ON THE EMPIRICAL BAYES PROCEDURE (1)

HIROSI HUDIMOTO

(Received Jan. 20, 1968)

1. Introduction

Let X be a random variable which is known to have one of the distinct distribution functions $F_1(x), \cdots,$ and $F_r(x)$, i.e. $P(X \leqq x \mid \theta = i) = F_i(x)$. When we denote by $L(\varphi(x) \mid \theta)$ the loss incurred in using a rule $\varphi(x)$ for the decision about unknown θ based on an observed value x on X, the risk function is given by $R_\theta(\varphi) = \int L(\varphi(x) \mid \theta) dF_\theta(x)$. When we consider θ as a random parameter, if the prior distribution $\boldsymbol{\omega}' = (\omega_1, \cdots, \omega_r)$ is given, i.e. $P(\theta = i) = \omega_i$ for $i = 1, \cdots, r$, the expected risk with respect to this $\boldsymbol{\omega}'$ is $R_{\boldsymbol{\omega}}(\varphi) = \sum_{i=1}^{r} \omega_i R_i(\varphi)$. We are interested in the rule $\varphi_{\boldsymbol{\omega}}(x)$ minimizing $R_{\boldsymbol{\omega}}(\varphi)$ for fixed $\boldsymbol{\omega}'$. The rule $\varphi_{\boldsymbol{\omega}}(x)$ is called a Bayes rule and the corresponding Bayes risk is given by $B(\boldsymbol{\omega}) = R_{\boldsymbol{\omega}}(\varphi_{\boldsymbol{\omega}})$.

On the other hand, in the case that $\boldsymbol{\omega}'$ is unknown, the general consideration on asking for $\varphi_n(x)$ such that $\lim_{n\to\infty} R_{\boldsymbol{\omega}}(\varphi_n) = B(\boldsymbol{\omega})$, on the basis of $(X_1, \cdots, X_n)$ considered to be a random sample from the population π with the distribution function $F(x) = \sum_{i=1}^{r} \omega_i F_i(x)$, was suggested by H. Robbins [12]. His paper contained more general case about $\boldsymbol{\omega}'$ than one as mentioned herein and the problem was reduced to asking for an estimator for $\boldsymbol{\omega}'$ that converges in probability to $\boldsymbol{\omega}'$ as $n \to \infty$. Especially, in the case of $\boldsymbol{\omega}' = (\omega_1, \cdots, \omega_r)$, he showed the general form of unbiased estimators for $\boldsymbol{\omega}'$.

In this paper, we shall be concerned with an embodiment of his idea for the case of $\boldsymbol{\omega}' = (\omega_1, \cdots, \omega_r)$. At the outset, we shall deal with the case of $\boldsymbol{\omega}' = (\omega_0, \omega_1)$ because of easy handling. Originally, H. Robbins brought forward a problem termed by him the compound decision problem which arose when one was repeatedly confronted with the same decision problem n times for such a case (of $r=2$), [10]. His aim was to bring the overall expected loss below the level attainable through n independent applications of the most powerful test in the case of testing type, and the possibility was shown for the case of normal distribution with known variance, say $\sigma^2=1$, and with an expected mean θ which was known to have one of only two values, either $\theta=-1$ or $\theta=1$. J. Neyman applaudingly mentioned his work "Robbins' breakthrough on the Bayes' front" in [9]. Consideration on the same or on the allied subject has appeared in [3], [4], [8], [14], and [15].

In section 2, we shall be concerned with estimating $\boldsymbol{\omega}'=(\omega_0, \omega_1)$ based on a random sample $(X_1, \cdots, X_n)$ from the population π with the distribution function $F(x)=\omega_0F_0(x)+\omega_1F_1(x)$, and we shall make use of the method proposed in [6]. Consider $\hat{p}_0=\frac{1}{n}\sum_{k=1}^{n}F_0(X_k)$ and $\hat{p}_1=\frac{1}{n}\sum_{k=1}^{n}F_1(X_k)$. Then, $\hat{\omega}_1=\frac{1}{2}+\left(\hat{h}-\frac{1}{2}\right)\Delta^{-1}$ is an unbiased estimator for unknown ω_1, where $\hat{h}=\frac{1}{2}\{\hat{p}_0+\hat{p}_1\}$ and $\Delta=\int F_0(x)dF_1(x)-\frac{1}{2}$. E. J. G. Pitman has given the maximum likelihood estimate for $\boldsymbol{\omega}'=(\omega_0, \omega_1)$, [13], but our method is applicable to the non-parametric case if past samples from $F_0(x)$ and $F_1(x)$ are available. When we take $\hat{\omega}_1$ as an estimator for ω_1, a confidence interval for ω_1 with the confidence coefficient not less than $(1-\alpha)$ for any given α $(1>\alpha>0)$ is given and an asymptotically confidence interval can be also obtained since it is shown that $\hat{\omega}_1$ is asymptotically normally distributed with the expected mean ω_1 and the variance $\frac{1}{12n\Delta^2}$ in section 3, and the empirical Bayes two-way decision procedure is discussed in section 4. The case of two p-variate normal distributions with equal covariance matrices is shown in section 5.

The case of $\boldsymbol{\omega}'=(\omega_1, \cdots, \omega_r)$ is considered in sections 6, 7, where $\boldsymbol{\omega}'$ denotes the transpose of the r-column vector $\boldsymbol{\omega}$. We shall take $\hat{\boldsymbol{\omega}}=A^{-1}\hat{\boldsymbol{p}}$ as an unbiased estimator for $\boldsymbol{\omega}$ if $|A|\neq 0$, where A and $\hat{\boldsymbol{p}}$ are the $(r\times r)$-matrix and the r-column vector such that $A=\left(\frac{1}{2}+\Delta_{ij}\right)=\left(\int F_i(x)\cdot dF_j(x)\right)$ and $\hat{\boldsymbol{p}}=(\hat{p}_i)=\left(\frac{1}{n}\sum_{k=1}^{n}F_i(X_k)\right)$ $(i, j=1, \cdots, r)$.

2. Estimating the prior distribution $\boldsymbol{\omega}'$

Consider a population π of individuals such that with each individual is associated a value of the observable one-dimensional random variable X. Let π be divided into two sub-populations π_0 and π_1, and let X have the distribution function $F_0(x)$ in π_0 and $F_1(x)$ in π_1. For convenience, it will be assumed that $F_0(x)$ and $F_1(x)$ are continuous and $F_0(x)>F_1(x)$ for all x. If the prior probabilities of an individual of π belonging to π_0 and π_1 are denoted by ω_0 and ω_1, X is considered to have the distribution function

$$F(x)=\omega_0 F_0(x)+\omega_1 F_1(x)\,. \tag{2.1}$$

But, in the usual case, $\boldsymbol{\omega}'=(\omega_0, \omega_1)$ is unknown.

Let

$$(X_1, \cdots, X_n) \tag{2.2}$$

be a random sample of size n from π. That is, (2.2) is a mixed sample drawn from π_0 and π_1 with (prior) probabilities ω_0 and ω_1. For instance, such a situation will arise in the two-way classification (or the two-way allocation) problem. Previously in [6] the author treated the case that (2.2) is a random sample from either π_0 or π_1. The procedure used in [6] will be extended in a similar manner in this paper. We shall be concerned with estimating them on the basis of the sample (2.2) in the case that ω_0 and ω_1 are unknown.

At the outset, we assume that $F_0(x)$ and $F_1(x)$ are completely known. Then, if we define

$$\begin{aligned}\Delta&=\int_{-\infty}^{\infty}[F_0(x)-F_1(x)]dF(x)\\&=\int_{-\infty}^{\infty}F_0(x)dF_1(x)-\frac{1}{2}\,,\end{aligned} \tag{2.3}$$

Δ is determined for specified $F_0(x)$ and $F_1(x)$, and takes on a positive value because of the assumption of stochastic ordering as described above.

Consider two statistics

$$\hat{p}_0=\frac{1}{n}\sum_{k=1}^{n}F_0(X_k)\,, \tag{2.4}$$

and

$$\hat{p}_1=\frac{1}{n}\sum_{k=1}^{n}F_1(X_k)\,. \tag{2.5}$$

Then, their expectations with respect to $F(x)$ are

$$\mathcal{E}\{\hat{p}_0\}=\int_{-\infty}^{\infty}F_0(x)dF(x)=\frac{1}{2}+\omega_1\Delta\,,$$

and

$$\mathcal{E}\{\hat{p}_1\}=\int_{-\infty}^{\infty}F_1(x)dF(x)=\frac{1}{2}-\omega_0\Delta\,.$$

Therefore, if we take

$$\hat{\omega}_0=\left(\frac{1}{2}-\hat{p}_1\right)\Delta^{-1} \quad\text{and}\quad \hat{\omega}_1=\left(\hat{p}_0-\frac{1}{2}\right)\Delta^{-1} \tag{2.6}$$

as estimators for ω_0 and ω_1, we have

$$\mathcal{E}\{\hat{\hat{\omega}}_0\}=\omega_0 \quad \text{and} \quad \mathcal{E}\{\hat{\hat{\omega}}_1\}=\omega_1 .$$

But these estimators have not always the properties (i) $\hat{\hat{\omega}}_0+\hat{\hat{\omega}}_1=1$ and (ii) $\hat{\hat{\omega}}_i\geqq 0$ $(i=0, 1)$, that must be satisfied by a probability distribution. However, so far as relating to (i), if we take

$$\tilde{\omega}_0=\left(\frac{1}{2}-\hat{p}_1\right)\tilde{\varDelta}^{-1} \quad \text{and} \quad \tilde{\omega}_1=\left(\hat{p}_0-\frac{1}{2}\right)\tilde{\varDelta}^{-1}, \tag{2.7}$$

where $\tilde{\varDelta}=\hat{p}_0-\hat{p}_1$, $\tilde{\omega}_0$ and $\tilde{\omega}_1$ have the property $\tilde{\omega}_0+\tilde{\omega}_1=1$. Since $\tilde{\varDelta}$ is an unbiased estimator for $\varDelta$ with finite variance, we have $\tilde{\omega}_i\rightarrow\omega_i$ $(i=0, 1)$ with probability one as $n\rightarrow\infty$.

A modification of (2.6) that satisfies (i) can also be given as follows. If we take

$$\begin{aligned}\hat{\omega}_0&=\frac{1}{2}-\left(\hat{h}-\frac{1}{2}\right)\varDelta^{-1}\\ \hat{\omega}_1&=\frac{1}{2}+\left(\hat{h}-\frac{1}{2}\right)\varDelta^{-1} .\end{aligned} \tag{2.8}$$

where $\hat{h}=\frac{1}{2}\{\hat{p}_0+\hat{p}_1\}$, we have

$$\mathcal{E}\{\hat{\omega}_i\}=\omega_i \qquad (i=0, 1) .$$

The assumption of ordering between $F_0(x)$ and $F_1(x)$ is not essential for our method and the argument is valid for $\varDelta\neq 0$. When $\varDelta<0$, replace F_0 and F_1 by each other. Then $\varDelta'=\int F_1 dF_0-\frac{1}{2}>0$. Hereafter, we shall, without loss of generality, assume that $\varDelta>0$.

Suppose $(X_1, \cdots, X_n)$ *is a random sample drawn from the population* π *with the distribution function* $F(x)$ *defined in* (2.1). (*It is assumed that* $F_0(x)$ *and* $F_1(x)$ *are continuous and* $\varDelta>0$, *where* $\varDelta$ *is defined by* (2.3).) *Then, if we take* $\hat{\hat{\boldsymbol{\omega}}}'=(\hat{\hat{\omega}}_0, \hat{\hat{\omega}}_1)$, $\hat{\boldsymbol{\omega}}'=(\hat{\omega}_0, \hat{\omega}_1)$ *or* $\tilde{\boldsymbol{\omega}}'=(\tilde{\omega}_0, \tilde{\omega}_1)$ *as estimators for components of* $\boldsymbol{\omega}'=(\omega_0, \omega_1)$, $\hat{\hat{\boldsymbol{\omega}}}'$ *and* $\hat{\boldsymbol{\omega}}'$ *are unbiased estimators for* $\boldsymbol{\omega}'$, *and* $\tilde{\boldsymbol{\omega}}'$ *is a consistent estimator for* $\boldsymbol{\omega}'$, *where* $\hat{\hat{\boldsymbol{\omega}}}'$, $\hat{\boldsymbol{\omega}}'$ *and* $\tilde{\boldsymbol{\omega}}'$ *are defined by* (2.6), (2.8) *and* (2.7).

The property (ii) that estimators for ω_0 and ω_1 take on non-negative values is usually more important than the property (i) from a viewpoint that they denote weights with which we consider that an observation will come from π_0 or π_1 in future experiment. We shall consider $\hat{\omega}_0$ and $\hat{\omega}_1$ given in (2.8) as to (ii), since other estimators can also be handled in the same manner. The property (ii) for $\hat{\omega}_0$ and $\hat{\omega}_1$ means

$$\frac{1}{2}-\frac{\Delta}{2}\leqq\hat{h}\leqq\frac{1}{2}+\frac{\Delta}{2}, \tag{2.9}$$

if $\Delta\geqq 0$. Expectation of $\hat{h}$ with respect to $F(x)$ is

$$\begin{aligned}\mathcal{E}\{\hat{h}\} &= \frac{1}{2}+\frac{\Delta}{2}-\omega_0\Delta \\ &= \frac{1}{2}-\frac{\Delta}{2}+\omega_1\Delta .\end{aligned} \tag{2.10}$$

We shall make use of the following theorem given by W. Hoeffding [5].

If random variables $X_1, \cdots, X_n$ are independent and $a_k \leqq X_k \leqq b_k$ for $k=1, \cdots, n$, then for $t>0$

$$P\{\bar{X}-\mu\geqq t\}\leqq\exp\left\{-2n^2t^2\Big/\sum_{k=1}^{n}(b_k-a_k)^2\right\}, \tag{2.11}$$

where $\bar{X}=\sum_{k=1}^{n}X_k/n$ and $\mu=\mathcal{E}\{\bar{X}\}$.

Since $\{F_0(X_1)+F_1(X_1)\}/2, \cdots, \{F_0(X_n)+F_1(X_n)\}/2$ are independent and $0\leqq\{F_0(X_k)+F_1(X_k)\}/2\leqq 1$ for $k=1, \cdots, n$, Hoeffding's inequality as mentioned above gives us

$$P\left\{\hat{h}-\left(\frac{1}{2}+\frac{\Delta}{2}-\omega_0\Delta\right)\geqq t\right\}\leqq e^{-2nt^2} \tag{2.12}$$

or

$$P\left\{\left(\frac{1}{2}-\frac{\Delta}{2}+\omega_1\Delta\right)-\hat{h}\geqq t\right\}\leqq e^{-2nt^2} \tag{2.13}$$

for $t>0$. (Note that an upper bound for $P\{\bar{X}-\mu\geqq t\}$ is also an upper bound for $P\{-\bar{X}+\mu\geqq t\}$.) Putting $t=\omega_0\Delta$ in (2.12) and $t=\omega_1\Delta$ in (2.13) if $\Delta>0$, we get

$$P\left\{\hat{h}\geqq\frac{1}{2}+\frac{\Delta}{2}\right\}\leqq e^{-2n(\omega_0\Delta)^2}$$

and

$$P\left\{\hat{h}\leqq\frac{1}{2}-\frac{\Delta}{2}\right\}\leqq e^{-2n(\omega_1\Delta)^2}.$$

Thus, if $0<c\leqq\min\{\omega_0, \omega_1\}$ and if we can take a sample size n such that $n\geqq\log\frac{1}{\alpha}\Big/2c^2\Delta^2$ for $0<\alpha<1$, we have

$$P\{\hat{\omega}_1 \leqq 0\} \leqq \alpha \quad \text{and} \quad P\{\hat{\omega}_1 \geqq 1\} \leqq \alpha .$$

In this sense, we shall always infer that $\omega_1=0$ when $\hat{\omega}_1 \leqq 0$ and that $\omega_1=1$ when $\hat{\omega}_1 \geqq 1$, and that $\omega_1=\hat{\omega}_1$ when $0 \leqq \hat{\omega}_1 \leqq 1$.

In the analogous form to (2.7), the method can be extended to the non-parametric case that the distribution functions $F_0(x)$ and $F_1(x)$ are unknown, but random samples of size n_0 and n_1 drawn from π_0 and π_1 are available. Such a situation will arise when our knowledge based on past experiments about distributions on π_0 and π_1 is not perfect.

Suppose that random samples $(X_1^{(0)}, \cdots, X_{n_0}^{(0)})$ and $(X_1^{(1)}, \cdots, X_{n_1}^{(1)})$ have been obtained from π_0 and π_1, respectively. Let $(X_1, \cdots, X_n)$ be a new random sample drawn from π. For these samples, define

$$(2.14) \quad \tilde{p}_0 = \frac{[\text{the number of pairs } (X_u^{(0)}, X_k) \text{ such that } X_u^{(0)} \leqq X_k]}{n_0 n} ,$$

and

$$(2.15) \quad \tilde{p}_1 = \frac{[\text{the number of pairs } (X_v^{(1)}, X_k) \text{ such that } X_v^{(1)} \leqq X_k]}{n_1 n} ,$$

$u=1, \cdots, n_0$, $v=1, \cdots, n_1$ and $k=1, \cdots, n$, and set

$$(2.16) \qquad \tilde{\tilde{\omega}}_0 = \left(\frac{1}{2} - \tilde{p}_1\right)\tilde{\tilde{\Delta}}^{-1} \quad \text{and} \quad \tilde{\tilde{\omega}}_1 = \left(\tilde{p}_0 - \frac{1}{2}\right)\tilde{\tilde{\Delta}}^{-1} ,$$

where $\tilde{\tilde{\Delta}} = \tilde{p}_0 - \tilde{p}_1$. This is (2.7) with $\tilde{p}_0$, $\tilde{p}_1$ replacing $\hat{p}_0$, $\hat{p}_1$. Since $\left(\frac{1}{2} - \tilde{p}_1\right)$, $\left(\tilde{p}_0 - \frac{1}{2}\right)$, and $\tilde{\tilde{\Delta}}$ are unbiased estimators with finite variances for $\omega_0 \Delta$, $\omega_1 \Delta$ and Δ respectively, the following corollary is obtained.

If $\Delta > 0$, $\tilde{\tilde{\boldsymbol{\omega}}}' = (\tilde{\tilde{\omega}}_0, \tilde{\tilde{\omega}}_1)$ *defined in* (2.16) *are consistent estimators for* $\boldsymbol{\omega}'$.

On the other hand, even if $F_0(x)$ and $F_1(x)$ are completely specified, it may happen that calculation of $\hat{p}_0$ and $\hat{p}_1$ is troublesome. In such a case, take random numbers $U_1, \cdots, U_{n_0}$ and $V_1, \cdots, V_{n_1}$ uniformly distributed on [0, 1], and transform them to $X_u^{(0)} = F_0^{-1}(U_u)$ and $X_v^{(1)} = F_1^{-1}(V_v)$ for $u=1, \cdots, n_0$ and $v=1, \cdots, n_1$. For these $X_u^{(0)}$'s, $X_v^{(1)}$'s and a sample (2.2), consider $\tilde{p}_0$ and $\tilde{p}_1$ given by (2.14) and (2.15). Then,

$$(2.17) \qquad \hat{\omega}_1^* = \frac{1}{2} + \left(\tilde{h} - \frac{1}{2}\right)\Delta^{-1}$$

is an unbiased estimator for ω_1, where $\tilde{h} = \frac{1}{2}\{\tilde{p}_0 + \tilde{p}_1\}$. The same can be said also of

(2.18) $$\hat{\omega}_0^* = \left(\frac{1}{2} - \tilde{p}_1\right)\Delta^{-1} \quad \text{and} \quad \hat{\omega}_1^* = \left(\tilde{p}_0 - \frac{1}{2}\right)\Delta^{-1}.$$

3. Confidence intervals for $\boldsymbol{\omega}'$

We shall consider to constitute confidence intervals for ω_1 in the case of $\boldsymbol{\omega}'=(\omega_0, \omega_1)$. In the case that $F_0(x)$ and $F_1(x)$ are completely specified, if we take $\hat{\boldsymbol{\omega}}'$ given in (2.8) as an estimator for $\boldsymbol{\omega}'$, Hoeffding's inequality as mentioned above can be used for this purpose. If $\Delta>0$, a slight modification of (2.12) and (2.13) gives

(3.1) $$P\{\omega_1 \leqq \hat{\omega}_1 - e(\Delta, n, \alpha/2)\} \leqq \alpha/2,$$

and

(3.2) $$P\{\omega_1 \geqq \hat{\omega}_1 + e(\Delta, n, \alpha/2)\} \leqq \alpha/2,$$

where $e(\Delta, n, \alpha/2) = \sqrt{\frac{1}{2n\Delta^2}\log\frac{2}{\alpha}}$. Thus we have

(3.3) $$P\{|\hat{\omega}_1 - \omega_1| < e(\Delta, n, \alpha/2)\} \geqq 1-\alpha.$$

Hence, we can state as follows.

If we take $\hat{\boldsymbol{\omega}}'=(\hat{\omega}_0, \hat{\omega}_1)$ as an estimator for $\boldsymbol{\omega}'=(\omega_0, \omega_1)$, a confidence interval for ω_1 with the confidence coefficient not less than $(1-\alpha)$ $(0<\alpha<1)$ is given by $(\hat{\omega}_1 - e(\Delta, n, \alpha/2), \hat{\omega}_1 + e(\Delta, n, \alpha/2))$, where $e(\Delta, n, \alpha/2) = \sqrt{\frac{1}{2n\Delta^2}\log\frac{2}{\alpha}}$.

The relation (3.3) with $\hat{\omega}_1^*$ replacing $\hat{\omega}_1$ also holds true, that is,

$$P\{|\hat{\omega}_1^* - \omega_1| < e(\Delta, n, \alpha/2)\} \geqq 1-\alpha,$$

where $\hat{\omega}_1^*$ is an expedient estimator for calculation and is defined in (2.17). The verification has been given in [6].

On the other hand, we can derive an asymptotic distribution of estimators for ω_1 that is expected to give a good approximation for our aim in the case of a fairly large sample size and is considered to be practically convenient. Now $\omega_0 F_0(X_k) + \omega_1 F_1(X_k)$ is distributed with the uniform distribution on $[0, 1]$ for each random variable X_k contained in a sample (2.2) because each X_k is distributed with the distribution function $\omega_0 F_0(x) + \omega_1 F_1(x)$, $k=1, \cdots, n$, so, $\omega_0 \hat{p}_0 + \omega_1 \hat{p}_1$ has a probability distribution which approaches to the normal distribution as the sample size increases, and the expected mean and the variance are $1/2$ and $1/12n$, respectively. Therefore, since

$$\omega_0\hat{p}_0+\omega_1\hat{p}_1-\frac{1}{2}=\tilde{\Delta}(\tilde{\omega}_1-\omega_1)\,,$$

the probability distribution of $\tilde{\Delta}(\tilde{\omega}_1-\omega_1)$ approaches to the normal distribution with the expected mean zero as $n\to\infty$, and since $\tilde{\Delta}\to\Delta$ with probability 1 as $n\to\infty$, if $\Delta>0$, $(\hat{\omega}_1-\omega_1)$ is asymptotically normally distributed with the expected mean 0 and variance $\dfrac{1}{12n\Delta^2}$ for large n.

As for the other estimators $\hat{\hat{\omega}}_1$ and $\hat{\omega}_1$, since we have

$$\frac{\omega_0\hat{p}_0+\omega_1\hat{p}_1-\dfrac{1}{2}}{\Delta}=\hat{\hat{\omega}}_1-\frac{\hat{\hat{\Delta}}}{\Delta}\omega_1$$

and

$$\frac{\omega_0\hat{p}_0+\omega_1\hat{p}_1-\dfrac{1}{2}}{\Delta}=\left(\hat{\omega}_1-\frac{1}{2}\right)-\frac{\tilde{\Delta}}{\Delta}\left(\omega_1-\frac{1}{2}\right)\,,$$

the result as mentioned above also holds for $(\hat{\hat{\omega}}_1-\omega_1)$ and $(\hat{\omega}_1-\omega_1)$. Thus, we obtain the following theorem.

When we take $\hat{\hat{\omega}}_1$, $\tilde{\omega}_1$ or $\hat{\omega}_1$ *given in* (2.6), (2.7) *or* (2.8) *as estimators for* ω_1, *if* $\Delta>0$, $\hat{\hat{\omega}}_1$, $\tilde{\omega}_1$ *and* $\hat{\omega}_1$ *are asymptotically normally distributed with the expected mean* ω_1 *and the variance* $\dfrac{1}{12n\Delta^2}$ *for large* n.

An asymptotically confidence interval for ω_1 with the confidence coefficient $(1-\alpha)$ is given by $\left(\hat{\hat{\omega}}_1-\dfrac{k_\alpha}{\Delta\sqrt{12n}},\hat{\hat{\omega}}_1+\dfrac{k_\alpha}{\Delta\sqrt{12n}}\right)$, $\left(\tilde{\omega}_1-\dfrac{k_\alpha}{\Delta\sqrt{12n}},\tilde{\omega}_1+\dfrac{k_\alpha}{\Delta\sqrt{12n}}\right)$ or $\left(\hat{\omega}_1-\dfrac{k_\alpha}{\Delta\sqrt{12n}},\hat{\omega}_1+\dfrac{k_\alpha}{\Delta\sqrt{12n}}\right)$, where k_α is determined from

$$\frac{1}{\sqrt{2\pi}}\int_{-k_\alpha}^{k_\alpha}e^{-t^2/2}dt=1-\alpha\,.$$

4. The empirical Bayes two-way decision procedure

Let π_0 and π_1 be two populations and $F_0(x)$, $F_1(x)$ the distribution functions of π_0 and π_1, respectively. Let X be a random variable which is known to have one of two distinct distribution functions $F_0(x)$ and $F_1(x)$. Consider the problem to decide whether an individual belongs to π_0 or π_1, i.e. the random variable X has $F_0(x)$ or $F_1(x)$, on the basis of an observed value x on X. Let the loss structure be as follows:

		decision π_0	decision π_1
population	π_0	0	$L(1\mid 0)$
	π_1	$L(0\mid 1)$	0

where $L(0\mid 1)$ and $L(1\mid 0)$ are two given positive numbers.

A randomized decision rule $\varphi(x)$ is a measurable function with respect to σ-field on the sample space with $0\leqq\varphi(x)\leqq 1$, that is, when $X=x$ is observed, one decides $\theta=1$ or $\theta=0$ with probabilities $\varphi(x)$ or $1-\varphi(x)$. Then, for any rule $\varphi(x)$, the expected losses due to an incorrect decision, called the risk functions of $\varphi(x)$, are given by

$$(4.1)\qquad \begin{aligned} &R_0(\varphi)=L(1\mid 0)\mathcal{E}_0\{\varphi(X)\} \\ \text{and}\quad &R_1(\varphi)=L(0\mid 1)\mathcal{E}_1\{1-\varphi(X)\}\,, \end{aligned}$$

where $\mathcal{E}_\theta$ denotes expectation with respect to $F_\theta(x)$ $(\theta=0, 1)$.

Now, suppose that π consists of sub-populations π_0 and π_1, and assume that the prior distribution is given by $\boldsymbol{\omega}'=(\omega_0, \omega_1)$, where $\omega_i=P(\theta=i)$ for $i=0$ and 1, that is, an individual of π drawn at random belongs to π_0 and π_1 with probabilities ω_0 and ω_1, respectively. Then, for the rule $\varphi(x)$, the expected risk with respect to this prior distribution is

$$(4.2)\qquad R_{\boldsymbol{\omega}}(\varphi)=(1-\omega_1)R_0(\varphi)+\omega_1R_1(\varphi)\,,$$

and one finds rules $\varphi_{\boldsymbol{\omega}}(x)$ minimizing (4.2) for fixed $\boldsymbol{\omega}'$.

Let $f_\theta(x)$ denote the likelihood function of θ given x. Then, for given $\boldsymbol{\omega}'$, (4.2) is minimized by $\varphi_{\boldsymbol{\omega}}(x)$ of the form

$$(4.3)\qquad \varphi_{\boldsymbol{\omega}}(x)=\begin{cases} 0 & \text{if } L(1\mid 0)(1-\omega_1)f_0(x)\leqq L(0\mid 1)\omega_1 f_1(x) \\ 1 & \text{if } L(1\mid 0)(1-\omega_1)f_0(x)> L(0\mid 1)\omega_1 f_1(x)\,. \end{cases}$$

The rule $\varphi_{\boldsymbol{\omega}}(x)$ is usually called a Bayes rule with respect to the prior distribution $\boldsymbol{\omega}'$, and the corresponding Bayes risk is

$$(4.4)\qquad B(\boldsymbol{\omega})=R_{\boldsymbol{\omega}}(\varphi_{\boldsymbol{\omega}})=\min_{\varphi} R_{\boldsymbol{\omega}}(\varphi)\,.$$

Suppose now that such decision problems occur repeatedly and independently with the same but unknown $\boldsymbol{\omega}'$. Let

$$(4.5)\qquad (X_1, \theta_1), \cdots, (X_n, \theta_n), (X_{n+1}, \theta_{n+1}), \cdots$$

be a sequence of pairs of random variables, each pair being independent of all the other pairs, where the value of θ_i $(i=1, \cdots, n, n+1, \cdots)$ appears according to the common prior distribution $\boldsymbol{\omega}'$. When the decision about parameter is made, we shall have observed values on $X_1, \cdots, X_n$

and X_{n+1}, but the values of $\theta_1, \cdots$, and θ_n are still unknown. However, we can use any estimate for $\boldsymbol{\omega}'$ based on $X_1, \cdots, X_n$ for the decision about θ_{n+1}.

Let $\varphi_{\hat{\boldsymbol{\omega}}}(x)$ denote the rule of the form (4.3) with $\boldsymbol{\omega}'$ replaced by the estimator $\hat{\boldsymbol{\omega}}'$ given by (2.8) based on $(X_1, \cdots, X_n)$. If we adopt the rule $\varphi_{\hat{\boldsymbol{\omega}}}(x)$, the actual expected risk with respect to $\boldsymbol{\omega}'$ will become

$$R_{\boldsymbol{\omega}}(\varphi_{\hat{\boldsymbol{\omega}}})=(1-\omega_1)R_0(\varphi_{\hat{\boldsymbol{\omega}}})+\omega_1 R_1(\varphi_{\hat{\boldsymbol{\omega}}}) . \tag{4.6}$$

Then, $R_{\boldsymbol{\omega}}(\varphi_{\hat{\boldsymbol{\omega}}})$ represents a straight line segment of a support for $B(\boldsymbol{\omega})$ which is continuous and concave as functions of $\boldsymbol{\omega}'$, at the fixed value of $\hat{\boldsymbol{\omega}}'$.

If we put $\omega_1=\hat{\omega}_1\pm e(\Delta, n, \alpha/2)$ in (4.6), we have

$$\begin{aligned}&(1-\hat{\omega}_1-e(\Delta, n, \alpha/2))R_0(\varphi_{\hat{\boldsymbol{\omega}}})+(\hat{\omega}_1+e(\Delta, n, \alpha/2))R_1(\varphi_{\hat{\boldsymbol{\omega}}})\\ &=B(\hat{\boldsymbol{\omega}})+e(\Delta, n, \alpha/2)(R_1(\varphi_{\hat{\boldsymbol{\omega}}})-R_0(\varphi_{\hat{\boldsymbol{\omega}}}))=B^+\end{aligned}$$

and

$$\begin{aligned}&(1-\hat{\omega}_1+e(\Delta, n, \alpha/2))R_0(\varphi_{\hat{\boldsymbol{\omega}}})+(\hat{\omega}_1-e(\Delta, n, \alpha/2))R_1(\varphi_{\hat{\boldsymbol{\omega}}})\\ &=B(\hat{\boldsymbol{\omega}})+e(\Delta, n, \alpha/2)(R_0(\varphi_{\hat{\boldsymbol{\omega}}})-R_1(\varphi_{\hat{\boldsymbol{\omega}}}))=B^- .\end{aligned}$$

Put $\hat{\boldsymbol{\omega}}^+=(1-\hat{\omega}_1-e(\Delta, n, \alpha/2), \hat{\omega}_1+e(\Delta, n, \alpha/2))$ and $\hat{\boldsymbol{\omega}}^-=(1-\hat{\omega}_1+e(\Delta, n, \alpha/2), \hat{\omega}_1-e(\Delta, n, \alpha/2))$, and define

$$K(\hat{\boldsymbol{\omega}}, \Delta, n, \alpha/2)=\max\{K^+, K^-\} , \tag{4.7}$$

where $K^+=B^+-B(\hat{\boldsymbol{\omega}}^+)$ and $K^-=B^--B(\hat{\boldsymbol{\omega}}^-)$. Then, if $|\hat{\omega}_1-\omega_1|<e(\Delta, n, \alpha/2)$, we have

$$0\leqq R_{\boldsymbol{\omega}}(\varphi_{\hat{\boldsymbol{\omega}}})-B(\boldsymbol{\omega})\leqq K(\hat{\boldsymbol{\omega}}, \Delta, n, \alpha/2) , \tag{4.8}$$

since $B(\boldsymbol{\omega})$ is a concave function of $\boldsymbol{\omega}$. Hence, we obtain:

If we take $\hat{\boldsymbol{\omega}}'=(\hat{\omega}_0, \hat{\omega}_1)$ given by (2.8) *as an estimator for $\boldsymbol{\omega}'=(\omega_1, \omega_2)$, we have*

$$P\{R_{\boldsymbol{\omega}}(\varphi_{\hat{\boldsymbol{\omega}}})-B(\boldsymbol{\omega})\leqq K(\hat{\boldsymbol{\omega}}, \Delta, n, \alpha/2)\}\geqq 1-\alpha \tag{4.9}$$

for any α $(0<\alpha<1)$, where $K(\hat{\boldsymbol{\omega}}, \Delta, n, \alpha/2)$ is defined by (4.7).

We can also give an approximate evaluation of the left-hand side of (4.9) in the same way, by using the result such that $\hat{\omega}_1$ is asymptotically normally distributed. On the other hand, let $\varphi^*(x)$ be a minimax rule such that $R_0(\varphi^*)=R_1(\varphi^*)$. The rule $\varphi^*(x)$ is a Bayes rule with respect to some $\boldsymbol{\omega}^{*\prime}=(\omega_0^*, \omega_1^*)$, the least favorable prior distribution, in the sense that the Bayes risk takes on a maximum for this value $\boldsymbol{\omega}^{*\prime}$ of $\boldsymbol{\omega}'$. Then, if $|\hat{\omega}_1-\omega_1^*|<e(\Delta, n, \alpha/2)$, we should take rather φ^* than

$\varphi_{\boldsymbol{\omega}}$ for the benefit of practical convenience.

5. The case of two known multivariate normal distributions

The method proposed in the previous sections is applicable to the case of multivariate probability distributions if we can select a suitable transformation to a one-dimensional random variable from a multi-dimensional one, i.e. a linear function of components of a vector random variable.

We shall deal with the case of two p-variate normal distributions with equal covariance matrices, namely, $N(\boldsymbol{m}^{(0)}, \boldsymbol{\Sigma})$ and $N(\boldsymbol{m}^{(1)}, \boldsymbol{\Sigma})$, where $\boldsymbol{m}^{(i)\prime}=(m_1^{(i)}, \cdots, m_p^{(i)})$ is the vector of expected means of the population π_i $(i=0, 1)$, ($\boldsymbol{m}^{(i)\prime}$ denotes the transpose of $\boldsymbol{m}^{(i)}$,) and $\boldsymbol{\Sigma}$ is the common matrix of variances and covariances of both populations.

Denote by $\boldsymbol{X}^{(i)}$ a p-column vector random variable distributed according to $N(\boldsymbol{m}^{(i)}, \boldsymbol{\Sigma})$ for the sake of the descriptive simplicity, and consider a linear function $\boldsymbol{X}^{(i)\prime}\boldsymbol{a}$ of components of $\boldsymbol{X}^{(i)}$ $(i=0, 1)$. Then, we have

$$P\{\boldsymbol{X}^{(0)\prime}\boldsymbol{a}<\boldsymbol{X}^{(1)\prime}\boldsymbol{a}\}=\Phi\left(\frac{1}{\sqrt{2}}\frac{\boldsymbol{m}^{(1)\prime}\boldsymbol{a}-\boldsymbol{m}^{(0)\prime}\boldsymbol{a}}{\sqrt{\boldsymbol{a}'\boldsymbol{\Sigma}\boldsymbol{a}}}\right), \tag{5.1}$$

where $\Phi(c)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{c} e^{-t^2/2}dt$.

Since our method tends to estimate $\boldsymbol{\omega}'$ more efficiently for pairs of two distribution functions $F_0(x)$ and $F_1(x)$ when the absolute value of $\Delta_{ij}=\int_{-\infty}^{\infty} F_i(x)dF_j(x)-\frac{1}{2}$ $(i\neq j)$ becomes larger, if $\boldsymbol{m}^{(1)\prime}\boldsymbol{a}>\boldsymbol{m}^{(0)\prime}\boldsymbol{a}$ (or if $\boldsymbol{m}^{(1)\prime}\boldsymbol{a}<\boldsymbol{m}^{(0)\prime}\boldsymbol{a}$), $\boldsymbol{a}$ $(i=0, 1)$ maximizing (5.1) (or minimizing (5.1)) will provide a suitable linear function for our aim. It will be easily seen from the right-hand term in (5.1) that such a linear function is the discriminant function.

On the other hand, the log-likelihood-ratio is

$$\log\frac{f_1(\boldsymbol{x})}{f_0(\boldsymbol{x})}=\boldsymbol{x}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{m}^{(1)}-\boldsymbol{m}^{(0)})-\frac{1}{2}(\boldsymbol{m}^{(1)}+\boldsymbol{m}^{(0)})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{m}^{(1)}-\boldsymbol{m}^{(0)}), \tag{5.2}$$

where

$$f_i(\boldsymbol{x})=\frac{1}{(2\pi)^{p/2}|\boldsymbol{\Sigma}|^{1/2}}\exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{m}^{(i)})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{m}^{(i)})\right].$$

Let

$$U=\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{m}^{(1)}-\boldsymbol{m}^{(0)})-\frac{1}{2}(\boldsymbol{m}^{(1)}+\boldsymbol{m}^{(0)})\boldsymbol{\Sigma}^{-1}(\boldsymbol{m}^{(1)}-\boldsymbol{m}^{(0)}) \tag{5.3}$$

for the random variable $\boldsymbol{X}$ distributed with either $N(\boldsymbol{m}^{(0)}, \boldsymbol{\Sigma})$ or $N(\boldsymbol{m}^{(1)}, \boldsymbol{\Sigma})$, and consider the "Mahalanobis' distance" between $N(\boldsymbol{m}^{(0)}, \boldsymbol{\Sigma})$ and $N(\boldsymbol{m}^{(1)}, \boldsymbol{\Sigma})$:

$$(5.4) \qquad (\boldsymbol{m}^{(1)}-\boldsymbol{m}^{(0)})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{m}^{(1)}-\boldsymbol{m}^{(0)})=d\,.$$

Then, U is distributed according to $N\left(\frac{1}{2}d,\, d\right)$ if $\boldsymbol{X}$ is distributed according to $N(\boldsymbol{m}^{(1)}, \boldsymbol{\Sigma})$. If $\boldsymbol{X}$ is distributed according to $N(\boldsymbol{m}^{(0)}, \boldsymbol{\Sigma})$, U is distributed according to $N\left(-\frac{1}{2}d,\, d\right)$. Thus, when $\boldsymbol{m}^{(0)}$, $\boldsymbol{m}^{(1)}$ and $\boldsymbol{\Sigma}$ are given, if a random sample $(\boldsymbol{X}_1, \cdots, \boldsymbol{X}_n)$ drawn from $\omega_0 N(\boldsymbol{x}\,|\,\boldsymbol{m}^{(0)}, \boldsymbol{\Sigma})+\omega_1 N(\boldsymbol{x}\,|\,\boldsymbol{m}^{(1)}, \boldsymbol{\Sigma})$ is available, we can estimate $\boldsymbol{\omega}'=(\omega_0, \omega_1)$ based on $(U_1, \cdots, U_n)$ transformed from $(\boldsymbol{X}_1, \cdots, \boldsymbol{X}_n)$ by (5.3), since $\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{m}^{(1)}-\boldsymbol{m}^{(0)})$ is the well-known discriminant function, (where $N(\boldsymbol{x}\,|\,\boldsymbol{m}^{(i)}, \boldsymbol{\Sigma})$ is the distribution function of the p-variate normal distribution with the expected mean $\boldsymbol{m}^{(i)}$ and the covariance matrix $\boldsymbol{\Sigma}$ $(i=0, 1)$). In this case, Δ defined by (2.3) becomes

$$(5.5) \qquad \Delta=\Delta_{01}=\Phi\left(\sqrt{\frac{d}{2}}\right)-\frac{1}{2}\,,$$

and the rule $\varphi_{\hat{\boldsymbol{\omega}}}$ is:

$$(5.6) \qquad \varphi_{\hat{\boldsymbol{\omega}}}(\boldsymbol{x})=\begin{cases} 0 & \text{if } \boldsymbol{x}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{m}^{(1)}-\boldsymbol{m}^{(0)}) \\ & \qquad -\frac{1}{2}(\boldsymbol{m}^{(1)}+\boldsymbol{m}^{(0)})\boldsymbol{\Sigma}^{-1}(\boldsymbol{m}^{(1)}-\boldsymbol{m}^{(0)})\leqq \log k\,, \\ 1 & \text{if } \boldsymbol{x}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{m}^{(1)}-\boldsymbol{m}^{(0)}) \\ & \qquad -\frac{1}{2}(\boldsymbol{m}^{(1)}+\boldsymbol{m}^{(0)})\boldsymbol{\Sigma}^{-1}(\boldsymbol{m}^{(1)}-\boldsymbol{m}^{(0)})> \log k\,, \end{cases}$$

where k is given by

$$k=\frac{\hat{\omega}_1 L(0\,|\,1)}{\hat{\omega}_0 L(1\,|\,0)}\,,$$

and $\hat{\boldsymbol{\omega}}'=(\hat{\omega}_0, \hat{\omega}_1)$ is an unbiased estimator for $\boldsymbol{\omega}'$.

6. Estimating the prior distribution $\boldsymbol{\omega}'$

Let the random variable X be known to have one of a finite number of specified distribution functions $F_1(x), \cdots, F_r(x)$ depending on the respective values of a random parameter θ which has an unknown prior distribution $\boldsymbol{\omega}'=(\omega_1, \cdots, \omega_r)$ such that $P(\theta=i)=\omega_i$ for $i=1, \cdots, r$. We shall be concerned with estimating the unknown prior distribution $\boldsymbol{\omega}'=$

$(\omega_1, \cdots, \omega_r)$. When a random sample $(X_1, \cdots, X_r)$ drawn from the distribution function

$$F(x)=\sum_{i=1}^{r} \omega_i F_i(x) \tag{6.1}$$

is available, the problem can be extended in the similar way as proposed in section 2.

Consider r statistics

$$\hat{p}_i=\frac{1}{n}\sum_{k=1}^{n} F_i(x_k) \qquad (i=1, \cdots, r)\,. \tag{6.2}$$

Then, their expectations with respect to $F(x)$ are

$$\mathcal{E}\{\hat{p}_i\} = \sum_{j=1}^{r} \omega_j \int_{-\infty}^{\infty} F_i(x)dF_j(x) \qquad (i=1, \cdots, r)\,, \tag{6.3}$$

respectively.

We shall define $\varDelta_{ij}$ by

$$\varDelta_{ij}=\int_{-\infty}^{\infty} F_i(x)dF_j(x)-\frac{1}{2}\,. \tag{6.4}$$

It will easily seen that $\varDelta_{ij}=-\varDelta_{ji}$. If the prior distribution that must be estimated and r statistics defined in (6.2) are denoted in vector forms by

$$\boldsymbol{\omega}=\begin{pmatrix} \omega_1 \\ \vdots \\ \omega_r \end{pmatrix} \quad \text{and} \quad \hat{\boldsymbol{p}}=\begin{pmatrix} \hat{p}_1 \\ \vdots \\ \hat{p}_r \end{pmatrix}, \tag{6.5}$$

and if matrix A is defined by

$$A=\begin{pmatrix} 1/2 & 1/2+\varDelta_{12} & \cdots & 1/2+\varDelta_{1r} \\ 1/2+\varDelta_{21} & 1/2 & \cdots & 1/2+\varDelta_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ 1/2+\varDelta_{r1} & 1/2+\varDelta_{r2} & \cdots & 1/2 \end{pmatrix}, \tag{6.6}$$

(6.3) is written as follows

$$\mathcal{E}\{\hat{\boldsymbol{p}}\} = A\boldsymbol{\omega}\,. \tag{6.7}$$

Hence, we obtain:

If we have $|A|\neq 0$, unbiased estimators for components of $\boldsymbol{\omega}$ are obtained from

$$\hat{\boldsymbol{\omega}}=A^{-1}\hat{\boldsymbol{p}}\,, \tag{6.8}$$

where $|A|$ *denotes the determinant of* A.

In most of the usual cases, it can be expected that $|A|\neq 0$, but we shall show the special cases of $r=3$ and 4 since it may be rather complicated to seek the conditions that have to be laid on the distributions for $|A|\neq 0$.

In the case of $r=3$,

$$|A|=\frac{1}{2}(\Delta_{12}+\Delta_{23}+\Delta_{31})^2 .$$

Thus, if we have

$$\Delta_{12}+\Delta_{23}+\Delta_{31}=\int(F_1(x)-F_2(x))dF_2(x)-\int(F_1(x)-F_2(x))dF_3(x) \neq 0 , \tag{6.9}$$

unbiased estimators for components of $\boldsymbol{\omega}'=(\omega_1, \omega_2, \omega_3)$ are given by

$$\hat{\omega}_1=\frac{1}{\Delta}\{(\hat{p}_3-\hat{p}_2)+\Delta_{23}\} ,$$

(6.10)

$$\hat{\omega}_2=\frac{1}{\Delta}\{(\hat{p}_1-\hat{p}_3)+\Delta_{31}\} ,$$

and

$$\hat{\omega}_3=\frac{1}{\Delta}\{(\hat{p}_2-\hat{p}_1)+\Delta_{12}\} ,$$

where $\Delta=\Delta_{12}+\Delta_{23}+\Delta_{31}$. In fact, we have

$$\mathcal{E}\{\hat{p}_3-\hat{p}_2\}=-\Delta_{23}+\omega_1\Delta ,$$

$$\mathcal{E}\{\hat{p}_1-\hat{p}_3\}=-\Delta_{31}+\omega_2\Delta ,$$

and

$$\mathcal{E}\{\hat{p}_2-\hat{p}_1\}=-\Delta_{12}+\omega_3\Delta ,$$

where $\mathcal{E}\{\ \}$ denotes expectation with respect to $F(x)$, so, their unbiasedness is quite clear.

In the case of $r=4$, we have

$$|A|=(\Delta_{12}\Delta_{43}+\Delta_{13}\Delta_{24}+\Delta_{23}\Delta_{41})^2 .$$

Thus, if $\Delta_{12}\Delta_{43}+\Delta_{13}\Delta_{24}+\Delta_{23}\Delta_{41}\neq 0$, unbiased estimators for components of $\boldsymbol{\omega}'=(\omega_1, \omega_2, \omega_3, \omega_4)$ can be obtained in the same way.

7. The empirical Bayes r-way decision procedure

Suppose that populations $\pi_1, \pi_2, \cdots, \pi_r$ are given and observations from π_θ $(\theta=1, 2, \cdots, r)$ have the distribution $F_\theta(x)$. Let X be a random variable which is known to have one of r distinct distribution functions $F_1(x), F_2(x), \cdots$ and $F_r(x)$, say $F_\theta(x)$.

We shall consider the same problem as discussed in section 4 for the case of r distributions. When the true parameter θ is i, if we make a decision $\theta=j$, assume that we incur the loss $L(j\,|\,i)(\geqq 0)$ $(i, j=1, \cdots, r)$. A randomized decision rule $\varphi(x)=(\varphi^{(1)}(x), \cdots, \varphi^{(r)}(x))$ is a vector of measurable functions such that $\sum_{j=1}^{r}\varphi^{(j)}(x)=1$, $\varphi^{(j)}(x)\geqq 0$ and $\varphi^{(j)}(x)$ is the probability of deciding X as coming from π_j when we observe $X=x$. When θ is the parameter, the risk function for any φ is

$$R_\theta(\varphi)=\mathcal{E}_\theta\left\{\sum_{j=1}^{r} L(j\,|\,\theta)\varphi^{(j)}(X)\right\}, \tag{7.1}$$

where $\mathcal{E}_\theta$ denotes expectation with respect to $F_\theta(x)$, and when the prior distribution of θ is $\boldsymbol{\omega}'=(\omega_1, \cdots, \omega_r)$, the expected risk with respect to $\boldsymbol{\omega}'$ is

$$R_{\boldsymbol{\omega}}(\varphi)=\sum_{i=1}^{r}\omega_i R_i(\varphi). \tag{7.2}$$

Let $f_\theta(x)$ denote the likelihood function of θ given x. Then, for given $\boldsymbol{\omega}'$, (7.2) is minimized by $\varphi_{\boldsymbol{\omega}}(x)=(\varphi_{\boldsymbol{\omega}}^{(1)}(x), \cdots, \varphi_{\boldsymbol{\omega}}^{(r)}(x))$ of the form

$$\varphi_{\boldsymbol{\omega}}^{(j)}(x)=\begin{cases} 0 & \text{if } \sum_{\theta=1}^{r} L(j\,|\,\theta)\omega_\theta f(x\,|\,\theta)\geqq \min_{u\neq j}\sum_{\theta=1}^{r} L(u\,|\,\theta)\omega_\theta f(x\,|\,\theta), \\ 1 & \text{if } \sum_{\theta=1}^{r} L(j\,|\,\theta)\omega_\theta f(x\,|\,\theta)< \min_{u\neq j}\sum_{\theta=1}^{r} L(u\,|\,\theta)\omega_{\hat\theta} f_\theta(x\,|\,\theta) \end{cases} \quad (j, u=1, \cdots, r). \tag{7.3}$$

The rule $\varphi_{\boldsymbol{\omega}}(x)$ is a Bayes rule with respect to $\boldsymbol{\omega}'$ and the corresponding Bayes risk is

$$B(\boldsymbol{\omega})=R_{\boldsymbol{\omega}}(\varphi_{\boldsymbol{\omega}})=\min_{\varphi} R_{\boldsymbol{\omega}}(\varphi). \tag{7.4}$$

The following lemma due to G. Suzuki [16] will be of avail in the next place.

Let $\varphi_{\boldsymbol{\eta}}(x)$ be the rule of the form (7.3) *with $\boldsymbol{\omega}'=(\omega_1, \cdots, \omega_r)$ replaced by any estimator $\boldsymbol{\eta}'=(\eta_1, \cdots, \eta_r)$ for $\boldsymbol{\omega}'$. Then we have*

$$0\leqq R_{\boldsymbol{\omega}}(\varphi_{\boldsymbol{\eta}})-B(\boldsymbol{\omega})\leqq \sum_{i=1}^{r}|\,\omega_i-\eta_i\,|\,L_i, \tag{7.5}$$

where $L_i=\max_j\{L(j\,|\,i)\}$ $(i, j=1, \cdots, r)$ and $B(\boldsymbol{\omega})$ is defined by (7.4).

We shall give a probability inequality for the relation between $R_{\boldsymbol{\omega}}(\varphi_{\hat{\boldsymbol{\omega}}})$

and $B(\boldsymbol{\omega})$ as mentioned in the lemma for the case of $r=3$, where $\varphi_{\hat{\boldsymbol{\omega}}}$ denotes the rule of form (7.3) with the estimator $\hat{\boldsymbol{\omega}}'=(\hat{\omega}_1, \hat{\omega}_2, \hat{\omega}_3)$ given by (6.10) replacing $\boldsymbol{\omega}'=(\omega_1, \omega_2, \omega_3)$. In the case of $r=3$, (7.5) is written as

$$0\leqq R_{\boldsymbol{\omega}}(\varphi_{\hat{\boldsymbol{\omega}}})-B(\boldsymbol{\omega})\leqq\sum_{i=1}^{3}|\hat{\omega}_i-\omega_i|L_i$$
$$=\sum_{i\neq\lambda}|\hat{\omega}_i-\omega_i|(L_i+L_\lambda),$$

since $(\hat{\omega}_\nu-\omega_\nu)+(\hat{\omega}_\mu-\omega_\mu)=(\omega_\lambda-\hat{\omega}_\lambda)$, where $L_\lambda=\min_i\{L_i\}$. Thus, we have

$$P\{R_{\boldsymbol{\omega}}(\varphi_{\hat{\boldsymbol{\omega}}})-B(\boldsymbol{\omega})\geqq\xi\}\leqq P\{\sum_{i\neq\lambda}|\hat{\omega}_i-\omega_i|(L_i+L_\lambda)\geqq\xi\}$$
$$\leqq\sum_{i\neq\lambda}P\{|\hat{\omega}_i-\omega_i|\geqq\eta/2\}$$

for any $\xi>0$, where $\eta=\min_{\substack{i\\ i\neq\lambda}}\left\{\frac{\xi}{L_i+L_\lambda}\right\}$.

On the other hand, if we make use of Hoeffding's probability inequality, we have

$$P\{(\hat{p}_\nu-\hat{p}_\mu)-(\omega_\lambda\Delta-\Delta_{\mu\nu})\geqq t\}\leqq e^{-nt^2/2} \qquad (\nu\neq\mu\neq\lambda;\ \nu, \mu, \lambda=1, 2, 3),$$

since $(F_\nu(X_1)-F_\mu(X_1)), \cdots,$ and $(F_\nu(X_n)-F_\mu(X_n))$ are independent, and $-1\leqq(F_\nu(X_k)-F_\mu(X_k))\leqq1$ for $k=1, \cdots, n$. We shall, without loss of generality, assume that $\Delta>0$ since Δ can be made to take on a positive value by interchanging subscripts ν, μ of $\Delta_{\nu\mu}$ for ν, $\mu=1, 2, 3$ unless $\Delta=0$. Then, we have

$$P\{\hat{\omega}_\lambda-\omega_\lambda\geqq t/\Delta\}\leqq e^{-nt^2/2}.$$

Hence, we have

$$P\{|\hat{\omega}_\lambda-\omega_\lambda|\geqq t/\Delta\}\leqq 2e^{-nt^2/2}$$

for any $t>0$. Thus, for the case of $r=3$ we can state:

When we take the rule $\varphi_{\hat{\boldsymbol{\omega}}}$ of the form (7.3) *with $\hat{\boldsymbol{\omega}}'$, replacing $\boldsymbol{\omega}'$ if $\Delta>0$, we have*

$$P\{R_{\boldsymbol{\omega}}(\varphi_{\hat{\boldsymbol{\omega}}})-B(\boldsymbol{\omega})<\xi\}\geqq1-4\exp\left\{-\frac{n}{6}\left(\frac{\xi\Delta}{L}\right)^2\right\}$$

where $L=\max_{i\neq\lambda}\{L_i+L_\lambda\}$, and $L_\lambda=\min_i\{L_i\}$ for $i=1, 2, 3$.

8. Acknowledgement

The author wishes to thank Dr. K. Matusita for his guidance. The author is also indebted to Dr. Y. Suzuki and Mr. G. Suzuki for valuable discussions. His sincere thanks are extended to Mrs. S. Kazi for her assistance in preparing the manuscript.

THE INSTITUTE OF STATISTICAL MATHEMATICS

REFERENCES

[1] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis,* John Wiley and Sons, New York, 1958.

[2] D. Blackwell and M. A. Girshick, *Theory of Games and Statistical Decisions,* John Wiley and Sons, New York, 1954.

[3] J. F. Hannan and H. Robbins, "Asymptotic solutions of the compound decision problem for two completely specified distributions," *Ann. Math. Statist.,* 26 (1955), 37-51.

[4] J. F. Hannan and J. R. Van Ryzin, "Rate of convergence in the compound decision problem for two completely specified distributions," *Ann. Math. Statist.,* 36 (1965), 1743-1752.

[5] W. Hoeffding, "Probability inequalities for sums of bounded random variables," *J. Amer. Statist. Ass.,* 58 (1963), 13-30.

[6] H. Hudimoto, "On a distribution-free two-way classification," *Ann. Inst. Statist. Math.,* 16 (1964), 247-253.

[7] H. Hudimoto, "On Bayesian inference," *Journal of the Japan Association for Philosophy of Science,* 8 (1966), 80-87.

[8] M. V. Johns, "An empirical Bayes approach to non-parametric two-way classification," *Studies in Item Analysis and Prediction,* Stanford Univ. Press, Stanford, California, (1961), 221-232.

[9] J. Neyman, "Two breakthroughs in the theory of statistical decision making," *Rev. Int. Statist. Inst.,* 30, 1 (1962), 11-27.

[10] H. Robbins, "Asymptotically subminimax solutions of compound statistical decision problems," *Proc. Second Berkeley Symp. on Math. Statist. and Prob.,* Univ. of California Press, New York, (1950), 131-148.

[11] H. Robbins, "An empirical Bayes approach to statistics," *Proc. Third Berkeley Symp. on Math. Statist. and Prob.,* Univ. of California Press, New York, (1955), 157-164.

[12] H. Robbins, "The empirical Bayes approach to statistical decision problems," *Ann. Math. Statist.,* 35 (1964), 1-20.

[13] E. J. G. Pitman, "Some remarks on statistical inference," *Bernoulli* $_{1713}$, *Bayes* $_{1763}$, *Laplace* $_{1813}$, Anniversary volume, Proceeding of an International Research Seminar, Statistical Laboratory, University of California, Berkeley Springer-Verlag, Berlin, Heidelberg, New York (1965), 209-216.

[14] E. Samuel, "Asymptotic solutions of the sequential compound decision problem," *Ann. Math. Statist.,* 34 (1963), 1079-1094.

[15] E. Samuel, "Convergence of the losses of certain decision rule for the sequential compound decision problem," *Ann. Math. Statist.,* 35 (1964), 1606-1621.

[16] G. Suzuki, "Asymptotic solutions of the compound decision problem for many completely specified distributions," *Research Memo.,* Inst. Statist. Math., No. 1 (1966).

[17] G. Suzuki, "Discrete compound decision problem," *Ann. Inst. Statist. Math.,* 18 (1966), 127-139.

CORRECTIONS TO
"ON THE EMPIRICAL BAYES PROCEDURE (1)"

HIROSI HUDIMOTO

The variances of $\widehat{\boldsymbol{\omega}}_1$ and $\hat{\omega}_1$ in the theorem on page 176 of the above paper (Ann. Inst. Statist. Math., 20 (1968), 169–185) which states that $\widehat{\boldsymbol{\omega}}_1$ and $\hat{\omega}_1$ are asymptotically normally distributed with the expected mean ω_1, should be read as

$$\mathrm{Var}(\widehat{\boldsymbol{\omega}}_1)=\frac{1}{12n\Delta^2}+\frac{1}{n\Delta^2}\left\{\omega_1\left(\int_{-\infty}^{\infty}F_0(x)dF_1(x)-\frac{1}{3}-\Delta\right)-\omega_1^2\Delta\right\},$$

and

$$\mathrm{Var}(\hat{\omega}_1)=\frac{1}{12n\Delta^2}+\frac{1}{4n\Delta^2}\left\{(2\omega_1-1)\left(\int_{-\infty}^{\infty}F_0^2(x)dF_1(x)-\int_{-\infty}^{\infty}F_1^2(x)dF_0(x)-2\Delta\right)-\Delta^2(2\omega_1-1)^2\right\},$$

respectively.
The second terms in the above expressions were left out.

The author wishes to thank Professor John W. Pratt, Harvard University, for pointing out these errors.