KŌDAI MATH. SEM. RFP.
19 (1967), 61—74

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 $E^{(1)}, \cdots, E^{(n)}$ and the sequence of costs till n-th step as $C^{(1)}, \cdots, C^{(n)}$. Of course we may select at i-th step $E^{(i)}$ from the two trials E_1, E_2 depending previous $i-1$ data $X_1, \cdots, X_{i-1}$ given by $E^{(1)}, \cdots, E^{(i-1)}$. A procedure $\mathfrak{L}$ 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 $\mathfrak{L}$ is asymptotically optimal in the same sense described above. In [3] we noted that our procedure $\mathfrak{L}$ reduced to a policy which does not depending on previous n data $X_1, \cdots, X_n$ 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 $E_1, \cdots, E_k$ having exponential distributions with one dimensional unknown parameter $\theta_1, \cdots, \theta_k$ 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 $\theta_j (j=1, \cdots, k)$ respectively. And we introduced the boundary π: $\mu\cdot\theta=p\,(\theta=(\theta_1, \cdots, \theta_k))$ as a hyperplane in k dimensional euclidean space where $\mu=(\mu_1, \cdots, \mu_k)$ 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 $c_j (j=1, \cdots, k)$ respectively. Then we shall show analogously that under the generalized procedure $\mathfrak{L}^*$ 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 $\mathfrak{L}$ given in [2], [3] the ratio is also asymptotically maximized in the sense of the generalized procedure $\mathfrak{L}^*$. 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 hyperplane 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 $(\mathfrak{X}, \Phi)$ so that

$$\nu_i(E)=\int_E f(x, \theta_i)dx \qquad E\in\Phi,\ (i=0, 1). \tag{2. 1}$$

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

$$I(\theta_1, \theta_0)=\int_R f(x, \theta_1) \log \frac{f(x, \theta_1)}{f(x, \theta_0)} dx. \tag{2. 2}$$

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

$$f(x, \theta_1)=e^{\tau(\theta_1)T(x)}f(x, \theta_0)/M(\tau(\theta_1)) \tag{2. 3}$$

where

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

We remark that

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

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

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

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 $[\theta_1, \theta_2]$ $(\theta_1<\theta_2)$

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

$$\leqq\int_{T(x)\geqq 0} f(x,\theta_0)\, e^{\tau(\theta_2)T(x)}dx$$

$$+\int_{T(x)<0} f(x,\theta_0)\, e^{\tau(\theta_1)T(x)}dx<\infty$$

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

$$\frac{d}{d\theta}\log\prod_{i=1}^n f(x_i,\theta)=0,$$

$$\frac{d\tau}{d\theta}\cdot\left\{\sum_{i=1}^n T(x_i)-n\frac{d}{d\tau}\log M(\tau(\theta))\right\}=0$$

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

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

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

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

$$\Theta=\Theta_1\otimes\cdots\otimes\Theta_k$$

which is a product space of k open intervals $\Theta_1,\cdots,\Theta_k$. 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.

$$H_1=\{\theta:\ \mu\cdot\theta>p,\ \theta\in\Theta\},\qquad H_2=\{\theta:\ \mu\cdot\theta<p,\ \theta\in\Theta\},$$

$$\pi=\{\theta:\ \mu\cdot\theta=p,\ \theta\in\Theta\}.$$

Then $\Theta=H_1+H_2+\pi$ is satisfied. Next let $E^{(i)}$ be i-th trial which is one of k elements $E_1,\cdots,E_k$ 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 $E^{(1)},\cdots,E^{(n)}$ as n_j. In the following line we suppose $n_j \geqq 1$ $(j=1, \cdots, k)$ then all n_j observations X_i from the trial E_j in the first n trials $E^{(1)}, \cdots, E^{(n)}$ is considered as n_j independent observations from the trial E_j. Therefore the likelihood function of the parameter θ_j 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 $1, \cdots, n$ satisfying $E^{(i)}=E_j$ for any fixed $j(j=1, \cdots, k)$. Then for given $E^{(1)}, \cdots, E^{(n)}$ the likelihood function of θ is expressed by the product as following

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

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

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 $\hat{\theta}_{nj}$. Then $\hat{\theta}_{nj}$ given by the n_j trials of E_j is uniquely expressed as followings from the discussion of exponential family

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

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

$$\int_R T_j(x) f(x, \theta, E_j) dx = \theta_j \qquad (j=1, \cdots, k)$$

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

$$\hat{\theta}_n = \{\hat{\theta}_{n1}, \cdots, \hat{\theta}_{nk}\}. \tag{2.7}$$

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

LEMMA 1. *Under any sequence $E^{(1)}, \cdots, E^{(n)}$ if $n_j \geqq 1$ $(j=1, \cdots, k)$ 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:

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

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 $\tilde{\theta}_n$ as follows

$$S_n(\hat{\theta}_n, \tilde{\theta}_n)=\sum_{i=1}^{n}\log\frac{f(x_i, \hat{\theta}_n, E^{(i)})}{f(x_i, \tilde{\theta}_n, E^{(i)})}$$

(2.9)

$$=\log\frac{\max_{\theta\in\Theta}\prod_{i=1}^{n}f(x_i, \theta, E^{(i)})}{\max_{\varphi\in a(\hat{\theta}_n)}\prod_{i=1}^{n}f(x_i, \varphi, E^{(i)})}.$$

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]

$$I(\theta, \varphi, E_j)=\int_R\left[\log\frac{f(x, \theta, E_j)}{f(x, \varphi, E_j)}\right]f(x, \theta, E_j)dx \tag{2.10}$$

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

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

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

so that

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

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

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

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

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

then

$$\sum_{j=1}^{k}\frac{\partial S_n(\hat{\theta}_n, \varphi)}{\partial\varphi_j}d\varphi_j=\frac{\partial S_n(\hat{\theta}_n, \varphi)}{\partial\varphi}\cdot d\varphi=0 \tag{2.11}$$ [2]

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

$$\mu\cdot d\varphi=0. \tag{2.12}$$ [3]

---

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

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

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

Therefore we get from (2. 11), (2. 12)

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

Or equivalently

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

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 $(j=1, \cdots, k)$, we can find $\tilde{\theta}_n$ 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 $\tilde{\theta}_n$ uniquely on π as to be proved.

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

Behavior of $\hat{\theta}_n$ *under any sequence* $E^{(1)}, E^{(2)}, \cdots$. Now we shall show the probability equals to zero that $\hat{\theta}_n$ does not converge for any sequence $E^{(1)}, E^{(2)}, \cdots$. The event that $\hat{\theta}_n$ does not converge is included in the event $\hat{\theta}_{nj}$ 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 $E^{(1)}, E^{(2)}, \cdots$.

LEMMA 3. *Under any sequence* $E^{(1)}, E^{(2)}, \cdots$, *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 $E^{(1)}, E^{(2)}, \cdots$, 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 $n\geqq N$. For any sequence $E^{(1)}, E^{(2)}, \cdots$, there exists some integer j in $1, \cdots, k$ satisfying $n_j\to\infty$ as $n\to\infty$. And evidently the event that $\hat{\theta}_n\in\pi$ for all $n\geqq N$ reduced to the event $\hat{\theta}_N=\hat{\theta}_{N+1}=\cdots$, that is, the event that $\hat{\theta}_{Nj}=\hat{\theta}_{N+1j}=\cdots$ for all $j=1, \cdots, k$. But for some j satisfying $n_j\to\infty$ as $n\to\infty$ the event that the maximum likelihood estimate $\hat{\theta}_{nj}$ 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 $n\geqq N$ would not occur at all. So that $\hat{\theta}_n$ does not exist on π frequentrly n for any sequence $E^{(1)}, E^{(2)}, \cdots$ 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 $\mathfrak{L}^*$.

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

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

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

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

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 $\boldsymbol{\lambda}_n = (\lambda_{n1}, \cdots, \lambda_{nk})$ having k positive components uniquely by the two conditions $\lambda_{n1} + \cdots + \lambda_{nk} = 1$ and

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

And if $\hat{\theta}_{n-1} \in \pi$ then $\hat{\theta}_{n-1} = \theta_{n-1}^*$ so we can not find $\boldsymbol{\lambda}_n$ uniquely. In this case, we put $\boldsymbol{\lambda}_n = \boldsymbol{\lambda}_{n-1}$, moreover we put $\boldsymbol{\lambda}_1 = (1, 0, \cdots, 0), \cdots, \boldsymbol{\lambda}_k = (0, \cdots, 0, 1)$. In this way we can get $\boldsymbol{\lambda}_n$ uniquely for all n. Hence $\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \cdots$ is uniquely defined for any sequence $E^{(1)}, E^{(2)}, \cdots$. 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 $\boldsymbol{\lambda}_n$ under any sequence $E^{(1)}, E^{(2)}, \cdots$. Next we shall investigate the behavior of the sequence $\boldsymbol{\lambda}_n$. If the limit point $\boldsymbol{\theta}_0$ of $\hat{\theta}_n$ did not exist on our plane π then $\boldsymbol{\lambda}_n$ converges to a positive vector $\boldsymbol{\lambda}_0 = (\lambda_{01}, \cdots, \lambda_{0k})$ [5] satisfying $\sum_{j=1}^{k} \lambda_{0j} = 1$ and the next equality analogously as we defined $\boldsymbol{\lambda}_n$ uniquely by $\hat{\theta}_n$

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

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

$$R_1^{\operatorname{sign} \mu_1(\theta_0)} \otimes \cdots \otimes R_k^{\operatorname{sign} \mu_k(\theta_0)}$$

satisfying that

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

So then, under any sequence $E^{(1)}, E^{(2)}, \cdots$, if $\theta_0 = \lim_{n\to\infty} \hat{\theta}_n$ does not exist on π, then $\lim_{n\to\infty} \boldsymbol{\lambda}_n = \boldsymbol{\lambda}_0 > 0$ with probability 1.

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

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

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

4) If $\hat{\theta}_{n-1} \notin \pi$ then we can define $\boldsymbol{\lambda}_n$ 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 $\boldsymbol{\lambda}_n$ as a function of $\hat{\theta}_{n-1}$: $\boldsymbol{\lambda}_n = \boldsymbol{\lambda}_n(\hat{\theta}_{n-1})$.

5) Where $\boldsymbol{\lambda}_0$ is uniquely given as a function of the limit point θ_0: $\boldsymbol{\lambda}_0 = \boldsymbol{\lambda}_0(\theta_0)$.

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

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

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

$$\lim_{n\to\infty}\left[\frac{d^2I(\hat{\theta}_n, \varphi, E_j)}{d\varphi_j^2}\right]_{\varphi=\bar{\theta}_n}=\lim_{n\to\infty}\left[\frac{d^2I(\hat{\theta}_n, \varphi, E_j)}{d\varphi_j^2}\right]_{\varphi=\bar{\bar{\theta}}_n}>0. \tag{3. 7}$$

So by the definition of $\boldsymbol{\lambda}_{n+1}$ (3. 2)

$$\begin{aligned}\left[\frac{\lambda_{n+1i}}{\lambda_{n+1j}}\right]^2&=\left[\frac{\mu_i}{\mu_j}\right]^2\left[\frac{dI(\hat{\theta}_n, \varphi, E_j)}{d\varphi_j}\right]^2_{\varphi=\theta_n^*}\Big/\left[\frac{dI(\hat{\theta}_n, \varphi, E_i)}{d\varphi_i}\right]^2_{\varphi=\theta_n^*}\\&=\left[\frac{\mu_i}{\mu_j}\right]^2\frac{p_j(\hat{\theta}_n-\theta_n^*)^2}{p_i(\hat{\theta}_n-\theta_n^*)^2}\left[\frac{d^2I(\hat{\theta}_n, \varphi, E_j)}{d\varphi_j^2}\right]^2_{\varphi=\bar{\theta}_n}\Big/\left[\frac{d^2I(\hat{\theta}_n, \varphi, E_i)}{d\varphi_i^2}\right]^2_{\varphi=\bar{\bar{\theta}}_n}.\end{aligned} \tag{3. 8}$$

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

$$\frac{p_j(\hat{\theta}_n-\theta_n^*)^2}{p_i(\hat{\theta}_n-\theta_n^*)^2}=\frac{c_j}{c_i}\left[\frac{d^2I(\hat{\theta}_n, \varphi, E_i)}{d\varphi_i^2}\right]_{\varphi=\bar{\theta}_n}\Big/\left[\frac{d^2I(\hat{\theta}_n, \varphi, E_j)}{d\varphi_j^2}\right]_{\varphi=\bar{\theta}_n}. \tag{3. 9}$$

Therefore from (3. 8)

$$\begin{aligned}\left[\frac{\lambda_{n+1i}}{\lambda_{n+1j}}\right]^2&=\left[\frac{\mu_i}{\mu_j}\right]^2\frac{c_j}{c_i}\left[\frac{d^2I(\hat{\theta}_n, \varphi, E_i)}{d\varphi_i^2}\right]_{\varphi=\bar{\theta}_n}\Big/\left[\frac{d^2I(\hat{\theta}_n, \varphi, E_j)}{d\varphi_j^2}\right]_{\varphi=\bar{\theta}_n}\\&\cdot\left[\frac{d^2I(\hat{\theta}_n, \varphi, E_j)}{d\varphi_j^2}\right]^2_{\varphi=\bar{\bar{\theta}}_n}\Big/\left[\frac{d^2I(\hat{\theta}_n, \varphi, E_i)}{d\varphi_i^2}\right]^2_{\varphi=\bar{\bar{\theta}}_n}.\end{aligned} \tag{3. 10}$$

So that

$$\lim_{n\to\infty}\left[\frac{\lambda_{ni}}{\lambda_{nj}}\right]^2=\left[\frac{\mu_i}{\mu_j}\right]^2\frac{c_j}{c_i}\left[\frac{d^2I(\theta_0, \varphi, E_j)}{d\varphi_j^2}\right]_{\varphi=\theta_0}\Big/\left[\frac{d^2I(\theta_0, \varphi, E_i)}{d\varphi_i^2}\right]_{\varphi=\theta_0}. \tag{3. 11}$$

Since the right hand limit value is positive, our sequence $\boldsymbol{\lambda}_n$ has a positive limit vector with probability 1, we put the vector as $\boldsymbol{\lambda}_0$ tentatively. As a conclusion we get in any sequence $E^{(1)}, E^{(2)}, \cdots$ there exists a positive vector $\boldsymbol{\lambda}_0$ such that $\lim_{n\to\infty}\boldsymbol{\lambda}_n=\boldsymbol{\lambda}_0>0$ is satisfied with probability 1.

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

The optimal procedure $\mathfrak{L}^*$. Using the sequence $\boldsymbol{\lambda}_n$ ($n\geqq 1$) defined above, we consider $E^{(n)}$ $(n\geqq 1)$ as a sequence of random variables which take values $E_1, \cdots, E_k$ in each step n and have probabilities as follows, for each $n\geqq 1$

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

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

Proprety of our procedure $\mathfrak{L}^*$. By Lemma 4 we observed the fact that in any sequence $E^{(1)}, E^{(2)}, \cdots$ there exists a positive vector $\boldsymbol{\lambda}_0$ such that $\boldsymbol{\lambda}_n$ converges to $\boldsymbol{\lambda}_0$. Using this fact we get under our procedure $\mathfrak{L}^*$ n_j/n converges to j-th component $\lambda_{0j}(>0)$ of the limit vector $\boldsymbol{\lambda}_0$ by the strong law of large numbers. Hence under our procedure $\mathfrak{L}^*$ $n_j\to\infty$ as $n\to\infty$ for any $j=1, \cdots, k$ is satisfied so that $\hat{\theta}_n$ converges to the unknown true parameter θ as to be proved.

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

Next we suppose the unknown true parameter θ is not an element of π. Then there exists a positive vector $\boldsymbol{\lambda}$ uniquely given by $\boldsymbol{\theta}$, as we defined $\boldsymbol{\lambda}_n$ uniquely by $\hat{\theta}_{n-1}$ and $\boldsymbol{\lambda}_0$ uniquely by θ_0, such that $\lim_{n\to\infty}\boldsymbol{\lambda}_n=\boldsymbol{\lambda}$ is satisfied with probability 1. This fact is shown as discussion of Lemma 4. Where the vector $\boldsymbol{\lambda}=(\lambda_1, \cdots, \lambda_k)$ is defined satisfying $\sum_{j=1}^{k}\lambda_j=1$ and analogousely as (3. 3)

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

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

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

satisfying that analogously as (3. 4)

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

Otherwise if the unknown true parameter θ is an element of π, then under our procedure $\mathfrak{L}^*$, from the result of Lemma 4, $\boldsymbol{\lambda}_n$ converges to a positive vector with probability 1. Hence under our procedure $\mathfrak{L}^*$ if $\theta\notin\pi$ then n_j/n converges to λ_j $(j=1, \cdots, k)$ 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* $\mathfrak{L}^*$, *if the true unknown parameter* θ *is not an element of* π *then* $(n_1/n, \cdots, n_k/n)$ *converges to our vector* $\boldsymbol{\lambda}(>0)$, *defined in* (3. 13), *is satisfied with probability* 1, *and otherwise if the true unknown parameter* θ *is an element of* π *then* $(n_1/n, \cdots, n_k/n)$ *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 $\mathfrak{L}^*$ that if $\theta\notin\pi$, then $n_j/n\to\lambda_j$ with probability 1, and otherwise if $\theta\in\pi$, then n_j/n converges to a positive value ($j=1, \cdots, k$) with probability 1. In the following line we call this proposition as *optimal condition* and the vector $\boldsymbol{\lambda}$ 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 $\lim_{n\to\infty}\tilde{\theta}_n=\theta^*$ with probability 1, from the two equalities (2. 14), (3. 13)

$$(2.\,14)' \qquad \frac{n_1}{\mu_1}\left[\frac{dI(\hat{\theta}_n, \varphi, E_1)}{d\varphi_1}\right]_{\varphi=\tilde{\theta}_n}=\cdots=\frac{n_k}{\mu_k}\left[\frac{dI(\hat{\theta}_n, \varphi, E_k)}{d\varphi_k}\right]_{\varphi=\tilde{\theta}_n},$$

$$(3.\,13)' \qquad \frac{\lambda_1}{\mu_1}\left[\frac{dI(\theta, \varphi, E_1)}{d\varphi_1}\right]_{\varphi=\theta^*}=\cdots=\frac{\lambda_k}{\mu_k}\left[\frac{dI(\theta, \varphi, E_k)}{d\varphi_k}\right]_{\varphi=\theta^*}.$$

So we get

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

with probability 1, where $I^*(\theta)$ is the value of (3. 14)

$$(3.\,14)' \qquad I^*(\theta)=\frac{I(\theta, \theta^*, E_1)}{c_1}=\cdots=\frac{I(\theta, \theta^*, E_k)}{c_k}.$$

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

$$(4.\,1) \qquad \frac{S_n(\hat{\theta}_n, \tilde{\theta}_n)}{\sum_{i=1}^n C^{(i)}}=\frac{n_1 I(\hat{\theta}_n, \tilde{\theta}_n, E_1)+\cdots+n_k I(\hat{\theta}_n, \tilde{\theta}_n, E_k)}{n_1 c_1+\cdots+n_k c_k}.$$

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 $\lim_{n\to\infty}\tilde{\theta}_n=\theta$ with probability 1. So we get

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

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*

$$\lim_{n\to\infty}\frac{S_n(\hat{\theta}_n, \tilde{\theta}_n)}{\sum_{i=1}^{n} C^{(i)}}=I^*(\theta) \tag{4. 2}$$

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

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

converges to zero with probability 1.

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

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

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

$$\frac{\lambda_1'}{\mu_1}\left[\frac{dI(\theta, \varphi, E_1)}{d\varphi_1}\right]_{\varphi=\theta^{**}}=\cdots=\frac{\lambda_k'}{\mu_k}\left[\frac{dI(\theta, \varphi, E_k)}{d\varphi_k}\right]_{\varphi=\theta^{**}}. \tag{4. 3}$$

From the equality (2. 14)′ defining $\tilde{\theta}_n$ uniquely on our hyperplane π we can verify $\lim_{n\to\infty}\tilde{\theta}_n=\theta^{**}$ with probability 1. Therefore

$$\lim_{n\to\infty}\frac{S_n(\hat{\theta}_n, \tilde{\theta}_n)}{\sum_{i=1}^{n} C^{(i)}}=\frac{\lambda_1'I(\theta, \theta^{**}, E_1)+\cdots+\lambda_k'I(\theta, \theta^{**}, E_k)}{\lambda_1'c_1+\cdots+\lambda_k'c_k} \tag{4. 4}$$

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

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

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

Therefore we have

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

Otherwise if the true parameter θ is an element of π then we have $\lim_{n\to\infty}\tilde{\theta}_n=\theta$ 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* $\mathfrak{L}^{**}$, *satisfying* $\min_j(n_j)\to\infty$ *as* $n\to\infty$ *and* $(n_1/n, \cdots, n_k/n)$ *converges to* $\boldsymbol{\lambda}'=(\lambda_1', \cdots, \lambda_k')$ *as* $n\to\infty$ $(\boldsymbol{\lambda}'\neq\boldsymbol{\lambda})$, *if the true unknown parameter* θ *is not an element of* π *then we get*

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

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 $\mathfrak{L}^{*}$ as *optimal condition* or having *optimal ratio vector.*

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

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

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

Under this procedure $\mathfrak{L}$ 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 $\boldsymbol{\lambda}_n$ converges to a *positive ratio vector* with probability 1. And also the procedure $\mathfrak{L}$ has a property $n_j/n-\lambda_{nj}$ converges to zero as $n\to\infty$ 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 $\mathfrak{L}$ we have $\min(n_1, n_2) \to \infty$ as $n \to \infty$ with probability 1. Hence under the procedure $\mathfrak{L}$ we have $\hat{\theta}_n \to \theta$ as $n \to \infty$ with probability 1. If the true parameter θ is not an element of π then $\boldsymbol{\lambda}_n$ converges to $\boldsymbol{\lambda}$ 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 $\mathfrak{L}$ 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 $\mathfrak{L}$ also has the optimal property, that is, most informative relative to the sum of costs than any other procedure having the two conditions $n_j \to \infty$ as $n \to \infty$, $(n_1/n, n_2/n)$ converges to $\boldsymbol{\lambda}'=(\lambda_1', \lambda_2')$ as $n \to \infty$ $(\boldsymbol{\lambda}' \neq \boldsymbol{\lambda})$.

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

$$f(x, \theta, E_j) = \frac{1}{\sqrt{2\pi}\,\sigma_j} \exp\left\{-\frac{(x-m_j)^2}{2\sigma_j^2}\right\} \qquad (j=1, 2) \tag{5. 2}$$

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

$$I(\theta, \varphi, E_j) = \frac{(m_j - m_j^*)^2}{2\sigma_j^2} \qquad (j=1, 2) \tag{5. 3}$$

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

$$H_1=\{\theta:\ \mu_1 m_1+\mu_2 m_2>p\}, \qquad H_2=\{\theta:\ \mu_1 m_1+\mu_2 m_2<p\},$$
$$\pi=\{\theta:\ \mu_1 m_1+\mu_2 m_2=p\}. \tag{5. 4}$$

And $\lim_{n\to\infty}\tilde{\theta}_n=\theta^*$ is given uniquely by the equality (3. 14) and $\theta^* \in \pi$.

$$\frac{(m_1-m_1^*)^2}{2c_1\sigma_1^2} = \frac{(m_2-m_2^*)^2}{2c_2\sigma_2^2} \tag{5. 5}$$

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

$$H_1=\{(m_1, m_2):\ m_1>m_2\}, \qquad H_2=\{(m_1, m_2):\ m_1<m_2\},$$
$$\pi=\{(m_1, m_2):\ m_1=m_2\}. \tag{5. 6}$$

In this case we can get the *optimal ratio* $\boldsymbol{\lambda}$ as following by (3. 13)

$$\frac{\lambda_1(m_1-m_1^*)}{\sigma_1^2} = -\frac{\lambda_2(m_2-m_2^*)}{\sigma_2^2} \tag{5. 7}$$

Therefore from the two relations (5. 5), (5. 7) we get the *optimal ratio* $\boldsymbol{\lambda}$ as next value uniformly on Θ.[6]

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

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

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

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 auther 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].