ON THE PROBABILITIES OF RANKINGS OF k POPULATIONS WITH APPLICATIONS

by

Raymond J. Carroll

Department of Statistics
University of North Carolina at Chapel Hill

and

Shanti S. Gupta

Purdue University

Institute of Statistics Mimeo Series #1024

August, 1975

SUMMARY

The method of approximating a continuous distribution by a discrete distribution is used to approximate certain multidimensional ranking integrals. In the location and scale parameter cases the method results in a simple iterative counting algorithm. A bound on the error term is given. The algorithm is applied to the problem of completely ranking normal means and shown to be quite accurate and fast. Applications of the above complete ranking problem are given, and the results are used to compute upper confidence bounds for mean differences in a trend situation.

INTRODUCTION

This paper is concerned with presenting approximations to the ranking probabilities

$$(1.1) \quad \Pr\{(X_1,\ldots,X_{t_1}) < (X_{t_1+1},\ldots,X_{t_2}) < \ldots < (X_{t_s+1},\ldots,X_k)\} ,$$

where the independent random variables $X_1,\ldots,X_k$ have the same distribution within each group and differ between groups only by scale or location parameters. The probabilities (1.1) have a number of applications in the ranking of populations according to their scale or location parameters. For example, if there are only two groups (that is, $k = t_2$), the probabilities reduce to the classical correct selection probabilities of the indifference zone formulation (Bechhofer (1954)) and the subset selection formulation (Gupta (1965)); in the context of the latter formulation, the probabilities become a lower bound on the probabilities of correct selection of a subset containing the $t(\geq 2)$ best populations (Deverman (1969), Carroll, Gupta, and Huang (1975)). Another example in which there are more than two groups (that is, $t_2 < k$) and for which tables are unavailable occurs in the problem of obtaining upper confidence bounds for mean differences in a trend situation; this is discussed more fully in Section 4. We feel that the application of ranking theory to the problem of ranking populations into groups, with special reference to the goal of completely ranking populations ($t_1=1$, $t_2=2$, ...), has suffered due to a lack of tables and even methods for computing (1.1). We discuss below why standard numerical computing techniques are not

really applicable to this problem, and propose an approximation algorithm.

The key feature is that the ranking probability (1.1) is a multidimensional iterated integral. In general, even in the special case where the random variables have normal distributions with a common known variance, the computation of iterated integrals is difficult and time consuming. We have found in the course of our study that simply combining Gaussian quadrative with Simpson's rule to compute probabilities in the complete ranking problem is very expensive for $k \geq 5$. Milton (1972) applied a method based on a multidimensional iterated Simpson's quadrative to the problem, but when applied to a special case of our problem (his section on the probabilities of rank orders when the random variables are i.i.d.) it seems that the number of computer evaluations and the time involved increases rapidly (possibly exponentially) as k increases. The tetrachoric series can sometimes be used to compute multivariate probability integrals but is generally unattractive for $k \geq 4$ (see Gupta (1963 a,b) for a discussion). Dutt (1973) transforms this series into a finite sum of multivariate Fourier transforms and then uses Gaussian quadrature but quadrature becomes increasingly difficult as k increases (see Milton (1966)). None of these approaches extends readily to non-normal distributions.

Somerville (1970)(1971) described a method for estimating the percentage point of a distribution. The important idea in his paper, which is used here to develop our algorithm, is to approximate a continuous distribution by that of a discrete random variable with r

mass points, each of mass r^{-1}. Because our algorithm is iterative in nature the computation becomes quite feasible. In his paper, Somerville (1970) shows that his method is often more efficient than Monte-Carlo methods with the same accuracy and conjectures that the error is $O(r^{-1})$. We will be able to show that the error in our algorithm for approximating (1.1) is no worse than kr^{-1}, and that in practice it is actually considerably smaller.

In Section 2 it is shown that in location and scale parameter families the discrete approximation algorithm given here admits a simple iterative counting representation. This representation has a number of appealing features:

(i) Its iterative nature allows the construction of fairly extensive tables.

(ii) Repeated evaluations of or approximations to distribution functions or density functions are not necessary.

(iii) The algorithm can be applied to any continuous distribution.

In Section 3, the algorithm is applied to the problem of completely ranking normal means and variances; it is shown to be surprisingly accurate in a number of cases. A method for using the algorithm in the general case is also given.

In Section 4, we discuss in detail the application of the tables and the algorithm to ranking problems and the computation of upper confidence bounds for mean differences in a trend situation. An application of this complete ranking problem to drug trials is given.

METHOD AND ERROR TERM

We will first restrict our attention to the case where the random variables $X_1,\ldots,X_k$ have continuous distribution functions $F(x-\theta_1),\ldots,F(x-\theta_k)$, where

$$(2.1)\quad \theta_1 = \ldots \theta_{t_1} = \theta_{t_1+1} - \delta = \ldots = \theta_{t_2} - \delta = \ldots = \theta_k - s\delta \quad (\delta>0) .$$

This is the equally - spaced means case; the more general case of non-equal spacing as well as the scale parameter case will be considered at the end of this section.

Now, in this location case under (2.1), the ranking probability (1.1) becomes

$$(2.2)\qquad \int \cdots \int_R dF(x_1) \ldots dF(x_k) ,$$

where $R = \{(x_1,\ldots,x_{t_1}) \le (x_{t_1+1},\ldots,x_{t_2}) + \delta \le \ldots \le (x_{t_s+1},\ldots,x_k) + s\delta\}$. This is an iterated integral, and the discrete approximation will become an iterated sum.

Choose (for $j = 0,1,\ldots,r$) a_j such that $F(a_j) = j/r$. Let $a_{j-1} \le y_j \le a_j$ $(j = 1,\ldots,r)$, and define G_r to be a distribution function with mass points $y_1,\ldots,y_r$ each of probability $1/r$. Since G_r converges to F as $r \to \infty$, and in fact $\sup_x |G_r(x) - F(x)| \le 1/r$, an approximation to (2.2) becomes

$$(2.3)\qquad \sum_R r^{-k} = B_r(\delta),$$

where the sum is taken over all possible k-truples of the mass points $y_1,\ldots,r_r$ of G_r which are elements of R.

Theorem 1 (Somerville (1970)):

$$\lim_{r\to\infty} B_r(\delta) = \int \cdots \int_R dF(x_1) \ldots dF(x_k) .$$

Theorem 2:

$$|B_r(\delta) - \int \cdots \int_R dF(x_1) \ldots dF(x_k)| \leq kr^{-1} .$$

Proof: See the appendix.

In the next section, it will be shown that the bound given above is somewhat conservative. In fact, if $r = 100$ and $k = 7$, in the complete ranking case the true error seems to be less than .01. This is not too surprising since Somerville (1970) found that for computing percentage points of one-dimensional statistics, $r = 20$ will usually yield an error of less than .01. Another point of interest is that if one defines y_i to satisfy $F(y_i) - F(a_{i-1}) = F(a_i) - F(y_i)$ then the bound in Theorem 2 becomes $k(2r)^{-1}$, since $\sup_x |G_r(x) - F(x)| \leq (2r)^{-1}$. However, computing y_i in this way can become time-consuming.

Let $Y_1,\ldots,Y_k$ be independent with distribution G_r. Before we can state the algorithm we need the following definitions:

(2.4) $L(\delta,i)$ = largest integer j such that $y_j \le y_i + \delta$

$$C(q,i,j) = \Pr\{y_i = \min(Y_1,\ldots,Y_q) \le \max(Y_1,\ldots,Y_q) \le y_j\}$$

$$= \left(\frac{j-i+1}{r}\right)^q - \left(\frac{j-i}{r}\right)^q \quad 1 \le i \le j \le n.$$

Note that if $q = 1$, $C(q,i,j) = r^{-1}$.

The algorithm is contained in the following result, the proof of which is given in the appendix.

Theorem 3: Let $Y_1,\ldots,Y_k$ be independent with distribution function G_r. Define for $\ell = 2,3,\ldots$

(2.5) $$P_i^2(\delta) = \Pr\{(Y_1,\ldots,Y_{t_1}) \le a_i + \delta\}$$

$$P_i^{\ell+1}(\delta) = \Pr\{(Y_1,\ldots,Y_{t_1}) \le (Y_{t_1+1},\ldots,Y_{t_2}) + \delta$$

$$\le \ldots \le (Y_{t_{\ell-1}+1},\ldots,Y_{t_\ell}) + (\ell-1)\delta \le a_i + \ell\delta\}.$$

Then

(2.6) $$P_i^2(\delta) = (L(\delta,i)/r)^{t_1}$$

$$P_i^{\ell+1}(\delta) = \sum_{j=1}^{L(\delta,i)} P_j^{\ell}(\delta)\ C(t_\ell - t_{\ell-1}, j, L(\delta,i))$$

and the ranking probability (1.1) is given by

$$\sum_{i=1}^{r} P_i^s(\delta)C(t_k-t_s,i,r) = P_r^{s+1}(\delta)$$

Note the iterative nature of the algorithm. If, for example, one wants to use the algorithm to compute $\Pr\{X_1 \leq X_2 \leq \ldots \leq X_5\}$ where $X_1,\ldots,X_5$ are normal random variables with common unit variance and means $\mu_1,\ldots,\mu_5$ satisfying $\mu_2-\mu_1 = \mu_3-\mu_2 = \ldots = \mu_5-\mu_4 = \delta$, one obtains in the process $\Pr\{X_1 \leq X_2\}$, $\Pr\{X_1 \leq X_2 \leq X_3\}$ and $\Pr\{X_1 \leq X_2 \leq X_3 \leq X_4\}$ in the terms $P_r^3(\delta)$, $P_r^4(\delta)$ and $P_r^5(\delta)$.

It is obvious that most of the time to be used in the implementation of the iterative algorithm in Theorem 3 revolves around the use of the functions $L(\delta,i)$ and $C(q,j,L(\delta,i))$. Fortunately a number of time-saving features are available. Suppose one wants to use the algorithm with a number of different δ values, say $\delta_0, \delta_1,\ldots, \delta_n$; this will be true if one is attempting to compute percentage points. For a given j there is a j^* such that $L(\delta_j,j^*) = r$ and as δ_j increases the threshhold value j^* decreases. This means that

(i) If $L(\delta_j,i) = r$, then $L(\delta_j,i^*) = \ldots = L(\delta_n,i^*) = r$ if $i^* \geq i$.

(ii) If $L(\delta_j,i) = r$, $P_{i^*}^{\ell}(\delta_j) = P_r^{\ell}(\delta_j)$ if $i^* \geq i$

Thus, as δ increases, it becomes less difficult and time consuming to compute $L(\delta,i)$ and $P_i^{\ell}(\delta)$.

The same is true of the function $C(q,j,L(\delta,i))$ because it is a function of q and $L(\delta,i) - j$ *alone*. Hence, for a given q, one need only an easily computed array of size r to have all the

possible values of $C(q,j,L(\delta,i))$ available. Note that in the important special case of probabilities of complete rankings, $C(1,j,L(\delta,i)) = r^{-1}$ if $j \le L(\delta,i)$, so that virtually no time is spent on computing this function.

In the scale parameter case with $F(x) = 0$ if $x \le 0$, $X_1,\ldots,X_k$ have distribution functions $F(x/\theta_1),\ldots,F(x/\theta_k)$ where

$$(2.7)\quad \theta_1 = \ldots = \theta_{t_1} = \theta_{t_1+1}/\Delta = \ldots = \theta_{t_2}/\Delta = \ldots = \theta_k/\Delta^s \quad (\Delta>1),$$

and the region of integration R in (2.2) becomes

$$(2.8)\quad R = \{(x_1,\ldots,x_{t_1}) \le (x_{t_1+1},\ldots,x_{t_2})/\Delta \le \circ\circ\circ \le (x_{t_s+1},\ldots,x_k)/\Delta^s\}.$$

Then Theorems 1 and 2 still hold while Theorem 3 holds if $L(\delta,i)$ is defined as the largest integer j such that $y_j \le y_i\Delta$.

The configuration (2.1) and (2.7) are the *equally-spaced* configurations; while these are important configurations, it is worthwhile to also study the general case. Suppose $\delta(1) \le \delta(2) \le \ldots \le \delta(n)$ $(n \ge s)$ and one wishes to approximate the probability (2.1) in the case

$$(2.9)\quad \theta_1 = \ldots = \theta_{t_1} = \theta_{t_1+1}{}^{-\delta(\beta_1)} = \ldots = \theta_{t_2}{}^{-\delta(\beta_2)} = \ldots = \theta_k{}^{-\delta(\beta_s)},$$

where $\{\beta_1,\ldots,\beta_s\}$ is a subset of $\{1,\ldots,n\}$ and $\beta_1 \le \beta_2 \le \ldots \le \beta_s$. Then Theorem 3 would still hold with

$$P_i^2(\delta,(\beta_1)) = (L(\delta(\beta_1),i)/r)^{t_1}$$

$$P_i^{\ell+1}(\delta(\beta_1),\ldots,\delta(\beta_\ell))$$

$$= \sum_{j=1}^{H(\ell,i)} P_j^{\ell}(\delta(\beta_1),\ldots,\delta(\beta_{\ell-1}))\ C(t_\ell - t_{\ell-1},j,H(\ell,i)) ,$$

where $H(\ell,i) = L(\delta(\beta_\ell) - \delta(\beta_{\ell-1}), i)$. The calculation of $L(\delta(\beta_1),i),\ldots,L(\delta(\beta_s) - \delta(\beta_{s-1}),i)$ can be accomplished quite easily by taking into account the monotonicity of $L(\delta,i)$ mentioned above (see the next section).

Finally, if it were not true that $\delta(\beta_1) \le \delta(\beta_2) \le \ldots \le \delta(\beta_s)$, then even in this case, the algorithm may be modified to yield an approximation to the probability (1.1), with no change in Theorem 3; of course if $y_i + \delta(\beta_j) - \delta(\beta_{j-1}) < y_1$, one should set $L(\delta(\beta_j) - \delta(\beta_{j-1}),i) = 0$.

COMPLETE RANKINGS IN NORMAL POPULATIONS

The algorithm given in Section 3 was used to calculate (1.1) in three special cases on the IBM 360 machine at the University of North Carolina and the CDC 6500 at Purdue University.

CASE I Compute $Pr\{X_1 \le X_2 \le \ldots \le X_k\}$, where $X_1,\ldots,X_k$ are independent normal random variables with common variance $\sigma^2 = 1$ and means given by (2.1)

(3.1) $\theta_i - \theta_{i-1} = \delta(i-1) \qquad (i=2,\ldots,k).$

CASE II Compute $\Pr\{X_1 \le X_2 \le \ldots \le X_k\}$, where $X_1/\theta_1,\ldots,X_k/\theta_k$ are independent chi-square random variables with 10 degrees of freedom and parameter configuration given by (2.7).

CASE III Compute $\Pr\{(X_1,X_2) \le (X_3,X_4) \le \ldots \le (X_{k-1},X_k)\}$, where $X_1,\ldots,X_k$ are independent normal random variables with common variance $\sigma^2 = 1$ and means given by

(3.2) $\theta_1 = \theta_2 = \theta_3-\delta = \theta_4-\delta = \theta_5-2\delta = \theta_6-2\delta \ldots$

The first case, the normal means case, was studied in the greatest detail. In Table 1a, results are given for the parameter configuration (2.1) using $r = 1000$, $k = 2(1)7$ and $\delta = 0.0(.10)4.20$. This table also provides selected approximate running times, including the computation of $L(\delta,i)$; for example, if $\delta = 1.6$ the computation of $\Pr\{X_1 \le X_2+\delta\}$, $\Pr\{X_1 \le X_2+\delta \le X_3+2\delta\}$, ..., $\Pr\{X_1 \le X_2+\delta \le \ldots \le X_{10}+9\delta\}$ took approximately 61 seconds total. In Table 1b, results are given for the parameter configuration (3.1) with $r = 1000$, $k = 2(1)6$, and $\delta = .25(.25)3.75$. The running times in Table 1b are smaller than those of Table 1a because of the addition of certain program-specific modifications, together with the fact that $L(i(i-1)\delta/2)$ is easy to calculate for $i \ge 3$ and $\delta \ge 2.50$. The apparant stability of the tables from $k = 5$ to $k = 6$ is due to the fact that $\mu_6 - \mu_5 = 15\delta$.

The complete results were spot-checked by a Monte-Carlo experiment with 5000 trials, with a maximal disagreement of .008.

For $k = 4$, the results were also checked by computing the probabilities under (2.1) and (3.1) by using a combination of Gauss-Hermite quadrature and Simpson's formula. The maximal difference between our algorithm's results and this quadrature method was .0005; most differences were of the order .0003 and smaller. This brings out an important point; that while the maximum possible error in the algorithm is kr^{-1}, the actual error in this case seems to be at most $k/(10r)$, which is very good.

The values of $y_1,\ldots,y_r$ were obtained by setting $y_1 = a_1$, $y_r = a_r$, $y_j = \frac{a_j+a_{j-1}}{2}$ if $a_j \leq 0$, and $y_j = .75\, a_j + .25\, a_{j-1}$ if $a_j > 0$. By Theorem 3, the results have a maximal possible error for $r = 1000$ of $k(.001)$, and it seems reasonable based on the above discussion to suppose that the true maximal error is $k(.0001)$.

The algorithm was also run in the configuration (2.1) with $r = 100$ and $r = 200$. A point to note is that the maximal difference between the results for $r = 100$ and $r = 1000$ is .007, so that the results for $r = 100$ are correct to the second decimal. This is much better than predicted by Theorem 3. Since the same conclusion holds for the variances problem (see below), choosing $r = 100$ or $r = 200$ will be acceptable in practice to get results correct to the second decimal place. The time for computation decreases appropriately as r decreases.

For the variances case, values of (1.1) for various k, δ, and ν (the degrees of freedom) are available in a set of tables computed by Schafer (1974) with 20,000 Monte-Carlo trials. The algorithm was used when $\nu = 10$, $\delta = 2,3$, $k = 2(1)7$, and $r = 100, 200$. The maximal deviation from these tables for $r = 100$ was .007 and the total running time (including compilation) was 5 seconds, or about .5 seconds per integral. When $r = 200$, the maximal deviation was .004 and the running time (including compilation) was 8 seconds or about .7 seconds per integral.

Case III was considered to see if the excellent performance of the algorithm (both in time and accuracy) in the cases of complete rankings is due to the special behavior of the function $C(q,i,j)$ (recall, this function is a constant in the complete rankings problems). Table 2 presents the results of our study for Case III. The running times are quite comparable to those of the complete ranking case. The results were spot-checked by a Monte-Carlo simulation with 20,000 trials; for $\delta \geq 2.00$, the maximal difference was .002, while for $1.60 \leq \delta < 2.00$, the maximal difference between the algorithm and the simulation experiment was .005.

APPLICATIONS OF THE TABLES

RANKING AND SELECTION: Suppose one has k normal populations $\pi_1, \ldots, \pi_k$ with unknown means $\mu_1, \ldots, \mu_k$ but a common variance σ^2.

The goal will be to find the correct (but unknown) ordering $\mu_{[1]} \leq \mu_{[2]} \leq \ldots \leq \mu_{[k]}$, where $\mu_{[i+1]} - \mu_{[i]}$ is expressible in σ units, i.e., either

$$(4.1) \qquad \mu_{[i+1]} - \mu_{[i]} \geq \Delta\sigma > 0, \quad i = 1, \ldots, k-1, \quad \text{or}$$

$$(4.2) \qquad \mu_{[i+1]} - \mu_{[i]} \geq i\Delta\sigma > 0, \quad i = 1, \ldots, k-1 .$$

Tables 1a and 1b may be utilized. One may use the work of Eaton (1966) or Savage (1957) to show that the "most likely" ordering of the sample averages $\bar{X}_1, \ldots, \bar{X}_k$ based on n observations from each of $\pi_1, \ldots, \pi_k$ is $\bar{X}_1 < \bar{X}_2 < \ldots < \bar{X}_k$ if $\mu_1 < \mu_2 < \ldots < \mu_k$, so that the best procedure is to rank the populations according to the observed rankings of the sample means. Table 1a presents as a function of $\delta = \sqrt{n}\,\Delta$ an exact lower bound on the probability of making a correct decision under the configuration (4.1), while Table 1b may be utilized for configuration (4.2). One chooses the sample size $n(\Delta)$ as the samllest integer n for which the lower bound given in the tables as a function of $\sqrt{n}\,\Delta$ is at least a value P^*. If $\mu_{[i+1]} - \mu_{[i]}$ is not expressible in σ units (i.e., $\mu_{[i+1]} - \mu_{[i]} \geq \Delta$ in equation (4.1)) and σ is known, the tables present the probabilities of making a correct decision as a function of $\delta = \sqrt{n}\,\Delta/\sigma$. If σ is unknown and $\mu_{[i+1]} - \mu_{[i]}$ is not expressible in σ units, one must either estimate σ or use a sequential procedure.

If the distributions of the populations $\pi_1, \ldots, \pi_k$ are not normal but there is a constant σ such that $\frac{\sqrt{n}}{\sigma}(\bar{X}_i - \mu_i)$ is

asymptotically normally distributed, then for n sufficiently large the results of this section are applicable.

We now show how Table 1a might be used in a drug-testing situation. In Table 3 data (collected for the National Cancer Institute by the Arthur D. Little Company) are given for the weight loss (in grams) of mice inoculated with a certain type of tumor. The mice have been given the same drug at five different dosage levels but are all on the same schedule; the number of mice in each group is eight and the pooled sample standard deviation was $\hat{\sigma} = 19.5$. To make a correct complete ranking at least 75% of the time under configuration (4.1) requires $\sqrt{n}\,\Delta \geq 2.134$ (by interpolating Table 1a), so that $\Delta \approx .754$ must be assumed and $\Delta\sigma \approx 14.7$. Thus, in order to guarantee that one makes a correct decision 75% of the time, one must assume that the means are spaced by an amount at least as large as 14.7; the data may cast some doubt on this assumption. If one were willing to assume $\Delta\sigma \approx 8.00$ then one needs $\Delta \approx 8/19.5$ and hence $\sqrt{n} \geq 19.5(2.134)/8$, which means that one should have taken at least $n = 28$ observations at each dosage level.

CONFIDENCE INTERVALS FOR CONTRASTS. Consider the problem of the previous subsection with the common variance σ^2 assumed for the moment to be known. In a trend situation where it is known that $\mu_1 \leq \mu_2 \leq \ldots \leq \mu_k$, it would be of interest to obtain upper confidence intervals on the amount of trend, i.e., on $\mu_2 - \mu_1, \mu_3 - \mu_2, \ldots, \mu_k - \mu_{k-1}$. It may be pointed out that Scheffé (1953)

and Tukey (1953) have considered the problem of simultaneous confidence bounds on the contrasts of unordered normal means. Table 1a may be used to find confidence intervals with coverage probability $(1-\alpha)$ of the form

$$\mu_2 - \mu_1 \le \bar{X}_2 - \bar{X}_1 + \delta\sigma/\sqrt{n}$$

$$\mu_3 - \mu_2 \le \bar{X}_3 - \bar{X}_2 + \delta\sigma/\sqrt{n}$$

$$\vdots$$

$$\mu_k - \mu_{k-1} \le \bar{X}_k - \bar{X}_{k-1} + \delta\sigma/\sqrt{n} \ ,$$

where $\delta > 0$ is read from Table 1a. Again, if σ^2 is unknown, it may be replaced by an estimate (with a small loss in coverage probability).

ACKNOWLEDGEMENT

The authors wish to thank Carol de Branges of Purdue University for computing the ranking probabilities for $k = 4$ using Gauss-Hermite quadrature and Simpson's rule. We also thank Dr. John M. Venditti of the National Cancer Institute and Sandra H. Stone of Arthur D. Little, Inc. for providing the data for the drug example of Section 4.

BIBLIOGRAPHY

[1] BECHHOFER, R.E., (1954). A single-sample multiple decision procedure for ranking means of normal populations with known variances. *Ann. Math. Statist.* (25) 16-39.

[2] CARROLL, R.J., GUPTA, S.S. and HUANG, D.T. (1975). Selection procedures for the t best populations. To appear in *Comm. Statist.*

[3] DEVERMAN, J.N. (1969). A general selection procedure relative to the t best populations. Ph.D. Thesis, Purdue University.

[4] DEVERMAN, J.N. and GUPTA, S.S. (1969). On a selection procedure concerning the t best populations. *Ann. Math. Statist.* (40) 1870 (abstract).

[5] DUTT, J.E., (1973). A representation of multivariate normal probability integrals by integral transforms. *Biometrika* (60) 637-645.

[6] EATON, M.L. (1966). Some optimum properties of ranking procedures. *Ann. Math. Statist.* (37) 124-137.

[7] GUPTA, S.S. (1963a). Probability integrals of multivariate normal and multivariate t. *Ann. Math. Statist.* (34) 792-828.

[8] GUPTA, S.S. (1963b). Bibliography on the multivariate normal integrals and related topics. *Ann. Math. Statist.* (34) 829-838.

[9] GUPTA, S.S. (1965). On some multiple decision (selection and ranking) rules. *Technometrics*, (7) 225-245.

[10] MILTON, R.C., (1970). *Rank Order Probabilities: Two Sample Normal Shift Alternatives.* John Wiley and Sons, New York.

[11] MILTON, R.C., (1972). Computer evaluation of the multivariate normal integral. *Technometrics* (14) 881-889.

[12] SAVAGE, I.RICHARD, (1957). Contributions to the theory of rank order statistics in the "trend" case. *Ann. Math. Statist.* (28) 968-977.

[13] SCHAFER, R.E., (1974). A single-sample complete ordering procedure for certain populations. In *Reliability and Biometry: Statistical Analysis of Lifelength*, eds. F. Proschan and R. J. Serfling, SIAM 1974.

[14] SCHEFFÉ, H. (1953). A method for judging all contrasts in the analysis of variance. *Biometrika* (40) 87-104.

[15] SOMERVILLE, P.N., (1970). A technique for computation of percentage points of a statistic. *Technometrics* (12) 378-382.

[16] SOMERVILLE, P.N., (1971). A technique for obtaining probabilities of correct selection in a two-stage selection problem. *Biometrika* (58) 615-623.

[17] TUKEY, J.W. (1953). The problem of multiple comparisons. Unpublished dittoed manuscript, Princeton University.

TABLE 1a

Approximations to $\Pr\{X_1 < X_2 < \ldots < X_k\}$ for $X_1, \ldots, X_k$ normal random variables with common variance $\sigma^2 = 1$ and means $\mu_1, \ldots, \mu_k$, where $\mu_{i+1} - \mu_i = \delta$.

Table 1a

δ \ k	2	3	4	5	6	7	8	9	10	Time in Seconds
0.00	.500	.167	.041	.008	.001	.000	.000	.000	.000	
0.10	.528	.196	.056	.014	.003	.001	.000	.000	.000	53
0.20	.556	.228	.077	.023	.006	.002	.000	.000	.000	
0.30	.584	.263	.101	.036	.012	.004	.001	.000	.000	57
0.40	.611	.299	.130	.052	.020	.008	.003	.001	.000	
0.50	.638	.337	.162	.074	.033	.014	.006	.003	.001	63
0.60	.664	.376	.192	.100	.050	.025	.012	.006	.003	
0.70	.690	.416	.237	.132	.073	.040	.022	.012	.006	63
0.80	.714	.456	.279	.168	.101	.060	.036	.021	.013	
0.90	.738	.496	.324	.208	.134	.086	.055	.035	.022	64
1.00	.760	.536	.369	.252	.172	.117	.080	.054	.037	
1.10	.782	.574	.415	.298	.214	.154	.110	.079	.057	63
1.20	.802	.612	.461	.346	.260	.195	.146	.110	.082	
1.30	.821	.647	.506	.395	.308	.240	.187	.146	.114	58
1.40	.839	.681	.550	.444	.358	.288	.232	.187	.151	
1.50	.855	.714	.593	.492	.408	.338	.281	.233	.193	

Table 1a (continued)

δ\k	2	3	4	5	6	7	8	9	10	Time in Seconds
1.60	.871	.744	.633	.539	.458	.390	.332	.282	.240	61
1.70	.885	.772	.671	.584	.507	.441	.384	.334	.290	
1.80	.898	.797	.707	.626	.555	.492	.436	.386	.342	63
1.90	.910	.821	.740	.667	.601	.541	.488	.439	.396	
2.00	.921	.843	.770	.704	.644	.589	.538	.492	.450	58
2.10	.931	.862	.798	.739	.684	.633	.586	.543	.503	
2.20	.940	.880	.824	.771	.722	.676	.632	.592	.554	56
2.30	.948	.856	.847	.800	.756	.715	.675	.638	.603	
2.40	.955	.910	.867	.826	.788	.750	.715	.681	.649	51
2.50	.961	.923	.886	.850	.816	.783	.752	.721	.692	
2.60	.967	.934	.902	.871	.841	.812	.785	.758	.732	47
2.70	.972	.944	.916	.890	.864	.839	.815	.791	.768	
2.80	.976	.952	.929	.906	.884	.863	.842	.821	.801	41
2.90	.978	.960	.940	.921	.902	.883	.865	.847	.830	
3.00	.983	.966	.950	.933	.917	.901	.886	.871	.856	36

Table 1a continued

δ\k	2	3	4	5	6	7	8	9	10	Time in Seconds
3.10	.936	.972	.958	.944	.931	.917	.904	.891	.878	31
3.20	.988	.977	.965	.953	.942	.931	.920	.909	.898	
3.30	.990	.981	.971	.961	.952	.942	.933	.924	.915	26
3.40	.992	.984	.976	.968	.960	.953	.945	.937	.930	
3.50	.993	.987	.980	.974	.968	.961	.955	.948	.942	21
3.60	.995	.989	.984	.979	.973	.968	.963	.958	.953	
3.70	.996	.991	.987	.983	.978	.974	.970	.966	.962	16
3.80	.997	.993	.990	.986	.983	.979	.976	.972	.969	
3.90	.997	.994	.991	.989	.986	.983	.981	.978	.975	13
4.00	.998	.996	.993	.991	.989	.987	.985	.982	.980	
4.10	.998	.996	.995	.993	.991	.990	.988	.986	.984	9
4.20	.999	.997	.996	.994	.993	.992	.990	.989	.988	
4.30	.999	.998	.997	.996	.995	.994	.992	.991	.990	8
4.40	.999	.998	.997	.997	.996	.995	.994	.993	.992	

TABLE 1b

Approximations to $Pr(X_1 < X_2 < \ldots < X_k)$ for $X_1, \ldots, X_k$ normal random variables with common variance $\sigma^2 = 1$ and means $\mu_1, \ldots, \mu_k$, where $\mu_{i+1} - \mu_i = \delta i(i+1)/2$.

Table 1b

δ \ k	2	3	4	5	6	Time in Seconds
.25	.570	.289	.158	.101	.074	31
.50	.638	.430	.344	.311	.289	33
.75	.702	.567	.529	.519	.517	26
1.00	.760	.684	.671	.699	.669	17
1.25	.811	.773	.770	.770	.770	14
1.50	.885	.839	.837	.837	.837	9
1.75	.892	.885	.885	.885	.885	4
2.00	.921	.917	.917	.917	.917	3
2.25	.944	.943	.943	.943	.943	1.0
2.50	.961	.961	.961	.961	.961	.36
2.75	.974	.974	.974	.974	.974	.18
3.00	.983	.983	.983	.983	.983	.09
3.25	.989	.989	.989	.989	.989	.03
3.50	.993	.993	.993	.993	.993	.04
3.75	.996	.996	.996	.996	.996	.04

TABLE 2

Approximations to $\Pr((X_1,X_2) < (X_3,X_4) < \ldots < (X_{k-1},X_k))$, where $X_1, \ldots, X_k$ are normal random variables with common variance $\sigma^2 = 1$ and means $\mu_1, \ldots, \mu_k$, with $\mu_1 = \mu_2$, $\mu_3 = \mu_4$, $\ldots$, $\mu_{k-1} = \mu_k$ and $\mu_3 - \mu_2 = \delta$, $\mu_5 - \mu_4 = \delta$, etc.

Table 2

$\delta \backslash^{k}$	4	6	8	10	Time in Seconds
1.60	.662	.411	.253	.155	50
1.80	.721	.499	.344	.237	50
2.00	.775	.585	.441	.332	49
2.20	.821	.664	.537	.433	47
2.40	.861	.735	.627	.535	44
2.60	.894	.796	.708	.630	41
2.80	.921	.847	.778	.715	37
3.00	.943	.887	.835	.786	33
3.20	.959	.919	.881	.844	28
3.40	.972	.943	.916	.890	23
3.60	.981	.961	.943	.924	19
3.80	.987	.974	.962	.949	15
4.00	.992	.983	.975	.967	12
4.20	.995	.989	.984	.979	9
4.40	.997	.994	.990	.987	6
4.60	.998	.996	.994	.992	4
4.80	.999	.998	.997	.996	3

TABLE 3

The weight loss in grams of mice inoculated with a certain type of tumor.

Drug Level	Average Weight Loss
A	23.25
B	46.25
C	56.25
D	29.25
E	11.00

APPENDIX

This appendix presents the proofs of Theorems 2 and 3. A general induction proof for Theorem 2 is possible but very messy; in order to illustrate the method, the proof is given only for the special case $k = 6$, $t_1 = t_2 = 2$. Define $a \wedge b = \min\{a,b\}$. Now, let $\varepsilon > 0$ be arbitrary, and let G^* be a continuous distribution function with $\sup\{|G^*(x) - G_r(x)|: -\infty < x < \infty\} \leq \varepsilon$. Then $\sup\{|F(x) - G^*(x)|: -\infty < x < \infty\} \leq r(\varepsilon) = r^{-1} + \varepsilon$. Since the joint distribution of $X_1, \ldots, X_k$ under G^* converges (as $\varepsilon \to 0$) weakly to the joint distribution of $X_1, \ldots, X_k$ under G, and since R is convex, the probability of the event R and under G^* converges to the probability of the event R under G. Thus, to show Theorem 2, it suffices to prove that

$$|P\{R|F\} - P\{R|G^*\}| \leq \mathbf{r}(\varepsilon). \tag{A1}$$

Let $F_i = F(x_i)$ and $G_i^* = G^*(x_i)$. Then, the error term (A1) is

$$= \left| \int_{-\infty}^{\infty} \int_{-\infty}^{x_5 \wedge x_6 + \delta} \int_{-\infty}^{x_3 \wedge x_4 + \delta} dF_1 \ldots dF_6 - dG_1^* \ldots dG_6^* \right|$$

$$\leq \left| \int_{-\infty}^{\infty} \int_{-\infty}^{x_5 \wedge x_6 + \delta} \int_{-\infty}^{x_3 \wedge x_4 + \delta} (dF_1 dF_2 - dG_1^* dG_2^*) dF_3 \ldots dF_6 \right|$$

$$+ \left| \int_{-\infty}^{\infty} \int_{-\infty}^{x_5 \wedge x_6 + \delta} \int_{-\infty}^{x_3 \wedge x_4 + \delta} dG_1^* dG_2^* (dF_3 \ldots dF_6 - dG_3^* \ldots dG_6^*) \right|.$$

By changing the order of integration,

$$\leq \Big| \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \int_{x_k \vee x_2 - \delta}^{\infty}\int_{x_3 \vee x_4 - \delta}^{\infty} dF_6 \ldots dF_3 (dF_2 dF_1 - dG_2^* dG_1^*) \Big|$$

$$+ \Big| \int_{-\infty}^{\infty}\int_{-\infty}^{x_5 \wedge x_6 + \delta} H(x_3, x_4, \delta)(dF_3 \ldots dF_6 - dG_3^* \ldots dG_6^*) \Big|,$$

where H is a monotone continuous function bounded by one.

Let $a \vee b = \max\{a,b\}$. Then,

$$\leq \Big| \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \int_{x_1 \vee x_2 - \delta}^{\infty}\int_{x_3 \vee x_4 - \delta}^{\infty} dF_6 \ldots dF_3 dG_1^* (dF_2 - dG_2^*) \Big|$$

$$+ \Big| \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \int_{x_1 \vee x_2 - \delta}^{\infty}\int_{x_3 \vee x_4 - \delta}^{\infty} dF_6 \ldots dF_3 dF_2 (dF_1 - dG_1^*) \Big|$$

$$+ \Big| \int_{-\infty}^{\infty}\int_{-\infty}^{x_5 \wedge x_6 + \delta} H(x_3, x_4, \delta)(dF_3 \ldots dF_6 - dG_3^* \ldots dG_6^*) \Big| .$$

Since G_1^*, G_2^* are continuous in x_1, x_2, each of the first two terms are bounded by $r(\varepsilon)$, so the term (A1) is

$$\leq 2r(\varepsilon) + \Big| \int_{-\infty}^{\infty}\int_{-\infty}^{x_5 \wedge x_6 + \delta} H(x_3, x_4, \delta)(dF_3 \ldots dF_6 - dG_3^* \ldots dG_6^*) \Big|$$

$$\leq 4r(\varepsilon) + \Big| \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} H^*(x_5, x_6, \delta)(dF_5 dF_6 - dG_5^* dG_6^*) \Big| \leq 6r(\varepsilon),$$

the last two inequalities following in a manner similar to the above.

Proof of Theorem 3: That $P_i^2(\delta) = (L(\delta,i)/r)^{t_1}$ follows from the independence of $(Y_1,\ldots,Y_{t_1})$. Now, by induction

$$P_i^{\ell+1}(\delta) = \sum_{j=1}^{L(\delta,i)} P_j^{\ell}(\delta)\ \Pr\{a_j=\min(Y_{t_{\ell-1}+1},\ldots,Y_{t_\ell}) \le \max(Y_{t_{\ell-1}+1},\ldots,Y_{t_\ell}) \le a_i + \delta\}$$

$$= \sum_{j=1}^{L(\delta,i)} P_j^{\ell}(\delta)\ \Pr\{a_j=\min(Y_{t_{\ell-1}+1},Y_{t_\ell}) \le \max(Y_{t_{\ell-1}+1},\ldots,Y_{t_\ell}) \le L(\delta,i)\}$$

$$= \sum_{j=1}^{L(\delta,i)} P_j^{\ell}(\delta)\ C(t_\ell - t_{\ell-1},\ j,\ L(\delta,i))\ ,$$

showing the first part of (2.6). Then

$$P\{(Y_1,\ldots,Y_{t_1}) \le \ldots \le (Y_{t_{\ell}+1},\ldots,Y_{t_{\ell+1}}) + \ell\delta\}$$

$$= \sum_{i=1}^{r} P_i^{\ell+1}(\delta)\ \Pr\{a_i=\min\{Y_{t_{\ell}+1},\ldots,Y_{t_{\ell+1}}\}\}$$

$$= \sum_{i=1}^{r} P_i^{\ell+1}(\delta)\ C(t_{\ell+1} - t_{\ell},\ i,\ r)\ .$$