A BAYES MULTIPLE DECISION PROCEDURE FOR SELECTING THE BEST ONE AMONG SEVERAL NORMAL POPULATIONS WITH COMMON KNOWN VARIANCE

By

Yukio Nomach and Takashi Matsuda

(Kochi University)

§ 1. Summary. A problem which selects the best one among several assigned normal populations with known variances, has been discussed by many authors such as Bechhofer, R. E. (1), Paulson, E. (2), Nomachi, Y. (3) and many others.

The purpose of this paper is to present a Bayes procedure for selecting the best one among \( k \) normal populations with common known variance, depending on \( a \) priori sample means. The mathematical model appears in quality control and analysis of variance model and etc.

In order to progress our comprehensions, let us consider \( k \) processes producing items continuously. In this case, it is often desired to chose the best lot (the “best” lot is the lot whose percentage of defective items in it is of least among these \( k \) lots), or to ranking \( k \) lot means in the ascending order of magnitude.

In what follows, for fixed \( n \) and \( r \), let us state the sampling scheme as follows:

1. For each \( i \) in \( 1 \leq i \leq k \), at the first stage of sampling, it is assumed that the unobservable sample mean \( \mu_1^{(i)} \) is drawn from a priori normal population \( N(\nu_1^{(i)}, \tau^2) \), where \( \tau^2 \) is known, but \( \nu_1^{(i)} \) is unknown to us. Then the observable sample \( \{x_1^{(i)}, x_2^{(i)}, \ldots, x_r^{(i)}\} \) is drawn from the normal population \( N(\mu_1^{(i)}, \sigma^2) \), where \( \sigma^2 \) is known to us. Then let us go to the next stage of sampling, and so on.

2. For each \( i \) in \( 1 \leq i \leq k \), at the \( n \)-th stage of sampling, it is assumed that the unobservable sample mean \( \mu_n^{(i)} \) is drawn from the normal population \( N(\nu_1^{(i)}, \tau^2) \). Then the observable sample \( \{x_1^{(i)}, x_2^{(i)}, \ldots, x_r^{(i)}\} \) is drawn from the normal population \( N(\mu_n^{(i)}, \sigma^2) \).

3. For each \( i \) in \( 1 \leq i \leq k \), at the \((n+1)\)-th stage of sampling after the sample \( \{x_1^{(i)}, x_2^{(i)}, \ldots, x_r^{(i)}\} \) was drawn from the normal population \( N(\mu_n^{(i)}, \tau^2) \), we want to rank unobservable means \( (i=1, 2, \ldots, k) \) in the ascending order of magnitude or to choose the largest mean among them.

In what follows, let us define the \((n+1)\)-th means by

\[ y_j^{(i)} = \sum_{h=1}^{r} x_h^{(i)}, n/r \quad (i=1, 2, \ldots, k). \]

It is well known to estimate the mean \( \mu_{n+1}^{(i)} \) at the present \((n+1)\)-th stage by use of
\[ y_j^{(i)} \quad (j = 1, 2, \ldots, n+1). \]

Let us assume here that the unobservable variables \( \mu_j^{(i)} \quad (j = 1, 2, \ldots, n+1) \) are independently and identically distributed according to the common normal distribution \( N(\nu^{(i)}, \sigma^2) \). Therefore all prior observations \( y_j^{(i)} \quad (j = 1, 2, \ldots, n+1) \) can supply certain informations to \( \mu_j^{(i)} \) through the distribution \( N(\nu^{(i)}, \sigma^2) \), \( (i = 1, 2, \ldots, k) \) respectively.

It is our object in the next section to present a Bayes procedure for ranking means \( \mu_j^{(i)} \quad (i = 1, 2, \ldots, k) \), depending upon the past \( n \) observations \( y_j^{(i)} \quad (j = 1, 2, \ldots, n; \quad i = 1, 2, \ldots, k) \), and to present certain numerical comparisons between our results and Bechhofer's (1). These comparisons may recommend us to take prior observations in the cases when our mathematical model is applicable.

\section*{§ 2. Notations and definitions.} In order to develope our theory, let us prepare the following preliminaries.

**Notations I.** Let us write a vector in \( k \) space \( \Omega \) by

\[ \mu_{n+1} = (\mu_{n+1}^{(1)}, \mu_{n+1}^{(2)}, \ldots, \mu_{n+1}^{(k)}). \]

For any set of non-negative values of \( d \quad (i = 1, 2, \ldots, k) \), let us define such three types of set of \( \mu_{n+1} \) that

\[ \Omega(\mu_{n+1}) = \{ \mu_{n+1} | \mu_{n+1}^{(i)} + d^{(i)} = \mu_{n+1}^{(i)} (i = 2, 3, \ldots, k) \} \]

\[ \Omega(d) = \{ \mu_{n+1} | \mu_{n+1}^{(i)} + d^{(i)} = \mu_{n+1}^{(i)} (i = 2, 3, \ldots, k) \} \]

and that

\[ \Omega(d^{(i)}) = \{ \mu_{n+1} | \mu_{n+1}^{(i)} = \mu_{n+1}^{(i)} \quad \text{for} \quad j \neq i \quad \text{and} \quad \mu_{n+1}^{(i)} + d^{(i)} = \mu_{n+1}^{(i)} \} \]

**Definition of decision rules.** Let \( \hat{\mu}_{n+1}^{(i)} \) be certain estimate of \( \mu_{n+1}^{(i)} \) based on observations drawn previously, \( i = 1, 2, \ldots, k \) respectively. Let us define that

\[ \mu_{n+1}^{(m)} = \max_{1 \leq i \leq k} \mu_{n+1}^{(i)} \quad \text{........................................................................... (4)} \]

is true, whenever

\[ \hat{\mu}_{n+1}^{(m)} = \max_{1 \leq i \leq k} \hat{\mu}_{n+1}^{(i)} \quad \text{........................................................................... (5)} \]

holds true. The event \( \hat{\mu}_{n+1}^{(i)} = \hat{\mu}_{n+1}^{(j)} \quad (i \neq j) \) is an event of probability zero and we may ignore in probability calculations. However, this event may occur in the actual applications. If this event occur the tied means may be ranked by means of a randomized technique which assigns equal probability to each ordering. Let us say such event that the relation (4) holds true when the relation (5) is true by the correct selection (CS) of means.

**Notation II.** Let \( (CS|\Omega(\cdot)) \) be the event of correct selection of the best mean in the case when \( \mu_{n+1} \in \Omega(\cdot) \), where we define that the best mean is the maximum mean among all \( k \) means \( \mu_{n+1}^{(i)} \quad (i = 1, 2, \ldots, k) \).
A Bayes multiple decision procedure for selecting the best one

Lemma 1. Under the assumptions stated previously for the distributions of $\mu_{j}^{(i)}$ and $y_{j}^{(i)}$ ($j=1,2,\ldots,n+1; i=1,2,\ldots,k$), we have that for each $i$ in $1 \leq i \leq k$, the conditional distribution of $\mu_{n+1}^{(i)}$, given $y_{n+1}^{(i)}$ is the normal a posteriori distribution

$$N\left( \frac{r y_{n+1}^{(i)} + c^2 y_{n+1}^{(i)}}{r+c^2}, \frac{\sigma^2}{r+c^2} \right).$$

where $c=\sigma/\tau$.

Proof. We can directly obtain the result from the conditional density function by use of the properties of normal density functions. (e.g. p. 380 in (4))

In Lemma 1, however, we cannot determine the normal a posteriori distribution, because we assumed that the value of $\nu^{(i)}$ ($i=1,2,\ldots,k$) were unknown to us.

In this case we will use the usual estimate

$$y_{n+1}^{(i)} = \sum_{j=1}^{n+1} y_{j}^{(i)}/(n+1)$$

for $\nu^{(i)}$ ($i=1,2,\ldots,k$), respectively. Then we have the following lemma:

Lemma 2. For each $i$ in $1 \leq i \leq k$, let us put that

$$\hat{\mu}_{n+1}^{(i)} = (r y_{n+1}^{(i)} + c^2 y_{n+1}^{(i)})/(r+c^2),$$

then $\hat{\mu}_{n+1}^{(i)}$ has the normal distribution

$$N\left( \mu_{n+1}^{(i)}, \frac{\sigma^2}{r+c^2} \left(1+\frac{c^2}{r(n+1)}\right) \right)$$

where we put that

$c=\sigma/\tau$ and $y_{n+1}^{(i)} = \sum_{j=1}^{n+1} y_{j}^{(i)}/(n+1)$.

Proof. We can directly obtain the result from the property of normal density functions. (e.g. p. 382 in (4))

Lemma 3. Let $X^{(1)}, X^{(2)}, \ldots, X^{(k)}$ are independent and normally distributed, $NID(\mu^{(i)}, \sigma^2)$ ($i=1,2,\ldots,k$), random variables. Then for any real $u$ we have

$$P\left\{ \max_{1 \leq i \leq k} X^{(i)} \leq u \right\} = \prod_{i=1}^{k} \Phi((u - \mu^{(i)})/\sigma),$$

where for the standard normal density function $\phi(t)$ we have put

$$\phi(x) = \int_{-\infty}^{x} \phi(t) dt.$$
\[
\sum_{t=1}^{k} P \left\{ X^{(t)} \leq u \right\} = \frac{1}{\sqrt{2\pi}} \frac{k}{\sigma} \Phi \left( \frac{u - \mu^{(t)}}{\sigma} \right).
\]

**Lemma 4.** Under the same conditions as in Lemma 3, we have
\[
P \left\{ CS \mid Q(\mu_{n+1}) \right\} = \int_{-\mu^{(k,1)}/\sqrt{2\sigma}}^{-\mu^{(k,1)}/\sqrt{2\sigma}} \exp \left\{ -\frac{1}{2} t' P^{-1} t \right\} dt,
\]
where
\[
y' = (y_1, y_2, \ldots, y_k), \quad \mu' = (\mu^{(1,1)}, \mu^{(2,2)}, \ldots, \mu^{(k,k)}), \quad t' = (y - \mu)' / \sqrt{2\sigma}
\]
\[
y_j = X^{(2)} - X^{(j-1)}, \quad u^{(j-1)} = \mu^{(j-1)} - \mu^{(j-1)}, \quad (j = 2, 3, \ldots, k)
\]
and where \( P = (\rho_{ij}) \) denotes the \( k-1 \) by \( k-1 \) correlation matrix with
\[
\rho_{ij} = \begin{cases}     1 & \text{for } i = j, \\     \frac{1}{2} & \text{for } i \neq j. \end{cases} (i, j = 2, 3, \ldots, k)
\]

**Proof.** By virtue of the transformations of variables
\[
y^{(j)} = X^{(j)} - X^{(j-1)} \quad (j = 2, 3, \ldots, k)
\]
the joint probability density function of \( y^{(j)} \) \((j = 2, 3, \ldots, k)\) is given by
\[
f(y^{(2)}, y^{(3)}, \ldots, y^{(k)}) = \frac{1}{(2\pi)^{(k-1)/2}} \exp \left\{ -\frac{1}{2} (y - \mu)' \Sigma^{-1} (y - \mu) \right\},
\]
where \( \mu \) and \( \Sigma \) are given by (13), and \( \Sigma = (s_{ij}) \) is the variance-covariance matrix of \( y_j \) \((j = 2, 3, \ldots, k)\) with
\[
s_{ij} = \begin{cases}     2 & \text{for } i = j, \\     1 & \text{for } |i - j| = 1, \\     0 & \text{for } |i - j| \geq 2. \end{cases} (i, j = 2, 3, \ldots, k)
\]
Therefore we obtain
\[
P \left\{ CS \mid Q(\mu) \right\} = P \left\{ X^{(1)} = X^{(2)} = \ldots = X^{(k)} \mid Q(\mu) \right\}
\]
\[
= P \left\{ Y^{(j)} \geq 0, \; j = 2, 3, \ldots, k \mid Q(\mu) \right\}
\]
\[
= \int_{0}^{\infty} \cdots \int_{0}^{\infty} \frac{1}{(2\pi)^{(k-1)/2}} \exp \left\{ -\frac{1}{2} (y - \mu)' \Sigma^{-1} (y - \mu) \right\} dy
\]
\[
= \int_{-\mu^{(2,1)}/\sqrt{2\sigma}}^{-\mu^{(2,1)}/\sqrt{2\sigma}} \cdots \int_{-\mu^{(k,k-1)}/\sqrt{2\sigma}}^{-\mu^{(k,k-1)}/\sqrt{2\sigma}} \frac{k^{-1/2}}{(2\pi)^{(k-1)/2}} \exp \left\{ -\frac{1}{2} t' P^{-1} t \right\} dt,
\]
where we put \( t' = (y - \mu)' / \sqrt{2\sigma} \), and \( P = (\rho_{ij}) \) is the \( k-1 \) by \( k-1 \) correlation matrix with
\[
\rho_{ij} = \begin{cases}     1 & \text{for } i = j, \\     \frac{1}{2} & \text{for } i \neq j. \end{cases} (i, j = 2, 3, \ldots, k)
**Theorem.** Under the same conditions and the same notations as in Lemma 3 and Lemma 4, we have

\[
P(CS|\Omega(\mu_{n+1})) \geq P(CS|\Omega(\mu)) \geq P(CS|\Omega(\mu^{(k)}))
\]

\[
= \int_{0}^{\infty} \cdots \int_{0}^{\infty} \frac{k^{-1/2}}{(2\pi)^{k-1/2}} \exp \left\{-\frac{1}{2} t' P^{-1} t\right\} dt, \tag{22}
\]

where \( P \) was given by (21) and

\[
\sigma^2 = \frac{\sigma^2}{r+c^2} \left\{1 + \frac{c^2}{r(n+1)}\right\}, \quad c = \sigma/r. \tag{23}
\]

**Proof.** Substituting \( \mu_i \) given by (8) into \( x^{(i)} \) (\( i=1,2,\ldots,k \)), respectively, and \( \sigma^2 \) given by (23) into \( \sigma \) used in Lemma 3 and Lemma 4 the resulting relation (23) directly follows.

**Corollary.** Under the same conditions and the same notations as in Theorem, for any preassigned value of \( \alpha \) in \( 0 < \alpha \leq 1/k \), there exists such unique value of \( \mu^{(k)}(\alpha) \) that

\[
P(CS|\Omega(\mu^{(k)}(\alpha))) = 1 - \alpha \tag{24}
\]

holds true.

**Proof.** The function \( P(CS|\Omega(\mu)) \) of non-negative value of \( \mu^{(k)} \) is monotone, continuous, increasing and bounded above by one. In special case when we know that \( \mu^{(k)} = 0 \), the correct selection of maximum mean is attained by use of certain chance mechanism which assigns equal probability \( 1/k \) to respective selection of mean. Therefore for any preassigned value of \( \alpha \) in \( 0 < \alpha \leq 1/k \),

\[
P(CS|\Omega(\mu^{(k)}(\alpha))) = 1 - \alpha
\]

holds true.

**§ 3. Conclusion** If we are in the position where a priori informations are available as was in our paper, it is recommended to use them all in such a manner as was stated by us.

While the numerical calculations of the result of the theorem in this paper are preparing now.

**References**


(Received September 24, 1970)