DISCRETE COMPOUND DECISION PROBLEM

GIITIRO SUZUKI

(Received June 20, 1966)

1. Introduction and summary

The compound decision problem, called by Robbins [7], consists in the simultaneous consideration of n decision problems which have an identical formal structure. All the papers [1]~[11] except [8] treated the continuous compound decision problem. Robbins [8] considered the decision problem between the binomial distributions and the distribution which is given by a unit step function. The problem of convergence rate was treated by Hannan-Van Ryzin [4], Van Ryzin [13] and Samuel [10] in the continuous case, by Johns [6] in the two-action case and by Robbins [8] in the binomial case.

In this paper, we shall treat the compound decision problem between $k+1$ (at most) m-variate multinomial distributions and present the construction of the "best" simple symmetric solution and the exact convergence rate of the asymptotically "good" non-simple solution, using the different approach from that for the continuous case. This is an extension of [8] and a modification of [1], [2], [5], [11] and [13] which treated the continuous compound decision problem involving multiple component decisions.

We do not consider the sequential case (cf. [4], [5], [9], [10] and [11]). But, in that case, a similar argument can be made by modifying only the part of the estimation of the true parameter vector.

2. Simple symmetric decision function for the compound decision problem

2.1. Multinomial compound decision problem. Suppose that we are confronted with the following situation. There are $(k+1)$ states of nature, $\{0, 1, \cdots, k\}$. We represent the state by the symbol θ, and the set $\{0, 1, \cdots, k\}$ by K. We do not know the true value of θ, but we can observe a random variable X with the value range $M=\{0, 1, \cdots, m\}$ $(m \geqq k)$, for which

$$(1) \qquad \Pr\{X=i \mid \theta=j\}=p(i, j), \qquad i \in M,\ j \in K,$$

where $(m+1)\times(k+1)$ matrix $\boldsymbol{P}=(p(i, j))$ is a known parameter matrix satisfying the following conditions:

$$(2) \qquad p(i, j)\geqq 0, \qquad \sum_{i=0}^{m} p(i, j)=1,$$

$$(3) \qquad \boldsymbol{P}'\boldsymbol{P} \text{ is non-singular.}$$

Suppose that when $\theta=j$ and we guess that $\theta=i$, we lose $a(i, j)$ with

$$(4) \qquad 0\leqq a(i, j)\leqq A(<\infty), \qquad \text{for each } i, j \in K.$$

Let $\mathcal{D}$ be the class of all non-randomized decision functions:

$$\mathcal{D}=\{\boldsymbol{d}=(d_0, d_1, \cdots, d_m);\ d_i \in K\ (i=0, 1, \cdots, m)\}.$$

When $\theta=j \in K$, the loss suffered by using $\boldsymbol{d}=(d_0, d_1, \cdots, d_m) \in \mathcal{D}$ (i.e. estimating $\theta=d_i$ according to the observation i of X), is given by

$$(5) \qquad l_j(\boldsymbol{d})=\sum_{i=0}^{m} a(d_i, j)p(i, j), \qquad j \in K.$$

Now, we shall consider the following discrete compound decision problem. Let n be a fixed positive integer and let $X_n=(X_1, \cdots, X_n)$ be the vector of independent random variables, where for each X_i we have

$$\Pr\{X_i=j\}=p(j, \theta_i), \qquad i=1, 2, \cdots, n$$

and the unknown parameter vector $\boldsymbol{\theta}_n=(\theta_1, \cdots, \theta_n)$ belongs to the set $K^n=\{\boldsymbol{\theta}_n=(\theta_1, \cdots, \theta_n);\ \theta_i \in K\ (i=1, \cdots, n)\}$. The problem is to find a $\boldsymbol{d} \in \mathcal{D}$ that minimizes the average loss

$$(6) \qquad l_n(\boldsymbol{d}\,; \boldsymbol{\theta}_n)=\frac{1}{n}\sum_{i=1}^{n} l_{\theta_i}(\boldsymbol{d})=\sum_{j=0}^{k} \eta_j(\boldsymbol{\theta}_n)l_j(\boldsymbol{d}),$$

where

$$(7) \qquad \eta_j(\boldsymbol{\theta}_n)=\frac{1}{n}\sum_{i=1}^{n} u(\theta_i, j), \qquad j=0, 1, \cdots, k,$$

$$u(i, j)=\delta_{i,j}=\begin{cases} 1, & \text{if } i=j, \\ 0, & \text{otherwise.}\end{cases}$$

The decision function of this type is called a simple symmetric (abbr. by s.s.) solution for the compound decision problem (cf. [3], [11]).

2.2. Construction of the best s.s. solution. We shall consider the vector-valued function on $\mathcal{D}$

$$(8) \qquad \boldsymbol{r}(\boldsymbol{d})=(r_0(\boldsymbol{d}), r_1(\boldsymbol{d}), \cdots, r_m(\boldsymbol{d}))=\boldsymbol{l}(\boldsymbol{d})\boldsymbol{Q},$$

where

$$\boldsymbol{l}(\boldsymbol{d})=(l_0(\boldsymbol{d}), l_1(\boldsymbol{d}), \cdots, l_k(\boldsymbol{d})),$$

each element of which is given by (5) and $\boldsymbol{Q}$ is the generalized inverse of the matrix $\boldsymbol{P}$. By the assumption (3) $\boldsymbol{Q}$ is uniquely determined by

$$(9) \qquad \boldsymbol{Q}=(\boldsymbol{P}'\boldsymbol{P})^{-1}\boldsymbol{P}'.$$

Let S_m be the m-dimensional simplex

$$S_m=\{\boldsymbol{s}=(s_0, s_1, \cdots, s_{m-1})\,;\, s_i\geqq 0,\ \sum_{i=0}^{m-1} s_i\leqq 1\}.$$

For any $\boldsymbol{s}\in S_m$, we put

$$(10) \qquad r(\boldsymbol{d}\,;\boldsymbol{s})=\sum_{i=0}^{m-1} s_i r_i(\boldsymbol{d})+(1-s_0-\cdots-s_{m-1})r_m(\boldsymbol{d}),$$

$$(11) \qquad r(\boldsymbol{s})=\min_{\boldsymbol{d}\in\mathcal{D}} r(\boldsymbol{d}\,;\boldsymbol{s}),$$

$$(12) \qquad \langle \boldsymbol{d}\rangle=\{\boldsymbol{s}\in S_m\,;\, r(\boldsymbol{d}\,;\boldsymbol{s})\leqq r(\tilde{\boldsymbol{d}}\,;\boldsymbol{s}),\ \text{for all } \tilde{\boldsymbol{d}}\in\mathcal{D}\}.$$

Then it is easily seen from (6), (7), (10) and (11) that

$$(13) \qquad \boldsymbol{s}^0=\boldsymbol{s}(\boldsymbol{\theta}_n)=\boldsymbol{\eta}(\boldsymbol{\theta}_n)\boldsymbol{P}'\in S_m$$

and

$$(14) \qquad l_n(\boldsymbol{d}\,;\boldsymbol{\theta}_n)=r(\boldsymbol{d}\,;\boldsymbol{s}^0)$$

$$(15) \qquad l_n(\boldsymbol{\theta}_n)=\min_{\boldsymbol{d}\in\mathcal{D}} l_n(\boldsymbol{d}\,;\boldsymbol{\theta}_n)=r(\boldsymbol{s}^0).$$

THEOREM 1. *Set*

$$(16) \qquad \langle j\,;i\rangle=\{\boldsymbol{s}\in S_m\,;\, w_i(j\,;\boldsymbol{s})\leqq w_i(h\,;\boldsymbol{s}),\ \textit{for all } h\in K\},$$

where

$$(17) \qquad w_i(j\,;\boldsymbol{s})=\sum_{h=0}^{k} a(j, h)p(i, h)q(h\,;\boldsymbol{s}), \qquad i\in M,\ j\in K,$$

$$(18) \qquad q(h\,;\boldsymbol{s})=q(h, m)-\sum_{i=0}^{m-1} s_i[q(h, m)-q(h, i)], \qquad h\in K,$$

and $q(i, j)$ *is the* (i, j) *element of the matrix* $\boldsymbol{Q}$ *given by* (9). *Then, for every* $\boldsymbol{d}\in\mathcal{D}_0=\{\boldsymbol{d}\in\mathcal{D}\,;\,\langle\boldsymbol{d}\rangle\neq\phi\}$, *we have*

$$(19) \qquad \langle\boldsymbol{d}\rangle=\langle d_0, \cdots, d_m\rangle=\bigcap_{i=0}^{m}\langle d_i\,;i\rangle.$$

Using the above proposition, we can construct the subset $\langle \boldsymbol{d} \rangle$ of S_m for every $\boldsymbol{d} \in \mathcal{D}_0$. Therefore, if we know the true value of $\boldsymbol{s}^0 = \boldsymbol{s}(\boldsymbol{\theta}_n)$ defined by (13), we can find such a $\boldsymbol{d}^0 = \boldsymbol{d}(\boldsymbol{\theta}_n) \in \mathcal{D}_0$ that $s^0 \in \langle \boldsymbol{d}^0 \rangle$. This $\boldsymbol{d}$ gives the minimum attainable risk.

PROOF. Since

$$r(\boldsymbol{d}\,; \boldsymbol{s}) = r(d_0, d_1, \cdots, d_m\,; \boldsymbol{s}) = \sum_{i=0}^{m} w_i(d_i\,; \boldsymbol{s}) \ ,$$

by the definitions (15) and (12) the relation

$$\langle \boldsymbol{d} \rangle \supset \bigcap_{i=0}^{m} \langle d_i\,; i \rangle$$

is evident. Therefore, we shall show the inverse inclusion. Let $\boldsymbol{s} \in \langle \boldsymbol{d} \rangle = \langle d_0, d_1, \cdots, d_m \rangle$. Then, for any $\tilde{\boldsymbol{d}} = (\tilde{d}_0, \tilde{d}_1, \cdots, \tilde{d}_m)$, we have

$$r(\boldsymbol{d}\,; \boldsymbol{s}) \leqq r(\tilde{\boldsymbol{d}}\,; \boldsymbol{s})$$

or equivalently,

$$\sum_{i=0}^{m} w_i(d_i\,; \boldsymbol{s}) \leqq \sum_{i=0}^{m} w_i(\tilde{d}_i\,; \boldsymbol{s}) \,.$$

When we put $\tilde{d}_i = d_i$ $(i = 1, 2, \cdots, m)$ especially, we have

$$w_0(d_0\,; \boldsymbol{s}) \leqq w_0(\tilde{d}_0\,; \boldsymbol{s}), \qquad \text{for any } \tilde{d}_0 \in K \,.$$

This shows that $\boldsymbol{s} \in \langle d_0\,; 0 \rangle$. In the same way, we have $\boldsymbol{s} \in \langle d_i\,; i \rangle$, $i = 1, 2, \cdots, m$.

2.3. Approximate s.s. solution. We shall show the following:

THEOREM 2. *For any* $\boldsymbol{\theta}_n \in K^n$ *and any* $\boldsymbol{s} \in S_m$, *we have*

$$l_n(\boldsymbol{d}\,; \boldsymbol{\theta}_n) - l_n(\boldsymbol{\theta}_n) \leqq B \,\| \boldsymbol{s} - \boldsymbol{s}(\boldsymbol{\theta}_n) \| \,, \tag{20}$$

where $\boldsymbol{d} \in \mathcal{D}_0$ *is such that* $\boldsymbol{s} \in \langle \boldsymbol{d} \rangle$, $\langle \boldsymbol{d} \rangle$ *being defined by* (12) *and*

$$B = 2A \max_{0 \leq j \leq m} \sum_{i=0}^{k} | q(i, m) - q(i, j) | \tag{21}$$

and the norm $\|\cdot\|$ *on* S_m *is given by*

$$\| \boldsymbol{s} - \tilde{\boldsymbol{s}} \| = \sum_{i=0}^{m-1} | s_i - \tilde{s}_i | \,. \tag{22}$$

From this theorem, when we know the approximate value $\tilde{\boldsymbol{s}}$ of $\boldsymbol{s}^0$

and we use a decision function $\tilde{\boldsymbol{d}}$ such that $\tilde{\boldsymbol{s}} \in \langle \tilde{\boldsymbol{d}} \rangle$, the risk $l_n(\tilde{\boldsymbol{d}}, \boldsymbol{\theta}_n)$ we suffer is not larger by more than $B \| \boldsymbol{s}^0 - \tilde{\boldsymbol{s}} \|$ than the minimum average risk $l_n(\boldsymbol{\theta}_n) = r(\boldsymbol{s}^0)$.

PROOF. By the definitions (10) (11) (12), it is easily shown that for any $\boldsymbol{s} \in S_m$ there exists at least one $\boldsymbol{d}^* = \boldsymbol{d}(\boldsymbol{s}) \in \mathcal{D}_0$ such that $\boldsymbol{s} \in \langle \boldsymbol{d}^* \rangle$ and $r(\boldsymbol{s}) = r(\boldsymbol{d}^* ; \boldsymbol{s})$. Furthermore, for any $\boldsymbol{s}, \tilde{\boldsymbol{s}} \in S_m$ there exists $\boldsymbol{d}, \tilde{\boldsymbol{d}} \in \mathcal{D}_0$ such that $\boldsymbol{s} \in \langle \boldsymbol{d} \rangle$ and $\tilde{\boldsymbol{s}} \in \langle \tilde{\boldsymbol{d}} \rangle$. If $\boldsymbol{d} = \tilde{\boldsymbol{d}}$, from (14), it is evident that the relation (20) holds. When $\boldsymbol{d} \neq \tilde{\boldsymbol{d}}$ there exist a finite number of $\boldsymbol{d}_1, \boldsymbol{d}_2, \cdots, \boldsymbol{d}_{l-1} \in \mathcal{D}_0$ such that

$$\boldsymbol{s}^i = (1-\alpha_i)\boldsymbol{s} + \alpha_i \tilde{\boldsymbol{s}} \in \langle \boldsymbol{d}_{i-1} \rangle \cap \langle \boldsymbol{d}_i \rangle, \qquad i = 1, \cdots, l,$$

where $0 = \alpha_0 < \alpha_1 < \cdots \alpha_l < \alpha_{l+1} = 1$ and $\boldsymbol{d}_0 = \boldsymbol{d}, \boldsymbol{d}_l = \tilde{\boldsymbol{d}}$. Since

$$r(\boldsymbol{d}^1 ; \boldsymbol{s}^*) = r(\boldsymbol{d}^2 ; \boldsymbol{s}^*)$$

for any $\boldsymbol{s}^* \in \langle \boldsymbol{d}^1 \rangle \cap \langle \boldsymbol{d}^2 \rangle$, we have

$$\begin{aligned} r(\tilde{\boldsymbol{d}} ; \boldsymbol{s}) - r(\boldsymbol{s}) &= r(\tilde{\boldsymbol{d}} ; \boldsymbol{s}) - r(\tilde{\boldsymbol{d}} ; \tilde{\boldsymbol{s}}) + r(\tilde{\boldsymbol{d}} ; \tilde{\boldsymbol{s}}) - r(\boldsymbol{d} ; \boldsymbol{s}) \\ &= r(\tilde{\boldsymbol{d}} ; \boldsymbol{s}) - r(\tilde{\boldsymbol{d}} ; \tilde{\boldsymbol{s}}) + \sum_{i=1}^{l+1} [r(\boldsymbol{d}_i ; \boldsymbol{s}^i) - r(\boldsymbol{d}_{i-1} ; \boldsymbol{s}^{i-1})] \\ &= r(\tilde{\boldsymbol{d}} ; \boldsymbol{s}) - r(\tilde{\boldsymbol{d}} ; \tilde{\boldsymbol{s}}) + \sum_{i=1}^{l+1} [r(\boldsymbol{d}_{i-1} ; \boldsymbol{s}^i) - r(\boldsymbol{d}_{i-1} ; \boldsymbol{s}^{i-1})] \\ &= \sum_{j=0}^{m-1} (\tilde{s}_i - s_i)[r_m(\tilde{\boldsymbol{d}}) - r_j(\tilde{\boldsymbol{d}})] + \sum_{j=0}^{m-1} \sum_{i=1}^{l+1} (s_j^{i-1} - s_j^i)[r_m(d_{i-1}) - r_j(d_{i-1})] . \end{aligned}$$

Noticing (4), (5), (8), (21) and (13)~(15), we get the inequality (20).

2.4. Example. Let

$$\text{(23)} \qquad \boldsymbol{P} = \begin{bmatrix} 1-p & p & 0 \\ p & 1-2p & p \\ 0 & p & 1-2p \end{bmatrix} \qquad \left(0 < p < \frac{1}{2},\ p \neq \frac{1}{3}\right),$$

$$a(i, j) = 1 - \delta_{ij} = \begin{cases} 0 & i = j, \\ 1 & i \neq j. \end{cases}$$

Then (9) becomes

$$\boldsymbol{Q} = \boldsymbol{P}^{-1} = \frac{1}{(1-p)(1-3p)} \begin{bmatrix} 1-3p+p^2 & -p(1-p) & p^2 \\ -p(1-p) & (1-p)^2 & -p(1-p) \\ p^2 & -p(1-p) & 1-3p+p^2 \end{bmatrix}.$$

Therefore, for $q(h ; \boldsymbol{s})$ defined by (18) we have

$$q(0\,;\boldsymbol{s})=\frac{1}{(1-p)(1-3p)}[p^2+(1-3p)s_0-ps_1]\,,$$

$$q(1\,;\boldsymbol{s})=\frac{1}{1-3p}[-p+s_1]\,,$$

$$q(2\,;\boldsymbol{s})=\frac{1}{(1-p)(1-3p)}[1-3p+p^2-(1-3p)s_0-(1-2p)s_1]\,.$$

For (17) we have

$$w_0(0\,;\boldsymbol{s})=p(0,1)q(1\,;\boldsymbol{s})+p(0,2)q(2\,;\boldsymbol{s})=\frac{-p(p-s_1)}{1-3p}\,,$$

$$w_0(1\,;\boldsymbol{s})=\frac{1}{1-3p}[p^2+(1-3p)s_0-ps_1]\,,$$

$$w_0(2\,;\boldsymbol{s})=w_0(0\,;\boldsymbol{s})+w_0(1\,;\boldsymbol{s})\,.$$

When $0<p<1/3$, we have

$$w_0(0\,;\boldsymbol{s})\leqq w_0(1\,;\boldsymbol{s})\Longleftrightarrow s_1\leqq p+\frac{1-3p}{2p}s_0\,,$$

$$w_0(1\,;\boldsymbol{s})\leqq w_0(2\,;\boldsymbol{s})\Longleftrightarrow s_1\geqq p\,,$$

$$w_0(2\,;\boldsymbol{s})\leqq w_0(0\,;\boldsymbol{s})\Longleftrightarrow s_1\geqq p+\frac{1-3p}{p}s_0\,.$$

If $1/3<p<1/2$, we have

$$w_0(0\,;\boldsymbol{s})\leqq w_0(1\,;\boldsymbol{s})\Longleftrightarrow s_1\geqq p+\frac{1-3p}{2p}s_0\,,$$

$$w_0(1\,;\boldsymbol{s})\leqq w_0(2\,;\boldsymbol{s})\Longleftrightarrow s_1\leqq p\,,$$

$$w_0(2\,;\boldsymbol{s})\leqq w_0(0\,;\boldsymbol{s})\Longleftrightarrow s_1\leqq p+\frac{1-3p}{p}s_0\,.$$

Thus we can obtain the following partition of 2-dimensional simplex S_2 (see Fig. 1-a and -b).

Along the same line we have the following relations and the Figs. 2 and 3.

$$w_1(0\,;\boldsymbol{s})\leqq w_1(1\,;\boldsymbol{s})$$

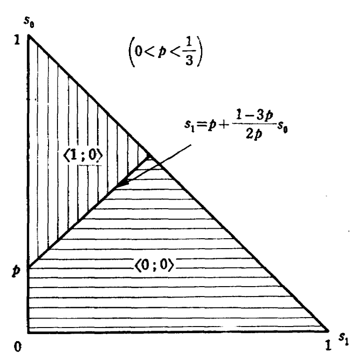
$$\Longleftrightarrow\begin{cases} s_1\leqq\dfrac{p(1-2p)^2}{1-3p+3p^2}+\dfrac{p(1-3p)}{1-3p+3p^2}s_0 & \left(0<p<\dfrac{1}{3}\right) \\ s_1\geqq\dfrac{p(1-2p)^2}{1-3p+3p^2}+\dfrac{p(1-3p)}{1-3p+3p^2}s_0 & \left(\dfrac{1}{3}<p<\dfrac{1}{2}\right)\end{cases}$$

$$w_1(1\,;\boldsymbol{s})\leqq w_1(2\,;\boldsymbol{s})$$

$$\Longleftrightarrow\begin{cases} s_1\geqq\dfrac{p(1-2p)^2}{1-3p+3p^2}+\dfrac{p(1-3p)}{1-3p+3p^2}(1-s_0) & \left(0<p<\dfrac{1}{3}\right)\\[2ex] s_1\leqq\dfrac{p(1-2p)^2}{1-3p+3p^2}+\dfrac{p(1-3p)}{1-3p+3p^2}(1-s_0) & \left(\dfrac{1}{3}<p<\dfrac{1}{2}\right)\end{cases}$$

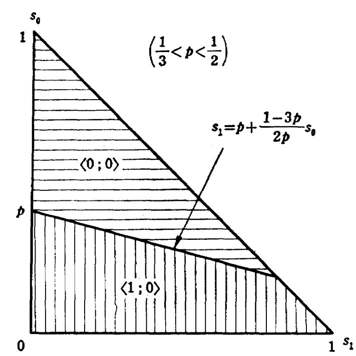
$$w_1(2\,;\boldsymbol{s})\leqq w_1(0\,;\boldsymbol{s})\Longleftrightarrow s_0\leqq\frac{1}{2}\qquad\left(0<p<\frac{1}{2}\right).$$

Combining Figs. 1∼3 according to the relation (19), we have such partitions of S_m as is given in Fig. 4 when $0<p<2-\sqrt{3}$ and as in Fig. 5 when $1/3<p<3/8$.



$\langle 2\,;0\rangle=\phi$

Fig. 1-a



$\langle 2\,;0\rangle=\phi$

Fig. 1-b

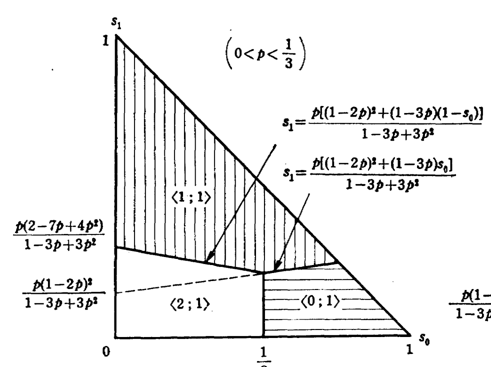


Fig. 2-a

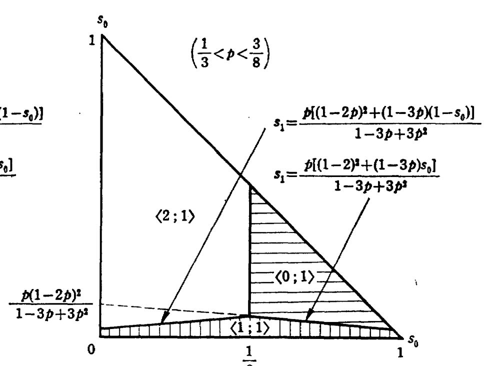
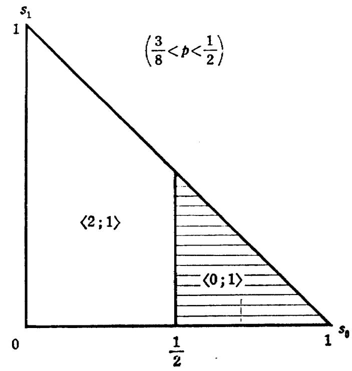
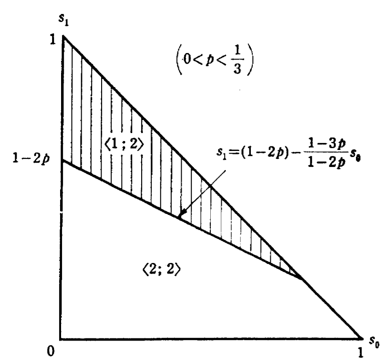


Fig. 2-b



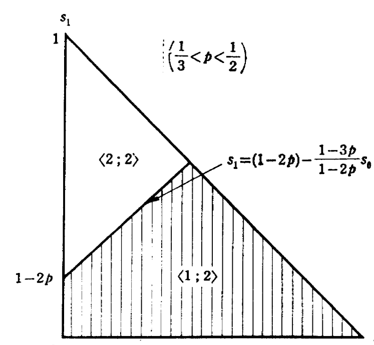
⟨1 ; 1⟩$=\phi$

Fig. 2-b′



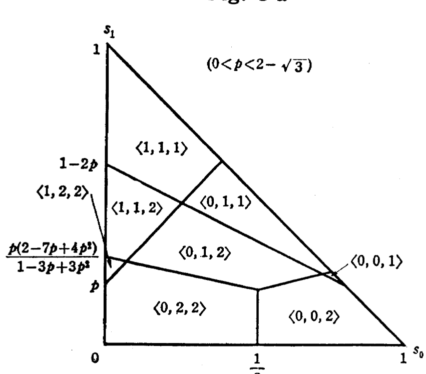
⟨0 ; 2⟩$=\phi$

Fig. 3-a



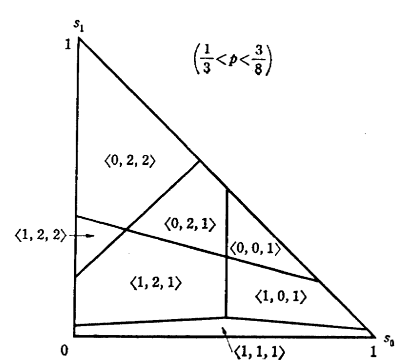
⟨0 ; 2⟩$=\phi$

Fig. 3-b



$$\left.\begin{array}{l}\langle 2,\cdot,\cdot\rangle\\ \langle\cdot,\cdot,0\rangle\\ \langle 1,0,\cdot\rangle\\ \langle\cdot,2,1\rangle\end{array}\right\}=\phi$$

Fig. 4



$$\left.\begin{array}{l}\langle 2,\cdot,\cdot\rangle\\ \langle\cdot,\cdot,0\rangle\\ \langle 0,1,\cdot\rangle\\ \langle\cdot,0,2\rangle\\ \langle 1,1,2\rangle\end{array}\right\}=\phi$$

Fig. 5

3. Non-simple decision function

3.1. Estimation of the proportion vector. When both the true value and an approximate value of $\boldsymbol{s}^0$ are not available, we must construct an estimator $\hat{\boldsymbol{s}}$ for $\boldsymbol{s}^0=\boldsymbol{s}(\boldsymbol{\theta}_n)$. If this be done, we can use the non-simple decision $\hat{\boldsymbol{d}}$ which is defined by

$$\hat{\boldsymbol{d}}=\boldsymbol{d}, \qquad \text{if } \hat{\boldsymbol{s}} \in \langle \boldsymbol{d} \rangle . \tag{24}$$

When the true parameter vector is $\boldsymbol{\theta}_n$, we shall denote the expected average loss suffered by using $\hat{\boldsymbol{d}}$ by

$$r_n(\hat{\boldsymbol{d}}\,;\boldsymbol{\theta}_n)=E\{l_n(\hat{\boldsymbol{d}}\,;\boldsymbol{\theta}_n)\mid \boldsymbol{\theta}_n\}.$$

Then from the theorem 1 we get

$$r_n(\hat{\boldsymbol{d}}\,;\boldsymbol{\theta}_n)-l_n(\boldsymbol{\theta}_n)\leqq B\gamma(\hat{\boldsymbol{s}}\,;\boldsymbol{\theta}_n) \tag{25}$$

where

$$\begin{aligned}\gamma(\hat{\boldsymbol{s}}\,;\boldsymbol{\theta}_n)&=\int_{S_m}\|\boldsymbol{s}-\boldsymbol{s}^0\|\,dF(\boldsymbol{s})\\&=\sum_{i=0}^{m-1}\int_0^1|s-s_i^0|\,dF_i(s)\\&=\sum_{i=0}^{m-1}\gamma(\hat{s}_i\,;\boldsymbol{\theta}_n)\end{aligned}$$

and F and F_i are the distribution function of $\hat{\boldsymbol{s}}=(\hat{s}_0,\cdots,\hat{s}_{m-1})$ and the marginal distribution function of $\hat{s}_i$, respectively.

In order to obtain an estimator $\hat{\boldsymbol{s}}$ for $\boldsymbol{s}^0=\boldsymbol{s}(\boldsymbol{\theta}_n)$, we may consider the following random vector

$$\boldsymbol{Z}=\boldsymbol{Z}(\boldsymbol{X}_n)=(Z_0, Z_1,\cdots, Z_{m-1}) \tag{26}$$

where

$$\begin{aligned}Z_j=Z_j(\boldsymbol{X}_n)&=\sum_{i=1}^{n}u(X_i-j)\\&=\{\text{no. of } i\,;\ X_i=j,\ i=1,\cdots,n\}.\end{aligned} \tag{27}$$

We put

$$M_j=\{i\in M\,; p(i,j)>0\},$$

$$m_j=\{\text{no. of elements of } M_j\}=\sum_{i=0}^{m}\operatorname{sgn}[p(i,j)],$$

$$\boldsymbol{p}(j)=(p(i,j)\,;i\in M_j),$$

$$K_i=\{j\in K\,;0<p(i,j)<1\}\,,$$

$$K_i^*=\{j\in K\,;p(i,j)=1\}\,.$$

Then, when the true parameter vector is $\boldsymbol{\theta}_n$, random vector $\boldsymbol{Z}$ can be expressed as

$$\boldsymbol{Z}=\sum_{j=0}^{k}\boldsymbol{y}_j\,,$$

where $\tilde{\boldsymbol{y}}_j=(y_{ij}\,;i\in M_j)$ of $\boldsymbol{y}_j=(y_{0j},y_{1j},\cdots,y_{m-1,j})$ is distributed as m_j-variate multinomial $M(n_j^0\,;\boldsymbol{p}(j))$ and $n_j^0=n\eta_j(\boldsymbol{\theta}_n)$. Since

$$Z_i=\sum_{j\in K}y_{ij}=\sum_{j\in K_i}y_{ij}+\sum_{j\in K_i^*}y_{ij}$$

and the marginal distribution of y_{ij} for $j\in K_i$ is the binomial $B(n_j^0\,;p(i,j))$, we have

$$E\{Z_i\,|\,\boldsymbol{\theta}_n\}=\sum_{j\in K}E\{y_{ij}\,|\,\boldsymbol{\theta}_n\}=\sum_{j\in K}n_j^0p(i,j)=ns_i^0$$

and

$$E\{|\,Z_i-ns_i^0\,|\}\leqq\sum_{j\in K_i}E\{|\,y_{ij}-n_j^0p(i,j)\,|\}$$

$$=\sum_{j\in K_i}\nu(n_j^0\,;p(i,j))\,,$$

where $\nu(n\,;p)$ denotes the mean deviation of the binomial $B(n\,;p)$. When we set

$$\hat{\boldsymbol{s}}=\frac{1}{n}\boldsymbol{Z}\,, \tag{28}$$

$\hat{\boldsymbol{s}}$ is an unbiased estimator of $\boldsymbol{s}^0=\boldsymbol{s}(\boldsymbol{\theta}_n)$.

3.2. Exact convergence rate. First, we shall state lemmas without proof (cf. [12]).

LEMMA 1. *Let $\nu(n,p)$ be the mean deviation of the binomial $B(n;p)$. Then, for $pn>1$, we have*

$$\nu(n,p)\leqq 2\Delta_2\exp\frac{\alpha(n,p)}{n}\left\{\sqrt{\frac{1-p}{p}}n+\frac{1}{\sqrt{1-p}}\right\},$$

where

$$\Delta_2=\left(\frac{2}{3}e\right)^{-3/2},$$

$$\alpha(n, p)=\frac{1+\frac{2}{n}}{\left(p-\frac{1}{n}\right)(1-p)} .$$

LEMMA 2. *For any positive integer l and any positive numbers $\alpha_0, \alpha_1, \cdots, \alpha_l$, we have*

$$\sum_{i=0}^{l} \sqrt{\alpha_i} \leqq \sqrt{(l+1) \sum_{i=0}^{l} d_i} .$$

From the lemma 1 it follows that for every n such that $\min_{i \in M} \min_{j \in K_i} n_j^0 p(i, j)>1$, we have

$$\begin{aligned} (29) \qquad E\{|z_i-ns_i^0|\} &\leqq \sum_{j \in K_i} \nu(n_j^0 ; p(i, j)) \\ &\leqq 2\Delta_2 \exp \frac{\alpha(n ; \boldsymbol{\theta}_n)}{n} \sum_{j \in K_i} \left\{\sqrt{\frac{1-p(i, j)}{p(i, j)} n_j^0}+\frac{1}{\sqrt{1-p(i, j)}}\right\}, \end{aligned}$$

where

$$(30) \qquad \alpha(n ; \boldsymbol{\theta}_n)=\max_{i \in M} \max_{j \in K_i} \alpha(n_j^0, p(i, j)) .$$

Then by the lemma 2, we get

$$\begin{aligned} (31) \quad \gamma(\hat{\boldsymbol{s}} ; \boldsymbol{\theta}_n) &= \sum_{i=0}^{m-1} \gamma(\hat{s}_i ; \boldsymbol{\theta}_n)=\frac{1}{n} \sum_{i=0}^{m-1} E\{| Z_i-ns_i^0|\} \\ &\leqq \frac{2}{n} \Delta_2 \exp \frac{\alpha(n ; \boldsymbol{\theta}_n)}{n} \sum_{i=0}^{m-1} \sum_{j \in K_i} \left\{\sqrt{\frac{1-p(i, j)}{p(i, j)} n_j^0}+\frac{1}{\sqrt{1-p(i, j)}}\right\} \\ &\leqq \frac{2}{n} \Delta_2 \exp \frac{\alpha(n ; \boldsymbol{\theta}_n)}{n} \sqrt{\sum_{i=0}^{m-1} k_i} \left\{\sqrt{\sum_{i, j} \frac{1-p(i, j)}{p(i, j)} n_j^0}+\sqrt{\sum_{i, j} \frac{1}{1-p(i, j)}}\right\} , \end{aligned}$$

where

$$\begin{aligned} k_i &= \{\text{no. of elements of } K_i\} \\ &= \sum_{j \in K} \operatorname{sgn}[p(i, j)] \operatorname{sgn}[1-p(i, j)] . \end{aligned}$$

Thus from (25) and (31), we have the following

THEOREM 3. *For every $\boldsymbol{\theta}_n \in K^n$ and every $n>\dfrac{1}{\min_{i \in M} \min_{j \in K_i} \eta_j(\boldsymbol{\theta}_n) p(i, j)}$, we have*

$$(32) \qquad r_n(\hat{\boldsymbol{d}} ; \boldsymbol{\theta}_n)-l_n(\boldsymbol{\theta}_n) \leqq \frac{C_1}{\sqrt{n}}+\frac{C_2}{n} ,$$

where

$$(33)\qquad C_1=2\Delta_2 B\exp\left[\frac{\alpha(n;\boldsymbol{\theta}_n)}{n}\right]\sqrt{\sum_{i=0}^{m-1}k_i}\cdot\sqrt{\max_j\sum_{i\in M_j}\frac{1-p(i,j)}{p(i,j)}}\,,$$

$$(34)\qquad C_2=2\Delta_2 B\exp\left[\frac{\alpha(n;\boldsymbol{\theta}_n)}{n}\right]\sqrt{\sum_{i=0}^{m-1}k_i\sum_{j\in K_i}\frac{1}{1-p(i,j)}}\,,$$

and $\hat{\boldsymbol{d}}=\hat{\boldsymbol{d}}(X_n)$ *is the non-simple decision given by (24) with (26), (27) and (28).*

We shall now examine how large the uniform bounds given by (33) and (34) are for the model given by (23) in section 2. Table 1 shows the results for $p=0.05(0.05)0.50$. The bounds are given by

$$C_i=J(\boldsymbol{\theta}_n)\times(\mathrm{III}_i)\,,\qquad i=1,2,$$

where

$$J(\boldsymbol{\theta}_n)=4\Delta_2 A\exp\frac{\alpha(n;\boldsymbol{\theta}_n)}{n}\,.$$

Table 1.

p	(I)	(II_1)	(III_1)	(II_2)	(III_2)
0.05	67.500	28.249	2610.187	11.379	648.583
0.10	5.079	15.923	288.390	15.312	277.310
0.15	5.987	12.694	146.105	13.031	149.983
0.20	7.500	10.747	88.664	11.779	97.178
0.25	10.667	9.416	59.637	11.015	69.763
0.30	20.636	8.446	43.034	10.629	54.159
0.35	43.077	7.716	32.815	10.239	43.543
0.40	13.333	7.165	26.271	10.083	36.971
0.45	9.351	6.758	22.050	10.040	32.758
0.50	8.000	6.000	18.000	8.944	26.883

$$\left\{\begin{aligned}
&(\mathrm{I})=\frac{B}{2A}=\max_{0\le j\le 2}\sum_{i=0}^{2}|q(i,2)-q(i,j)|,\\
&(\mathrm{II}_1)=\sqrt{\sum_{i=0}^{1}k_i\max_{0\le j\le 2}\sum_{i\in M_j}\frac{1-p(i,j)}{p(i,j)}}\,,\\
&(\mathrm{II}_2)=\sqrt{\sum_{i=0}^{1}k_i\sum_{j\in K_i}\frac{1}{1-p(i,j)}}\,,\\
&(\mathrm{III}_i)=(\mathrm{I})\times(\mathrm{II}_i)\qquad(i=1,2)\,.
\end{aligned}\right.$$

For small p these bounds are considerably large. On the other hand, the three distributions in the model (23) become more and more distinguishable from each other so that we may take smaller values as C_1 and C_2.

Thus, the above example shows that there is possibility of improving the theorem 3.

4. Acknowledgements

The author expresses his hearty thanks to Misses Y. Nogami and S. Hiraisi for their help in calculating the numerical examples and typing the manuscript.

THE INSTITUTE OF STATISTICAL MATHEMATICS

REFERENCES

[1] J. F. Hannan, "The dynamic statistical decision problem when the component problem involves a finite number, *m*, of distributions," *Ann. Math. Statist.*, 27 (1956), p. 212 (abstract).

[2] J. F. Hannan, "Approximation to Bayes risk in repeated play," *Contribution to the Theory of Games*, 3 (1957), 97-139.

[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] M. V. Johns, Jr., "The parametric and non-parametric compound decision problems in the sequential case," *Ann. Math. Statist.*, 34 (1963), 1620-1621 (abstract).

[6] M. V. Johns, Jr., "Two-action compound decision problems," Dept. Statist. Stanford Univ., *Technical Report*, No. 87 (1966).

[7] H. Robbins, "Asymptotically subminimax solutions of compound statistical decision problems," *Proc. 2nd Berkeley Symp. Math. Statist. Prob.*, Univ. of California Press, (1951), 131-148.

[8] H. Robbins, "Some numerical results on a compound decision problem," *Recent Developments in Information and Decision Processes*, The Macmillian Co., New York, (1962), 56-62.

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

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

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

[12] G. Suzuki, "Discrete compound decision problem," *Research Memo.*, No. 5 (1966), Inst. Stat. Math.

[13] J. R. Van Ryzin, "The compound decision problem with $m \times n$ finite loss matrix," *Ann. Math. Statist.*, 37 (1966), 412-424.