AN OPTIMUM SOLUTION TO THE k-SAMPLE SLIPPAGE PROBLEM FOR THE NORMAL DISTRIBUTION[1]

By Edward Paulson

University of Washington

0. Summary. A slippage problem for normal distributions is formulated as a multiple decision problem, and a solution is obtained which has certain optimum properties. The discussion is confined to the fixed sample case with the same number of observations from each distribution, and the normal distributions involved are assumed to have a common but unknown variance.

1. Introduction. This paper will consider the problem of how to compare k categories, such as k varieties of wheat, k machines, k teaching methods, etc., so as to decide on the basis of a random sample of n observations with each category whether or not the categories are equal, and if not which is the 'best' one. A problem of this type has been discussed by Mosteller [1] for the nonparametric case. In previous papers [2], [3], we had considered some different types of multiple-decision problems arising in the comparison of k categories, and the emphasis had been on studying the distribution problems involved when the statistical procedures used were suggested by intuitive considerations. In this paper we will be primarily concerned with finding a statistical procedure which in some reasonable sense is an 'optimum' one.

In this paper we will restrict our attention to the case where the n observations $x_{i1}, x_{i2}, \cdots, x_{in}$ in the ith category Π_i are assumed to be normally and independently distributed with mean m_i and a common standard deviation σ, and the best category is (for convenience) defined to be the one associated with the greatest mean value. Let D_0 denote the decision that the k means are all equal, and let D_j $(j = 1, 2, \cdots, k)$ denote the decision that D_0 is incorrect and $m_j = \max(m_1, m_2, \cdots, m_k)$. Our problem is to find a statistical procedure for choosing one of the $k + 1$ decisions $(D_0, D_1, \cdots, D_k)$ which will be in some sense an optimum one. At this point, instead of introducing a weight function as required in the general theory as developed by Wald [4], we will follow a simpler plan which is somewhat analogous to the classical Neyman-Pearson theory of testing a hypothesis, and attempt to find a statistical procedure which, subject to certain restrictions, will in certain instances maximize the probability of making the correct decision.

In order to give a more precise formulation and the solution to the problem let $x_{i\alpha}$ denote the αth observation in the sample from Π_i $(i = 1, 2, \cdots, k;$ $\alpha = 1, 2, \cdots, n)$, let $\bar{x}_i = \sum_{\alpha=1}^{n} (x_{i\alpha}/n)$, $\bar{x} = \sum_{i=1}^{k} (\bar{x}_i/k)$, $s^2 = \sum_{i=1}^{k} \sum_{\alpha=1}^{n} (x_{i\alpha} - \bar{x}_i)^2/[k(n - 1)]$, and let M be the subscript of the category with the greatest sample mean, so that $\bar{x}_M = \max\{\bar{x}_1, \bar{x}_2, \cdots, \bar{x}_k\}$. We will say that the category Π_i has slipped to the right by an amount Δ $(\Delta > 0)$ if $m_1 = m_2 =$

[1] Work done under the sponsorship of the Office of Naval Research.

$\cdots = m_{i-1} = m_{i+1} = \cdots = m_k$ and $m_i = m_1 + \Delta$. The first formulation of the problem is the following: to find a statistical procedure for selecting one of the decisions $(D_0, D_1, \cdots, D_k)$ which will maximize the probability of making the correct decision when some category has slipped to the right subject to the restriction (a) when all the means are equal, D_0 should be selected with probability $1 - \alpha$ (where α is some small positive number fixed in advance of the experiment). In this formulation, the class of allowable decision procedures seems to be too large to admit of an optimum solution and we will, therefore, limit the class of allowable statistical procedures by the following additional restrictions. (b) The decision procedure must be invariant if a constant is added to all the observations, (c) the decision procedure must be invariant when all the observations are multiplied by a positive constant, and (d) the decision procedure must be symmetric in the sense that the probability of making the correct decision when category Π_i has slipped to the right by an amount Δ must be the same for $i = 1, 2, \cdots, k$. These additional restrictions are rather weak and seem to be reasonable requirements to impose in many practical problems. The problem is now reformulated as follows: to find a statistical procedure for selecting one of the set $(D_0, D_1, \cdots, D_k)$ which, subject to restrictions (a), (b), (c) and (d), will maximize the probability of making the correct decision when one of the categories has slipped to the right. The optimum solution will be shown to be the following procedure:

$$\text{(1)}\qquad \begin{aligned} &\text{if } \frac{n(\bar{x}_M - \bar{x})}{\sqrt{\sum_{i=1}^{k}\sum_{\alpha=1}^{n}(x_{i\alpha} - \bar{x})^2}} > \lambda_\alpha, \text{ select } D_M;\\ &\text{if } \frac{n(\bar{x}_M - \bar{x})}{\sqrt{\sum_{i=1}^{k}\sum_{\alpha=1}^{n}(x_{i\alpha} - \bar{x})^2}} \leqq \lambda_\alpha, \text{ select } D_0. \end{aligned}$$

Here λ_α is a constant whose precise value is determined by requirement (a), and does not depend on Δ or σ. Since for a given k and n the value of λ_α depends only on α, the optimum property of (1) holds uniformly in Δ and σ.

2. Derivation of the optimum procedure. There is obviously no loss of generality in only considering statistical procedures which depend on the set $(\bar{x}_1, \bar{x}_2, \cdots, \bar{x}_k, s^2)$ since these constitute a set of sufficient statistics for the unknown parameters $(m_1, m_2, \cdots, m_k, \sigma^2)$. Making use of this in connection with restrictions (b) and (c) it is easy to see that any allowable decision procedure will depend only on the $k-1$ statistics $(\bar{x}_1 - \bar{x}_k)/s, (\bar{x}_2 - \bar{x}_k)/s, \cdots, (\bar{x}_{k-1} - \bar{x}_k)/s$. Let $w_\alpha = (\bar{x}_\alpha - \bar{x}_k)/s$ and let $a_\alpha = (m_\alpha - m_k)/\sigma$ for $\alpha = 1, 2, \cdots, k-1$. The joint probability distribution of the set $(w_1, w_2, \cdots, w_{k-1})$ depends only on the parameters $(a_1, a_2, \cdots, a_{k-1})$. Let $\bar{D}_0$ denote the decision tha $a_1 = a_2 = \cdots = a_{k-1} = 0$, and for $1 \leqq j \leqq k - 1$ let $\bar{D}_j$ denote the decision that $a_1 = a_2 = \cdots = a_{j-1} = a_{j+1} = \cdots = a_{k-1} = 0$ and $a_j = \Delta/\sigma$, while $\bar{D}_k$ denotes the decision that $a_1 = a_2 = \cdots = a_{k-1} = -\Delta/\sigma$. Since any allowable decision procedure for selecting one of the set $(D_0, D_1, \cdots, D_k)$ must be a function only of $(w_1, w_2, \cdots, w_{k-1})$ it can be transformed in a natural manner into a decision procedure for selecting one of the decisions $(\bar{D}_0, \bar{D}_1, \cdots, \bar{D}_k)$ by making D_i correspond to $\bar{D}_i$ for $i = 0, 1, 2, \cdots, k$; that is, whenever the original decision procedure selects D_i the transformed decision procedure is to select $\bar{D}_i$. Because of restriction (a), the probability that any transformed allowable decision procedure will select $\bar{D}_0$ when $a_1 = a_2 = \cdots = a_{k-1} = 0$ will be equal to $1 - \alpha$; in addition the probability that any allowable decision procedure will select D_i when Π_i has slipped to the right by an amount Δ is equal to the probability that the transformed procedure select $\bar{D}_i$ when $\bar{D}_i$ is the correct decision, and this last probability must be the same for each i because of restriction (d).

The proof that (1) is the optimum solution consists mainly in showing that for any Δ and σ there exist a set of nonzero a priori probabilities $g_0, g_1, \cdots, g_k$ which are functions of Δ and σ so that when (1) is transformed in the manner indicated above into a decision procedure for selecting one of $(\bar{D}_0, \bar{D}_1, \cdots, \bar{D}_k)$, it will maximize the probability of making the correct decision among the set $(\bar{D}_0, \bar{D}_1, \cdots, \bar{D}_k)$ when g_i is the a priori probability that $\bar{D}_i$ is the correct decision. Assuming this has been demonstrated, it follows easily that (1) must be the optimum solution. For suppose there existed an allowable decision procedure D^*, which for some Δ and σ had a greater probability than (1) of making the correct decision when some category had slipped to the right by an amount Δ. Then D^*, which must be a function only of $(w_1, w_2, \cdots, w_{k-1})$ when transformed in the indicated manner into a decision procedure for selecting one of $(\bar{D}_0, \bar{D}_1, \cdots, \bar{D}_k)$ will have a greater probability than (1) of making the correct decision among $(\bar{D}_0, \bar{D}_1, \cdots, \bar{D}_k)$ with respect to any set of non-zero a priori probabilities, which would be a contradiction.

To show that the required a priori distribution exists, first let $u_\alpha = (\bar{x}_\alpha - \bar{x}_k)/\sigma$ $(\alpha = 1, 2, \cdots, k-1)$ so that $w_\alpha = (u_\alpha \sigma/s)$. The random variables $(u_1, u_2, \cdots, u_{k-1})$ can easily be verified to have a $(k-1)$ dimensional multivariate normal distribution with common variance $= 2/n$, common correlation $= \frac{1}{2}$, and mean values $(a_1, a_2, \cdots, a_{k-1})$. By an elementary calculation, the joint probability density function of $u_1, u_2, \cdots, u_{k-1}$ is given by $C_1 \exp [-\frac{1}{2}\{A \sum_{\alpha=1}^{k-1}(u_\alpha - a_\alpha)^2 + B\sum_{\alpha\neq\beta} (u_\alpha - a_\alpha)(u_\beta - a_\beta)\}]$ where $A = ((k-1)n/k)$, $B = -n/k$, and C_1 is a constant whose precise value is not needed. Using this result plus the known facts that $n's^2/\sigma^2$ has the χ^2 distribution with $n' = k(n-1)$ degrees of freedom and is independent of the set $u_1, u_2, \cdots, u_{k-1}$, the joint probability density function $f(w_1, w_2, \cdots, w_{k-1})$ of $w_1, \cdots, w_{k-1}$ is easily found to be given by

$$f(w_1, w_2, \cdots, w_{k-1}) = C_2 \int_0^\infty y^{n'+k-2} \exp\left[\quad \left\{ n' y^2 + A \sum_{\alpha=1}^{k-1} (w_\alpha y - a_\alpha)^2 + B \sum_{\alpha\neq\beta} (w_\alpha y - a_\alpha)(w_\beta y - a_\beta) \right\} \right] dy. \tag{2}$$

Let $f_i = f(w_1, \cdots, w_{k-1} \mid \bar{D}_i)$ denote the joint probability density function of $w_1, \cdots, w_{k-1}$ when $\bar{D}_i$ is the correct decision. The decision procedure which will maximize the probability of making the correct decision among the set $(\bar{D}_0, \bar{D}_1, \cdots, \bar{D}_k)$ when the a priori probability distribution is $(p_0, p_1, p_2, \cdots, p_k)$, that is, the Bayes solution with respect to $(p_0, p_1, \cdots, p_k)$, is known [4] to be given by the rule: for each j $(j = 0, 1, \cdots, k)$ select $\bar{D}_j$ for all points in the $w_1 \cdots w_{k-1}$ space where $p_j f_j = \max \{p_0 f_0, p_1 f_1, \cdots, p_k f_k\}$. For the problem at hand, this is the unique Bayes solution except possibly for a set of measure zero according to all f_i. Using (2) it is easy to calculate for each j the region where $\bar{D}_j$ is selected for the special a priori distribution $p_0 = (1 - kp)$, $p_1 = p_2 \cdots = p_k = p$. For example the region where $\bar{D}_1$ is selected is given by the points in the w space where $f_1 > f_2, f_1 > f_3, \cdots, f_1 > f_k$, and $pf_1 > (1 - kp)f_0$. For any j with $1 < j < k$, the region where $f_1 > f_j$ is given by

$$\int_0 y^{n'+k-2} \exp\left[-\tfrac{1}{2}\left(n' y^2 + Ay^2 \sum_{\alpha=1}^{k-1} w_\alpha^2 + By^2 \sum_{\alpha \neq \beta} w_\alpha w_\beta + A \frac{\Delta^2}{\sigma^2} - 2B\frac{\Delta}{\sigma} y \sum_{\alpha=1}^{k-1} w_\alpha\right)\right] \cdot \left\{\exp\left[A\frac{\Delta}{\sigma} yw_1 - B\frac{\Delta}{\sigma} yw_1\right] - \exp\left[A\frac{\Delta}{\sigma} yw_j - B\frac{\Delta}{\sigma} yw_j\right]\right\} dy > 0.$$

The integrand is positive for all y in the range $0 < y < \infty$ if $w_1 > w_j$, and the integrand is negative for all y in this range when $w_1 < w_j$, (since $A - B > 0$) so that $f_1 > f_j$ for $1 < j < k$ if and only if $w_1 > w_j$. In a similar manner, it is easy to show that $f_1 > f_k$ if and only if $w_1 > 0$. The region where $pf_1 > (1 - kp)f_0$ is given by

$$\int_0^\infty y^{n'+k-2} \exp\left[-\frac{y^2}{2}\left(n' + A\sum_{\alpha=1}^{k-1} w_\alpha^2 + B\sum_{\alpha\neq\beta} w_\alpha w_\beta\right)\right] \cdot \left\{p \exp\left(-\frac{A\Delta^2}{2\sigma^2}\right) \exp\left[\left\{(A - B)\frac{\Delta}{\sigma} w_1 + B\frac{\Delta}{\sigma}\sum_{\alpha=1}^{k-1} w_\alpha\right\} y\right] - (1 - kp)\right\} dy > 0.$$

Making a change of variable, this region is equivalent to

$$\int_0^\infty t^{n'+k-2} \exp\left(-\frac{t^2}{2}\right)\left\{p \exp\left(-\frac{A\Delta^2}{2\sigma^2}\right) \exp\left[\frac{\Delta}{\sigma} h(w_1, w_2, \cdots, w_{k-1})\, t\right] - (1 - kp)\right\} dt > 0,$$

where

$$h(w_1, w_2, \cdots, w_{k-1}) = \frac{(A - B)w_1 + B\sum_{\alpha=1}^{k-1} w_\alpha}{\sqrt{n' + A\sum_{\alpha=1}^{k-1} w_\alpha^2 + B\sum_{\alpha\neq\beta} w_\alpha w_\beta}}.$$

The integrand on the left hand side is for all t a monotonically increasing function of $h(w_1, \cdots, w_{k-1})$, so the region where $pf_1 > (1 - kp)f_0$ must be of the type

$h(w_1, \cdots, w_{k-1}) > L$ where L is a number which depends on Δ/σ and p. The other regions can be calculated explicitly in a similar manner, and the Bayes solution is the following procedure: for $1 \leqq j \leqq k - 1$ select $\bar{D}_j$ if $w_j > 0$ and $w_j > \max (w_1, \cdots, w_{j-1}, w_{j+1}, \cdots, w_{k-1})$ and

$$(A - B)w_j + B \sum_{\alpha=1}^{k-1} w_\alpha > L \sqrt{n' + A \sum_{\alpha=1}^{k-1} w_\alpha^2 + B \sum_{\alpha \neq \beta} w_\alpha w_\beta};$$

select $\bar{D}_k$ if $w_j < 0$ for $j = 1, 2, \cdots, k - 1$ and

$$[-A - B(k - 2)] \sum_{\alpha=1}^{k-1} w_\alpha > L \sqrt{n' + A \sum w_\alpha^2 + B \sum w_\alpha w_\beta};$$

otherwise select $\bar{D}_0$. Define the function $F(p)$ by the equation

$$F(p) = \int_0^\infty t^{n'+k-2} \exp\left(-\frac{t^2}{2}\right)\left\{p \exp\left(-\frac{A\Delta^2}{2\sigma^2}\right) \exp\left(\frac{\Delta}{\sigma} \lambda_\alpha t\right) - (1 - kp)\right\} dt,$$

where λ_α is the constant used in (1). It is obvious that $F(p)$ is a continuous function of p with $F(0) < 0$ and $F(1/k) > 0$. Hence there exists a value p^* with $0 < p^* < 1/k$ which is a function of Δ/σ so that $F(p^*) = 0$. Once the Bayes solution relative to $(1 - kp, p, p, \cdots, p)$ has been worked out, it is obvious that to get the Bayes solution relative to $(1 - kp^*, p^*, \cdots, p^*)$ it is only necessary to replace L by λ_α. If we now substitute $w_i = (x_i - x_k)/s$ and replace A and B by their values, we find after some algebraic simplifications that the Bayes solution relative to $(1 - kp^*, p^*, \cdots, p^*)$ reduces to (1) when $\bar{D}_i$ is made to correspond to D_i. Since (1) is an allowable procedure, this proves that it is an optimum one.

3. The calculation of λ_α. The calculation of the exact value of λ_α required in order to have $P\{n(\bar{x}_M - \bar{x}) > \lambda_\alpha \sqrt{\sum_{i=1}^k \sum_{j=1}^n (x_{ij} - \bar{x})^2}\} = \alpha$ when all k means are equal will be extremely difficult until tables are made available, and therefore some approximation is required at present. For this purpose let A_i denote the event $[n(\bar{x}_i - \bar{x}) > \lambda_\alpha \sqrt{\sum_{i=1}^k \sum_{j=1}^n (x_{ij} - \bar{x})^2}]$, so that

$$P\{n(\bar{x}_M - \bar{x}) > \lambda_\alpha \sqrt{\textstyle\sum_{i=1}^k \sum_{j=1}^n (x_{ij} - \bar{x})^2}\}$$

will be equal to the probability of the occurrence of at least one A_i ($i = 1, 2, \cdots, k$). The approximation to be suggested is of a familiar type, and consists in determining λ_α so that $P(A_1) = \alpha/k$. For this purpose, it is clearly legitimate to take $m_1 = m_2 \cdots = m_k = 0$ and $\sigma = 1$. Next, let $y_j = \sqrt{n}\,\bar{x}_j$ ($j = 1, 2, \cdots, k$) so that $\{y_j\}$ constitute a set of independent and standardized normal variables, and let $\bar{y} = (\sum_{i=1}^k y_i/k)$. Then

$$P(A_1) = P\left\{y_1 - \bar{y} > \frac{\lambda_\alpha}{\sqrt{n}} \sqrt{\sum_{i=1}^k \sum_{j=1}^n (x_{ij} - \bar{x}_i)^2 + \sum_{i=1}^k (y_i - \bar{y})^2}\right\}.$$

Now we introduce an orthogonal transformation given by

$$t_1 = \frac{\sum_{i=1}^{k} y_i}{\sqrt{k}},$$

$$t_r = \frac{\sum_{i=r}^{k} y_i - (k - r + 1)y_{r-1}}{\sqrt{(k - r + 1)(k - r + 2)}}, \qquad r = 2, 3, \cdots, k.$$

The new variables $t_1, \cdots, t_k$ also constitute an independent set of normally distributed random variables with zero means and unit variances and are obviously independent of $\sum_{i=1}^{k} \sum_{j=1}^{n} (x_{ij} - \bar{x}_i)^2$. We now have

$$\begin{aligned} P(A_1) &= P\left\{\sqrt{\frac{k-1}{k}}\,(-t_2) > \frac{\lambda_\alpha}{\sqrt{n}} \sqrt{\sum_{i=1}^{k} \sum_{j=1}^{n} (x_{ij} - \bar{x}_i)^2 + \sum_{i=2}^{k} t_i^2}\right\} \\ &= \tfrac{1}{2} P\left\{\frac{(k-1)}{k}\, t_2^2 > \frac{\lambda_\alpha^2}{n}\left(\sum_{i=1}^{k} \sum_{j=1}^{n} (x_{ij} - \bar{x}_i)^2 + \sum_{i=2}^{k} t_i^2\right)\right\} \\ &= \tfrac{1}{2} P\left\{\left(\frac{k-1}{k} - \frac{\lambda_\alpha^2}{n}\right) t_2^2 > \frac{\lambda_\alpha^2}{n} \chi_{n''}^2\right\}, \end{aligned}$$

where $n'' = k(n - 1) + k - 2$ and $\chi_{n''}^2 = \sum_{i=1}^{k} \sum_{j=1}^{n} (x_{ij} - \bar{x}_i)^2 + \sum_{i=3}^{k} t_i^2$ has the chi-square distribution with n'' degrees of freedom and is independent of t_2. If F_0 is used for the value of the F distribution with $n_1 = 1$ and $n_2 = n''$ degrees of freedom which is exceeded with probability $2\alpha/k$, it is a simple matter to verify that the desired approximation is given by

$$\lambda_\alpha = \sqrt{\frac{n(k-1)F_0}{k(n'' + F_0)}}.$$

If λ_α is determined by the above formula so that $P(A_1) = \alpha/k$, it follows at once from Bonferoni's inequality [5] that the probability of not selecting D_0 when all the means are equal will be less than α by an amount which cannot exceed $\frac{1}{2}k(k - 1)P(A_1A_2)$. This quantity is still difficult to evaluate, but in the limit as $n \to \infty$, $\frac{1}{2}k(k - 1)P(A_1A_2)$ can be obtained from tables of the normal bivariate distribution, and is easily shown to be less than $\frac{1}{2}\alpha^2$ for n large enough. Even for small n it seems plausible on an intuitive basis that this bound will be small for values of α ordinarily of interest (say $\alpha \leqq .05$), although further investigation on this point would obviously be desirable. In any event, if the approximation $\lambda_\alpha = \sqrt{n(k - 1)F_0/[k(n'' + F_0)]}$ is used, it can be asserted that for any n the probability of not selecting D_0 when all the means are equal is less than α, and for large n the difference between the true probability and α will be less than $\frac{1}{2}\alpha^2$.

REFERENCES

[1] Frederick Mosteller, "A k-sample slippage test for an extreme population," *Annals of Math. Stat.*, Vol 19 (1948), pp. 58–65.
[2] Edward Paulson, "A multiple decision procedure for certain problems in the analysis of variance," *Annals of Math. Stat.*, Vol. 20 (1949), pp. 95–98.
[3] Edward Paulson, "The comparison of several experimental categories with a control," *Annals of Math. Stat.*, Vol. 23 (June 1952), pp. 239–246.
[4] Abraham Wald, *Statistical Decision Functions*, John Wiley and Sons, 1950.
[5] William Feller, *An Introduction to Probability Theory and its Applications*, John Wiley and Sons, 1950.