi, E_j).$$

Then we must find $\check{\theta}_n$ on $\alpha(\hat{\theta}_n)$ which minimizes $\sum_{j=1}^k n_j I(\hat{\theta}_n, \varphi, E_j)$ with respect to φ . First we shall show the fact that $\check{\theta}_n \in \pi$ for all n . But this fact will be given evidently from the property of $\check{\theta}_n$ minimizing

$$\sum_{j=1}^k n_j I(\hat{\theta}_n, \varphi, E_j)$$

on $\alpha(\hat{\theta}_n)$ with respect to φ , and the convexity of $I(\hat{\theta}_n, \varphi, E_j)$ with respect to φ_j ($j=1, \dots, k$) where $\varphi = (\varphi_1, \dots, \varphi_k)$. Therefore we can search $\check{\theta}_n$ on π as to minimizing $S_n(\hat{\theta}_n, \varphi)$ with respect to φ . Put

$$dS_n(\hat{\theta}_n, \varphi) = 0,$$

then

$$(2.11) \quad \sum_{j=1}^k \frac{\partial S_n(\hat{\theta}_n, \varphi)}{\partial \varphi_j} d\varphi_j = \frac{\partial S_n(\hat{\theta}_n, \varphi)}{\partial \varphi} \cdot d\varphi = 0$$

and we have from $\varphi \in \pi: \mu \cdot \varphi = p$

$$(2.12) \quad \mu \cdot d\varphi = 0.$$

1) In this paper we assumed as a restriction of density function of E_j ($j=1, \dots, k$) that continuity of $d^2 I(\theta, \varphi, E_j) / d\varphi_j^2$ with respect to φ_j for any fixed θ_j in the interval Θ_j ($j=1, \dots, k$) respectively.

2) $\partial S_n(\hat{\theta}_n, \varphi) / \partial \varphi \cdot d\varphi$ is the innerproduct of two vectors $\partial S_n(\hat{\theta}_n, \varphi) / \partial \varphi = (\partial S_n(\hat{\theta}_n, \varphi) / \partial \varphi_1, \dots, \partial S_n(\hat{\theta}_n, \varphi) / \partial \varphi_k)$ and $d\varphi = (d\varphi_1, \dots, d\varphi_k)$.

3) $\mu \cdot d\varphi$ is also the innerproduct of the vectors $\mu = (\mu_1, \dots, \mu_k)$ and $d\varphi = (d\varphi_1, \dots, d\varphi_k)$.

Therefore we get from (2.11), (2.12)

$$(2.13) \quad \frac{\partial S_n(\hat{\theta}_n, \varphi)}{\partial \varphi} = \text{Constant} \cdot \mu.$$

Or equivalently

$$(2.14) \quad \frac{n_1}{\mu_1} \frac{dI(\hat{\theta}_n, \varphi, E_1)}{d\varphi_1} = \dots = \frac{n_k}{\mu_k} \frac{dI(\hat{\theta}_n, \varphi, E_k)}{d\varphi_k}.$$

Hence from the relation (2.14) and the convexity of $I(\hat{\theta}_n, \varphi, E_j)$ with respect to φ_j , that is, $dI(\hat{\theta}_n, \varphi, E_j)/d\varphi_j$ is strictly increasing function of φ_j ($j=1, \dots, k$), we can find $\hat{\theta}_n$ on π uniquely for any fixed n . Therefore if $\hat{\theta}_n$ is uniquely given for any fixed n then from (2.14) and $\mu \cdot \varphi = p$ we can find $\hat{\theta}_n$ uniquely on π as to be proved.

LEMMA 2. *Under any sequence $E^{(1)}, \dots, E^{(n)}$, if $n_j \geq 1$ ($j=1, \dots, k$) is satisfied then there exists $\hat{\theta}_n$ uniquely on our hyperplane π .*

Behavior of $\hat{\theta}_n$ under any sequence $E^{(1)}, E^{(2)}, \dots$. Now we shall show the probability equals to zero that $\hat{\theta}_n$ does not converge for any sequence $E^{(1)}, E^{(2)}, \dots$. The event that $\hat{\theta}_n$ does not converge is included in the event $\hat{\theta}_{n_j}$ does not converge for some j . But this event would not occur from the strong law of large numbers. Hence $\hat{\theta}_n$ converges with probability 1 for any sequence $E^{(1)}, E^{(2)}, \dots$.

LEMMA 3. *Under any sequence $E^{(1)}, E^{(2)}, \dots$, our $\hat{\theta}_n$ converges with probability 1.*

In the following line we put the limit point as θ_0 tentatively. Next under any sequence $E^{(1)}, E^{(2)}, \dots$, we shall prove the probability equals to zero that there exists some integer N such that $\hat{\theta}_n \in \pi$ is satisfied for all $n \geq N$. For any sequence $E^{(1)}, E^{(2)}, \dots$, there exists some integer j in $1, \dots, k$ satisfying $n_j \rightarrow \infty$ as $n \rightarrow \infty$. And evidently the event that $\hat{\theta}_n \in \pi$ for all $n \geq N$ reduced to the event $\hat{\theta}_N = \hat{\theta}_{N+1} = \dots$, that is, the event that $\hat{\theta}_{N_j} = \hat{\theta}_{N_{j+1}} = \dots$ for all $j=1, \dots, k$. But for some j satisfying $n_j \rightarrow \infty$ as $n \rightarrow \infty$ the event that the maximum likelihood estimate $\hat{\theta}_{n_j}$ gives an identical value for sufficiently large n occurs with probability zero by the zero one law. Hence the event that there exists some integer N such that $\hat{\theta}_n \in \pi$ is satisfied for all $n \geq N$ would not occur at all. So that $\hat{\theta}_n$ does not exist on π frequently n for any sequence $E^{(1)}, E^{(2)}, \dots$ with probability 1 as to be proved. This property of $\hat{\theta}_n$ will play an important part in following section.

§ 3. The optimal procedure \mathfrak{F}^* .

Definition of sequence $\{\lambda_n\}$. In this section, at first we shall define k dimensional ratio vector λ_n in each step n . For fixed $\hat{\theta}_n$, we define θ_n^* on π in subspace of θ

$$R_1^{\text{sign } \mu_1(\hat{\theta}_n)} \otimes \dots \otimes R_k^{\text{sign } \mu_k(\hat{\theta}_n)}$$

where $R_j^-(\hat{\theta}_n) = (-\infty, \hat{\theta}_{n_j}]$ and $R_j^+(\hat{\theta}_n) = [\hat{\theta}_{n_j}, +\infty)$ such that following equality is satisfied

$$(3.1) \quad \frac{I(\hat{\theta}_n, \theta_n^*, E_1)}{c_1} = \dots = \frac{I(\hat{\theta}_n, \theta_n^*, E_k)}{c_k}.$$

Unique existence of θ_n^* for fixed $\hat{\theta}_n$ is given by the convexity of $I(\hat{\theta}_n, \varphi, E_j)$ with respect to φ_j for all j . If $\hat{\theta}_{n-1} \notin \pi$ then $\hat{\theta}_{n-1} \neq \theta_{n-1}^*$. Using this $\hat{\theta}_{n-1}$ and θ_{n-1}^* we shall define k dimensional vector $\lambda_n = (\lambda_{n1}, \dots, \lambda_{nk})$ having k positive components uniquely by the two conditions $\lambda_{n1} + \dots + \lambda_{nk} = 1$ and

$$(3.2) \quad \frac{\lambda_{n1}}{\mu_1} \left[\frac{dI(\hat{\theta}_{n-1}, \varphi, E_1)}{d\varphi_1} \right]_{\varphi=\theta_{n-1}^*} = \dots = \frac{\lambda_{nk}}{\mu_k} \left[\frac{dI(\hat{\theta}_{n-1}, \varphi, E_k)}{d\varphi_k} \right]_{\varphi=\theta_{n-1}^*} \quad (4)$$

And if $\hat{\theta}_{n-1} \in \pi$ then $\hat{\theta}_{n-1} = \theta_{n-1}^*$ so we can not find λ_n uniquely. In this case, we put $\lambda_n = \lambda_{n-1}$, moreover we put $\lambda_1 = (1, 0, \dots, 0), \dots, \lambda_k = (0, \dots, 0, 1)$. In this way we can get λ_n uniquely for all n . Hence $\lambda_1, \lambda_2, \dots$ is uniquely defined for any sequence $E^{(1)}, E^{(2)}, \dots$. In the following line we shall call a vector having all positive components *positive vector* and denote a positive vector V as $V > 0$.

Behavior of λ_n under any sequence $E^{(1)}, E^{(2)}, \dots$. Next we shall investigate the behavior of the sequence λ_n . If the limit point θ_0 of $\hat{\theta}_n$ did not exist on our plane π then λ_n converges to a positive vector $\lambda_0 = (\lambda_{01}, \dots, \lambda_{0k})$ ⁵⁾ satisfying $\sum_{j=1}^k \lambda_{0j} = 1$ and the next equality analogously as we defined λ_n uniquely by $\hat{\theta}_n$

$$(3.3) \quad \frac{\lambda_{01}}{\mu_1} \left[\frac{dI(\theta_0, \varphi, E_1)}{d\varphi_1} \right]_{\varphi=\theta_0^*} = \dots = \frac{\lambda_{0k}}{\mu_k} \left[\frac{dI(\theta_0, \varphi, E_k)}{d\varphi_k} \right]_{\varphi=\theta_0^*}$$

where θ_0^* is uniquely defined by θ_0 on π in subspace

$$R_1^{\text{sign } \mu_1(\theta_0)} \otimes \dots \otimes R_k^{\text{sign } \mu_k(\theta_0)}$$

satisfying that

$$(3.4) \quad \frac{I(\theta_0, \theta_0^*, E_1)}{c_1} = \dots = \frac{I(\theta_0, \theta_0^*, E_k)}{c_k}.$$

So then, under any sequence $E^{(1)}, E^{(2)}, \dots$, if $\theta_0 = \lim_{n \rightarrow \infty} \hat{\theta}_n$ does not exist on π , then $\lim_{n \rightarrow \infty} \lambda_n = \lambda_0 > 0$ with probability 1.

Otherwise, if $\theta_0 = \lim_{n \rightarrow \infty} \hat{\theta}_n$ exists on π , we shall show λ_n has a positive limit vector with probability 1, as follows. From the assumption $\lim_{n \rightarrow \infty} \hat{\theta}_n = \theta_0 \in \pi$ we get $\lim_{n \rightarrow \infty} \theta_n^* = \theta_0$ with probability 1 so we have $\lim_{n \rightarrow \infty} \hat{\theta}_n = \lim_{n \rightarrow \infty} \theta_n^*$ and $\lim_{n \rightarrow \infty} \hat{\theta}_n = \theta_0$ with probability 1. By Taylor's expansion

$$(3.5) \quad I(\hat{\theta}_n, \theta_n^*, E_j) = \frac{p_j(\hat{\theta}_n - \theta_n^*)^2}{2!} \left[\frac{d^2 I(\hat{\theta}_n, \varphi, E_j)}{d\varphi_j^2} \right]_{\varphi=\bar{\theta}_n},$$

$$(3.6) \quad \left[\frac{dI(\hat{\theta}_n, \varphi, E_j)}{d\varphi_j} \right]_{\varphi=\theta_n^*} = p_j(\hat{\theta}_n - \theta_n^*) \left[\frac{d^2 I(\hat{\theta}_n, \varphi, E_j)}{d\varphi_j^2} \right]_{\varphi=\bar{\theta}_n}$$

4) If $\hat{\theta}_{n-1} \notin \pi$ then we can define λ_n uniquely by (3.2), where θ_{n-1}^* in (3.2) is defined uniquely by $\hat{\theta}_{n-1}$ from (3.1). Hence we can consider λ_n as a function of $\hat{\theta}_{n-1}$: $\lambda_n = \lambda_n(\hat{\theta}_{n-1})$.

5) Where λ_0 is uniquely given as a function of the limit point θ_0 : $\lambda_0 = \lambda_0(\theta_0)$.

where $p_j(\cdot)$ is a projection of the vector in (\cdot) to j -th coordinate, and $\bar{\theta}_n, \bar{\bar{\theta}}_n$ exists on the intervals

$$0 < p_j(\hat{\theta}_n - \bar{\theta}_n) < p_j(\hat{\theta}_n - \theta_n^*), \quad 0 < p_j(\hat{\theta}_n - \bar{\bar{\theta}}_n) < p_j(\hat{\theta}_n - \theta_n^*),$$

respectively, for any fixed j . Then $\bar{\theta}_n$ and $\bar{\bar{\theta}}_n$ converge to θ_0 with probability 1. Therefore

$$(3.7) \quad \lim_{n \rightarrow \infty} \left[\frac{d^2 I(\hat{\theta}_n, \varphi, E_j)}{d\varphi_j^2} \right]_{\varphi = \bar{\theta}_n} = \lim_{n \rightarrow \infty} \left[\frac{d^2 I(\hat{\theta}_n, \varphi, E_j)}{d\varphi_j^2} \right]_{\varphi = \bar{\bar{\theta}}_n} > 0.$$

So by the definition of λ_{n+1} (3.2)

$$(3.8) \quad \begin{aligned} \left[\frac{\lambda_{n+1i}}{\lambda_{n+1j}} \right]^2 &= \left[\frac{\mu_i}{\mu_j} \right]^2 \left[\frac{dI(\hat{\theta}_n, \varphi, E_j)}{d\varphi_j} \right]_{\varphi = \theta_n^*}^2 \bigg/ \left[\frac{dI(\hat{\theta}_n, \varphi, E_i)}{d\varphi_i} \right]_{\varphi = \theta_n^*}^2 \\ &= \left[\frac{\mu_i}{\mu_j} \right]^2 \frac{p_j(\hat{\theta}_n - \theta_n^*)^2}{p_i(\hat{\theta}_n - \theta_n^*)^2} \left[\frac{d^2 I(\hat{\theta}_n, \varphi, E_j)}{d\varphi_j^2} \right]_{\varphi = \bar{\theta}_n}^2 \bigg/ \left[\frac{d^2 I(\hat{\theta}_n, \varphi, E_i)}{d\varphi_i^2} \right]_{\varphi = \bar{\bar{\theta}}_n}^2. \end{aligned}$$

And by the definition of θ_n^* (3.1)

$$\frac{I(\hat{\theta}_n, \theta_n^*, E_i)}{c_i} = \frac{I(\hat{\theta}_n, \theta_n^*, E_j)}{c_j}.$$

Then we have

$$(3.9) \quad \frac{p_j(\hat{\theta}_n - \theta_n^*)^2}{p_i(\hat{\theta}_n - \theta_n^*)^2} = \frac{c_j}{c_i} \left[\frac{d^2 I(\hat{\theta}_n, \varphi, E_i)}{d\varphi_i^2} \right]_{\varphi = \bar{\theta}_n} \bigg/ \left[\frac{d^2 I(\hat{\theta}_n, \varphi, E_j)}{d\varphi_j^2} \right]_{\varphi = \bar{\bar{\theta}}_n}.$$

Therefore from (3.8)

$$(3.10) \quad \begin{aligned} \left[\frac{\lambda_{n+1i}}{\lambda_{n+1j}} \right]^2 &= \left[\frac{\mu_i}{\mu_j} \right]^2 \frac{c_j}{c_i} \left[\frac{d^2 I(\hat{\theta}_n, \varphi, E_i)}{d\varphi_i^2} \right]_{\varphi = \bar{\theta}_n} \bigg/ \left[\frac{d^2 I(\hat{\theta}_n, \varphi, E_j)}{d\varphi_j^2} \right]_{\varphi = \bar{\bar{\theta}}_n} \\ &\cdot \left[\frac{d^2 I(\hat{\theta}_n, \varphi, E_j)}{d\varphi_j^2} \right]_{\varphi = \bar{\bar{\theta}}_n}^2 \bigg/ \left[\frac{d^2 I(\hat{\theta}_n, \varphi, E_i)}{d\varphi_i^2} \right]_{\varphi = \bar{\bar{\theta}}_n}^2. \end{aligned}$$

So that

$$(3.11) \quad \lim_{n \rightarrow \infty} \left[\frac{\lambda_{ni}}{\lambda_{nj}} \right]^2 = \left[\frac{\mu_i}{\mu_j} \right]^2 \frac{c_j}{c_i} \left[\frac{d^2 I(\theta_0, \varphi, E_j)}{d\varphi_j^2} \right]_{\varphi = \theta_0} \bigg/ \left[\frac{d^2 I(\theta_0, \varphi, E_i)}{d\varphi_i^2} \right]_{\varphi = \theta_0}.$$

Since the right hand limit value is positive, our sequence λ_n has a positive limit vector with probability 1, we put the vector as λ_0 tentatively. As a conclusion we get in any sequence $E^{(1)}, E^{(2)}, \dots$ there exists a positive vector λ_0 such that $\lim_{n \rightarrow \infty} \lambda_n = \lambda_0 > 0$ is satisfied with probability 1.

LEMMA 4. *Under any sequence $E^{(1)}, E^{(2)}, \dots, \lambda_n$ converges to a positive vector as $n \rightarrow \infty$ with probability 1.*

The optimal procedure \mathfrak{G}^* . Using the sequence λ_n ($n \geq 1$) defined above, we

consider $E^{(n)}$ ($n \geq 1$) as a sequence of random variables which take values E_1, \dots, E_k in each step n and have probabilities as follows, for each $n \geq 1$

$$(3.12) \quad \text{Prob} \{E^{(n)} = E_j\} = \lambda_{nj}.$$

In the following line we call this randomized policy *procedure* \mathfrak{P}^* .

Proprety of our procedure \mathfrak{P}^* . By Lemma 4 we observed the fact that in any sequence $E^{(1)}, E^{(2)}, \dots$ there exists a positive vector λ_0 such that λ_n converges to λ_0 . Using this fact we get under our procedure \mathfrak{P}^* n_j/n converges to j -th component $\lambda_{0j} (> 0)$ of the limit vector λ_0 by the strong law of large numbers. Hence under our procedure \mathfrak{P}^* $n_j \rightarrow \infty$ as $n \rightarrow \infty$ for any $j=1, \dots, k$ is satisfied so that $\hat{\theta}_n$ converges to the unknown true parameter θ as to be proved.

THEOREM 1. *Under our procedure* \mathfrak{P}^* , $\hat{\theta}_n$ *converges to the unknown true parameter* θ *as* $n \rightarrow \infty$ *with probability* 1.

Next we suppose the unknown true parameter θ is not an element of π . Then there exists a positive vector λ uniquely given by θ , as we defined λ_n uniquely by $\hat{\theta}_{n-1}$ and λ_0 uniquely by θ_0 , such that $\lim_{n \rightarrow \infty} \lambda_n = \lambda$ is satisfied with probability 1. This fact is shown as discussion of Lemma 4. Where the vector $\lambda = (\lambda_1, \dots, \lambda_k)$ is defined satisfying $\sum_{j=1}^k \lambda_j = 1$ and analogously as (3.3)

$$(3.13) \quad \frac{\lambda_1}{\mu_1} \left[\frac{dI(\theta, \varphi, E_1)}{d\varphi_1} \right]_{\varphi=\theta^*} = \dots = \frac{\lambda_k}{\mu_k} \left[\frac{dI(\theta, \varphi, E_k)}{d\varphi_k} \right]_{\varphi=\theta^*}$$

where θ^* is uniquely defined by θ on π in subspace

$$R_1^{\text{sign } \mu_1(\theta)} \otimes \dots \otimes R_k^{\text{sign } \mu_k(\theta)}$$

satisfying that analogously as (3.4)

$$(3.14) \quad \frac{I(\theta, \theta^*, E_1)}{c_1} = \dots = \frac{I(\theta, \theta^*, E_k)}{c_k}.$$

Otherwise if the unknown true parameter θ is an element of π , then under our procedure \mathfrak{P}^* , from the result of Lemma 4, λ_n converges to a positive vector with probability 1. Hence under our procedure \mathfrak{P}^* if $\theta \notin \pi$ then n_j/n converges to λ_j ($j=1, \dots, k$) with probability 1, and if $\theta \in \pi$ then n_j/n converges to a positive value with probability 1 as to be proved.

COROLLARY 1. *Under our procedure* \mathfrak{P}^* , *if the true unknown parameter* θ *is not an element of* π *then* $(n_1/n, \dots, n_k/n)$ *converges to our vector* $\lambda (> 0)$, *defined in* (3.13), *is satisfied with probability* 1, *and otherwise if the true unknown parameter* θ *is an element of* π *then* $(n_1/n, \dots, n_k/n)$ *converges to a positive vector is satisfied with probability* 1.

Optimal condition and optimal ratio vector. As a conclusion of Corollary 1, we have shown under our procedure \mathfrak{P}^* that if $\theta \notin \pi$, then $n_j/n \rightarrow \lambda_j$ with probability 1,

and otherwise if $\theta \in \pi$, then n_j/n converges to a positive value ($j=1, \dots, k$) with probability 1. In the following line we call this proposition as *optimal condition* and the vector λ defined in (3.13) as *optimal ratio vector*. The meaning of optimality will be given in following main theorems of next section.

§ 4. Main theorems and the proofs.

Under the *optimal condition* given in the preceding section, if the true unknown parameter θ is not an element of π , then we get $\lim_{n \rightarrow \infty} \tilde{\theta}_n = \theta^*$ with probability 1, from the two equalities (2.14), (3.13)

$$(2.14)' \quad \frac{n_1}{\mu_1} \left[\frac{dI(\hat{\theta}_n, \varphi, E_1)}{d\varphi_1} \right]_{\varphi=\tilde{\theta}_n} = \dots = \frac{n_k}{\mu_k} \left[\frac{dI(\hat{\theta}_n, \varphi, E_k)}{d\varphi_k} \right]_{\varphi=\tilde{\theta}_n},$$

$$(3.13)' \quad \frac{\lambda_1}{\mu_1} \left[\frac{dI(\theta, \varphi, E_1)}{d\varphi_1} \right]_{\varphi=\theta^*} = \dots = \frac{\lambda_k}{\mu_k} \left[\frac{dI(\theta, \varphi, E_k)}{d\varphi_k} \right]_{\varphi=\theta^*}.$$

So we get

$$\lim_{n \rightarrow \infty} \frac{I(\hat{\theta}_n, \tilde{\theta}_n, E_j)}{c_j} = I^*(\theta)$$

with probability 1, where $I^*(\theta)$ is the value of (3.14)

$$(3.14)' \quad I^*(\theta) = \frac{I(\theta, \theta^*, E_1)}{c_1} = \dots = \frac{I(\theta, \theta^*, E_k)}{c_k}.$$

And from the definitions of $S_n(\hat{\theta}_n, \tilde{\theta}_n)$ and $\sum_{i=1}^n C^{(i)}$

$$(4.1) \quad \frac{S_n(\hat{\theta}_n, \tilde{\theta}_n)}{\sum_{i=1}^n C^{(i)}} = \frac{n_1 I(\hat{\theta}_n, \tilde{\theta}_n, E_1) + \dots + n_k I(\hat{\theta}_n, \tilde{\theta}_n, E_k)}{n_1 c_1 + \dots + n_k c_k}.$$

Hence the sum of information relative to the sum of costs for the first n trials to discriminate $\mu \cdot \theta > p$ or not

$$\frac{S_n(\hat{\theta}_n, \tilde{\theta}_n)}{\sum_{i=1}^n C^{(i)}}$$

converges to the value $I^*(\theta)$ with probability 1 as to be proved. And otherwise if the true unknown parameter θ is an element of π , then we get $\lim_{n \rightarrow \infty} \tilde{\theta}_n = \theta$ with probability 1. So we get

$$\lim_{n \rightarrow \infty} \frac{I(\hat{\theta}_n, \tilde{\theta}_n, E_j)}{c_j} = 0$$

with probability 1. Hence our

$$\frac{S_n(\hat{\theta}_n, \tilde{\theta}_n)}{\sum_{i=1}^n C^{(i)}}$$

converges to zero with probability 1 as to be proved.

THEOREM 2. *Under our optimal condition if the true unknown parameter θ is not an element of π then we get*

$$(4.2) \quad \lim_{n \rightarrow \infty} \frac{S_n(\hat{\theta}_n, \check{\theta}_n)}{\sum_{i=1}^n C^{(i)}} = I^*(\theta)$$

with probability 1, and otherwise if the true unknown parameter θ is an element of π then we get

$$\frac{S_n(\hat{\theta}_n, \check{\theta}_n)}{\sum_{i=1}^n C^{(i)}}$$

converges to zero with probability 1.

By Corollary 1 we have shown that our procedure \mathfrak{F}^* has a property of *optimal condition*. Therefore we have next corollary.

COROLLARY 2. *Under our procedure \mathfrak{F}^* we get the same result as given in Theorem 2.*

Meaning of optimality. In the following line we consider a class of procedure satisfying $\min_j(n_j) \rightarrow \infty$ as $n \rightarrow \infty$, and $(n_1/n, \dots, n_k/n)$ converges to vector $\lambda' = (\lambda_1', \dots, \lambda_k')$ where $\sum_{j=1}^k \lambda_j' = 1$. In this class of procedure we shall show that our procedure \mathfrak{F}^* is asymptotically most informative one relative to the sum of costs to discriminate $\mu \cdot \theta < p$ or not. That is, under another procedure \mathfrak{F}^{**} having the limit ratio λ' different from our ratio λ we can get asymptotically less information relative to the sum of costs to discriminate $\mu \cdot \theta > p$ or not than we get using our procedure \mathfrak{F}^* . Under the procedure \mathfrak{F}^{**} , fixed by the limit ratio λ' different from our ratio λ , how much information to discriminate $\mu \cdot \theta > p$ or not we can get asymptotically relative to the sum of costs? From the first condition that $\min_j(n_j) \rightarrow \infty$ as $n \rightarrow \infty$ we get $\lim_{n \rightarrow \infty} \hat{\theta}_n = \theta$ with probability 1. If we assume the true parameter θ does not exist on π then for a given ratio in the second condition: $(n_1/n, \dots, n_k/n)$ converges to ratio $\lambda' = (\lambda_1', \dots, \lambda_k')$ as $n \rightarrow \infty$ ($\lambda' \neq \lambda$), we can define θ^{**} uniquely as an element of π and satisfying

$$(4.3) \quad \frac{\lambda_1'}{\mu_1} \left[\frac{dI(\theta, \varphi, E_1)}{d\varphi_1} \right]_{\varphi=\theta^{**}} = \dots = \frac{\lambda_k'}{\mu_k} \left[\frac{dI(\theta, \varphi, E_k)}{d\varphi_k} \right]_{\varphi=\theta^{**}}$$

From the equality (2.14)' defining $\check{\theta}_n$ uniquely on our hyperplane π we can verify $\lim_{n \rightarrow \infty} \check{\theta}_n = \theta^{**}$ with probability 1. Therefore

$$(4.4) \quad \lim_{n \rightarrow \infty} \frac{S_n(\hat{\theta}_n, \check{\theta}_n)}{\sum_{i=1}^n C^{(i)}} = \frac{\lambda_1' I(\theta, \theta^{**}, E_1) + \dots + \lambda_k' I(\theta, \theta^{**}, E_k)}{\lambda_1' c_1 + \dots + \lambda_k' c_k}$$

is satisfied with probability 1. So we denote the limit value under the procedure

\mathfrak{Q}^{**} as $I^{**}(\theta)$ in the following line. From the inequality $\lambda' \neq \lambda$ in the second condition we have $\theta^{**} \neq \theta^*$ so we get the next inequality

$$(4.5) \quad \lambda_1'(\theta, \theta^{**}, E_1) + \cdots + \lambda_k'(\theta, \theta^{**}, E_k) < \lambda_1'(\theta, \theta^*, E_1) + \cdots + \lambda_k'(\theta, \theta^*, E_k).$$

Therefore we have

$$(4.6) \quad I^{**}(\theta) < I^*(\theta)$$

Otherwise if the true parameter θ is an element of π then we have $\lim_{n \rightarrow \infty} \tilde{\theta}_n = \theta$ with probability 1. So that

$$\frac{S_n(\hat{\theta}_n, \tilde{\theta}_n)}{\sum_{i=1}^n C^{(i)}}$$

converges to zero with probability 1 as to be proved.

THEOREM 3. *Under any procedure \mathfrak{Q}^{**} , satisfying $\min_j(n_j) \rightarrow \infty$ as $n \rightarrow \infty$ and $(n_1/n, \dots, n_k/n)$ converges to $\lambda' = (\lambda_1', \dots, \lambda_k')$ as $n \rightarrow \infty$ ($\lambda' \neq \lambda$), if the true unknown parameter θ is not an element of π then we get*

$$(4.7) \quad \lim_{n \rightarrow \infty} \frac{S_n(\hat{\theta}_n, \tilde{\theta}_n)}{\sum_{i=1}^n C^{(i)}} = I^{**}(\theta) < I^*(\theta)$$

is satisfied with probability 1. And otherwise if the true unknown parameter θ is an element of π then

$$\frac{S_n(\hat{\theta}_n, \tilde{\theta}_n)}{\sum_{i=1}^n C^{(i)}}$$

converges to zero with probability 1.

Having expected the meaning of Theorem 3 we have called the property of our procedure \mathfrak{Q}^* as *optimal condition* or having *optimal ratio vector*.

§ 5. Original procedure \mathfrak{Q} in the case $k=2$.

Original procedure \mathfrak{Q} . We consider two exponential trials E_1, E_2 and use the procedure \mathfrak{Q} given in [2], [3] that is $E^{(1)} = E_1, E^{(2)} = E_2$ and for $n \geq 2$ we define successively

$$(5.1) \quad E^{(n+1)} \in \left\{ E: \text{which maximizes } \frac{I(\hat{\theta}_n, \tilde{\theta}_n, E_i)}{c_i} \quad (i=1, 2) \right\}.$$

Under this procedure \mathfrak{Q} we shall show the *optimal condition*. Following to Lemma 3, under any sequence we have $\hat{\theta}_n$ converges with probability 1. By Lemma 4 λ_n converges to a *positive ratio vector* with probability 1. And also the procedure \mathfrak{Q} has a property $n_j/n - \lambda_{n_j}$ converges to zero as $n \rightarrow \infty$ with probability 1 as the proof

given in our paper [2]. Therefore n_j/n converges to a positive value with probability 1. So that under the procedure \mathfrak{P} we have $\min(n_1, n_2) \rightarrow \infty$ as $n \rightarrow \infty$ with probability 1. Hence under the procedure \mathfrak{P} we have $\hat{\theta}_n \rightarrow \theta$ as $n \rightarrow \infty$ with probability 1. If the true parameter θ is not an element of π then λ_n converges to λ with probability 1. Then n_j/n converges to λ_j with probability 1. And otherwise n_j/n converges to a positive value with probability 1 ($j=1, 2$) respectively. Therefore we can verify the procedure \mathfrak{P} also have the *optimal condition* as to be proved. Then by Theorem 2 we get the limit equality (4.2) with probability 1. Hence this procedure \mathfrak{P} also has the optimal property, that is, most informative relative to the sum of costs than any other procedure having the two conditions $n_j \rightarrow \infty$ as $n \rightarrow \infty$, $(n_1/n, n_2/n)$ converges to $\lambda'=(\lambda_1', \lambda_2')$ as $n \rightarrow \infty$ ($\lambda' \neq \lambda$).

Example of the case $k=2$. We consider two trials E_1, E_2 depending normal distributions with unknown means and known variances respectively. Then the density functions of E_1, E_2 are expressed as follows

$$(5.2) \quad f(x, \theta, E_j) = \frac{1}{\sqrt{2\pi} \sigma_j} \exp \left\{ -\frac{(x-m_j)^2}{2\sigma_j^2} \right\} \quad (j=1, 2)$$

where $\theta=(m_1, m_2)$ is the pair of unknown means, then we get

$$(5.3) \quad I(\theta, \varphi, E_j) = \frac{(m_j - m_j^*)^2}{2\sigma_j^2} \quad (j=1, 2)$$

where $\theta=(m_1, m_2)$, $\varphi=(m_1^*, m_2^*)$. And the subspaces H_1, H_2 and π are given by $\mu=(\mu_1, \mu_2)$ and p , that is

$$(5.4) \quad \begin{aligned} H_1 &= \{\theta: \mu_1 m_1 + \mu_2 m_2 > p\}, & H_2 &= \{\theta: \mu_1 m_1 + \mu_2 m_2 < p\}, \\ \pi &= \{\theta: \mu_1 m_1 + \mu_2 m_2 = p\}. \end{aligned}$$

And $\lim_{n \rightarrow \infty} \tilde{\theta}_n = \theta^*$ is given uniquely by the equality (3.14) and $\theta^* \in \pi$.

$$(5.5) \quad \frac{(m_1 - m_1^*)^2}{2\sigma_1^2} = \frac{(m_2 - m_2^*)^2}{2\sigma_2^2}$$

where $\theta=(m_1, m_2)$ is the true parameter and $\theta^*=(m_1^*, m_2^*)$ is the limit point of $\tilde{\theta}_n$. And the limit value of (4.2) $I^*(\theta)$ is given as the value of (5.5). Moreover we put $\mu_1=1/\sqrt{2}$, $\mu_2=-1/\sqrt{2}$ and $p=0$ then our subspaces become as following

$$(5.6) \quad \begin{aligned} H_1 &= \{(m_1, m_2): m_1 > m_2\}, & H_2 &= \{(m_1, m_2): m_1 < m_2\}, \\ \pi &= \{(m_1, m_2): m_1 = m_2\}. \end{aligned}$$

In this case we can get the *optimal ratio* λ as following by (3.13)

$$(5.7) \quad \frac{\lambda_1(m_1 - m_1^*)}{\sigma_1^2} = -\frac{\lambda_2(m_2 - m_2^*)}{\sigma_2^2}$$

Therefore from the two relations (5.5), (5.7) we get the *optimal ratio* λ as next

value uniformly on Θ .⁶⁾

$$(5.8) \quad \lambda = \left(\frac{\sqrt{c_2} \sigma_1}{\sqrt{c_1} \sigma_2 + \sqrt{c_2} \sigma_1}, \frac{\sqrt{c_1} \sigma_2}{\sqrt{c_1} \sigma_2 + \sqrt{c_2} \sigma_1} \right)$$

Here we additionally suppose $\sigma_1^2 = \sigma_2^2$, then the *optimal ratio* λ becomes

$$(5.9) \quad \lambda = \left(\frac{\sqrt{c_2}}{\sqrt{c_1} + \sqrt{c_2}}, \frac{\sqrt{c_1}}{\sqrt{c_1} + \sqrt{c_2}} \right)$$

We can see this result in our paper [3], that is, this result of Section 5 is a generalization of papers [2], [3] as to be proved.

ACKNOWLEDGEMENT. The author expresses hearty thanks to Professor K. Kunisawa for many useful suggestions.

REFERENCES

- [1] CHERNOFF, H., Sequential design of experiments. Ann. Math. Stat. **30** (1959), 755-770.
- [2] KAWAMURA, K., Asymptotic behavior of sequential design with costs of experiments. Kōdai Math. Sem. Rep. **16** (1964), 169-182.
- [3] KAWAMURA, K., Asymptotic behavior of sequential design with costs of experiments (The case of normal distribution). Kōdai Math. Sem. Rep. **17** (1965), 48-52.
- [4] KULLBACK, S., Information theory and statistics. (1959), Wiley.
- [5] KUNISAWA, K., Modern probability theory. 12th Ed. (1963), Iwanami Co. (In Japanese)
- [6] KUNISAWA, K., Introduction to information theory for operations research. 4th Ed. (1963), J. U. S. E. (In Japanese)

6) Generally λ is a function of θ on Θ ($\theta \notin \pi$) but, in this special example of trials E_1, E_2 with normal distributions of the case $k=2$, λ does not depending on θ , that is, λ equals to a constant value uniformly on Θ ($\theta \notin \pi$).

This is the reason why, in this case of special example, the optimal procedure \mathcal{P} does not depending on previous observations till now but only on sample sizes till now as we showed in the note in [3].