On the Conditional Distribution of a Statistic that Arises in Tests of Homogeneity and Selection and Ranking of Binomial Populations*

by

Shanti S. Gupta
Purdue University

and

Wen-Tao Huang
Institute of Mathematics
Academia Sinica

Department of Statistics
Division of Mathematical Sciences
Mimeograph Series #329
August 1973

*This research was supported by the Office of Naval Research under Contract N00014-67-A-0226-00014 at Purdue University.

On the Conditional Distribution of a
Statistic that Arises in Tests of Homogeneity and
Selection and Ranking of Binomial Populations*

by

Shanti S. Gupta
Purdue University

and

Wen-Tao Huang
Institute of Mathematics
Academia Sinica

1. Introduction and Summary

Let $\pi_1, \pi_2, \ldots, \pi_k$ be k binomial populations such that X_i, the random observation from π_i, has density $b(n;p_i)$ for $i=1,2,\ldots,k$. In some practical situations one wishes to know whether p_i are significantly different or not. This is the problem of a test of homogeneity of the binomial populations. For this problem, Siotani [6] studied the distribution of the range $\max_{1\le i\le k} X_i - \min_{1\le j\le k} X_j$ considering all p_i to have a common known p. He used this range as a criterion to judge whether p_i are all equal to a known p. When the value of the range is large, the hypothesis of homogeneity is rejected. Motivated by this test problem, we are interested in deriving the conditional distribution of $\max_{1\le i\le k} X_i - X_j$ given that $\sum_{i=1}^{k} X_i = T$, where j is some fixed integer between 1 and k. It is shown that the distribution of this statistic is independent of the common value p. We study a test, based on this statistic, which is conservative. This test does not depend on the

*This research was supported by the Office of Naval Research under Contract N00014-67-A-0226-00014 at Purdue University. Reproduction in whole or part is permitted for any purpose of the United States Government.

common value of p which is usually not assumed to be known. The test is conservative in the sense that its size is less than or equal to the level of significance α. However, for k=2, it is an exact test. Some comparisons with Siotani's test [6] are made. The conditional distribution of the statistic is derived in Section 2 and tables for the percentage points of the above statistic are constructed. The moments of the statistic are also discussed (Section 3). Application to the selection and ranking problem is discussed in Section 5. In Section 6, the computations of the values of the percentage points or critical values of the test are discussed. Tables of values for the critical region for the test of homogeneity are also given.

2. Definitions and Notations

Let $[\alpha]$ ($\langle\alpha\rangle$) denote the largest (smallest) integer $\leq$ ($\geq$) α.

Define

(2.1) $A(k,\alpha;t,n) = \Sigma\binom{n}{r_1}\binom{n}{r_2}\cdots\binom{n}{r_k}$ where the summation is over the set of all non-negative integers r_i such that $\sum_{i=1}^{k} r_i = t$ and $\max_{1\leq i\leq k} r_i = \alpha$ for some integer α.

(2.2) $N(k,c;t,n) = \Sigma\binom{n}{s_1}\binom{n}{s_2}\cdots\binom{n}{s_k}$ where the summation is over the set of all non-negative integers s_i such that $\sum_{i=1}^{k} s_i = t$ and $s_k \geq \max_{1\leq j\leq k-1} s_j - c$ for some constant c.

(2.3) $M(k;c,t,n) = \Sigma\binom{n}{t_1}\binom{n}{t_2}\cdots\binom{n}{t_k}$ where the summation is over all non-negative integer t_i such that $\sum_{i=1}^{k} t_i = t$ and $t_k = \max_j t_j - c$.

When any of the summations is over an empty set, the sum will be defined as zero.

Lemma 2.1. Let k and n be any two positive integers. Suppose t and c are non-negative integers such that $0 \le t \le kn$ and $0 \le c \le \min(t,n)$. Then the following statements hold.

(a) $$N(k;c,t,n) = \sum_{i=I_1}^{I_2} A(k-1,\ i+c;\ t-i,\ n)\ \binom{n}{i},$$
$$= 0 \quad \text{if } I_1 > I_2$$

where $I_1 = \max(0, \langle (t-(k-1)c)/k \rangle)$ and $I_2 = \min(n,t)$. In particular, for k=2, $N(2;c,t,n) = \sum_{j=j_0}^{j_1} \binom{n}{j}\binom{n}{t-j}$ where $j_0 = \langle \frac{t-c}{2} \rangle$, $j_1 = \min(t,n)$.

(b) If $\langle \frac{t+c}{k} \rangle \le \min(t,n)$
$$M(k;c,t,n) = \sum_{j=J_1}^{J_2} A(k-1,j;t-j+c,n)\ \binom{n}{j-c}$$
$$= 0 \quad \text{otherwise},$$

where $J_1 = \langle \frac{t+c}{k} \rangle$ and $J_2 = \min(t,n,\ [\frac{t+c}{2}])$. For k=2,

$$M(2;c,t,n) = 0 \qquad \text{for } t+c \text{ odd}$$
$$= \binom{n}{\frac{t+c}{2}}\binom{n}{\frac{t-c}{2}} \qquad \text{for } t+c \text{ even.}$$

(c) $$A(k,\alpha;t,n) = \sum_{i=I_1}^{I_2} A(k-1,\alpha;t-i,n)\ \binom{n}{i} \text{ where}$$

$I_1 = \max(0, t-(k-1)\alpha, t-(k-1)n)$ and

$I_2 = \min(n,t,\alpha)$.

For k=2, $A(2,\alpha;t,n) = \sum_{i=I_3}^{I_4} \binom{n}{i}\binom{n}{t-i}$ where

$I_3 = \max(0, t-\alpha, t-n)$

$I_4 = \min(n,t,\alpha)$.

Proof:

(a) By the definition (2.2), fix $s_k=\alpha$. For the evaluation of $N(k,c;t,n)$ we have $s_i \leq \alpha+c$ for $i=1,2,\ldots,k-1$ and $\sum_{i=1}^{k-1} s_i = t-\alpha$. Therefore $t-\alpha \leq (k-1)(\alpha+c)$. This leads to $k\alpha \geq t-(k-1)c$. Now (a) follows from (2.1), noting that $0 \leq \alpha \leq \min(t,n)$.

When $k=2$, fix $s_2=j$. Then, $j \geq t-j-c$, or equivalently, $j \geq <\frac{t-c}{2}> \geq 0$. Since $j \leq \min(t,n)$, the result follows.

(b) According to the definition (2.3), let $\max_{1\leq j\leq k-1} t_j=\alpha$, then $t_k = \alpha-c$ and $\sum_{i=1}^{k-1} t_i = t-(\alpha-c)$. By definition (2.2) we have $M(k;c,t,n) = A(k-1,\alpha;t-\alpha+c,n)\binom{n}{\alpha-c}$. To find the range of α, we see that $t-\alpha+c \leq (k-1)\alpha$ or, equivalently, $\alpha \geq <(t+c)/k>$. Since $\alpha \leq n$ and $\alpha \leq t$, we have thus $\alpha \geq <\frac{t+c}{k}>$ in case $<\frac{t+c}{k}> \leq \min(n,t)$. If $<\frac{t+c}{k}> > \min(n,t)$, $M(k;c,t,n) = 0$, i.e. the summation in (2.3) is over an empty set. On the other hand, α attains its maximum when $t_i=\alpha$, $t_j=0$ and $t_k=\alpha-c$ for all $j\neq i,k$. Under this case we have $\alpha+(\alpha-c) \leq t$ or $\alpha \leq [\frac{t+c}{2}]$. Since α cannot exceed n and t, we have thus $\alpha \leq \min(n,t,[\frac{t+c}{2}])$.

When $k=2$, we see that $t_1=\alpha$ and $t_2=\alpha-c$. Hence $\alpha+(\alpha-c)=t$ or $2\alpha = t+c$. Hence, if $t+c$ is odd, the equality is impossible and the summation in (2.3) is thus over an empty set. When $t+c$ is even $t_1=\alpha= \frac{t+c}{2}$ and $t_2=\alpha-c= \frac{t-c}{2}$.

(c) To set up a recursive formula for $A(k,\alpha;t,n)$ we fix $r_1=i$. By definition (2.1) we see that $A(k,\alpha;t,n)=A(k-1,\alpha;t-i,n)\binom{n}{i}$. To find the range of i, we note that when $r_2 \leq r_3 \leq\ldots\leq r_k \leq \alpha$, $i \geq t-(k-1)\alpha$. Also, since $r_i \leq n$, we have $i \geq t-(k-1)n$ in case $\alpha > n$. Hence, $i \geq \max(0,t-(k-1)\alpha,t-(k-1)n)$. On the other hand, i never exceeds n, t or α. Hence, $i \le \min(n,t,\alpha)$. When k=2, the result is immediate.

Let X_i denote the observation from π_i, i=1,2,...,k. Define

$$(2.4) \qquad R_i = \max_{1 \le j \le k} X_j - X_i \qquad \text{and}$$

$$(2.5) \qquad \eta(R_i) = \begin{cases} 1 & \text{if } R_i < c(t), \\ 1-\rho & \text{if } R_i = c(t), \quad \text{given } \sum_{i=1}^{k} X_i = t . \\ 0 & \text{if } R_i > c(t), \end{cases}$$

Let $\Omega_0 = \{\omega=(p_1,p_2,\ldots,p_k) | p_1=p_2=\ldots=p_k=p\}$. Then, we have

<u>Theorem 2.1</u> For $\omega \in \Omega_0$ and i=1,2,...,k,

(a) $P_\omega(R_i \le c | \sum_{i=1}^{k} X_i = t) = N(k;c,t,n)/\binom{kn}{t}$ for $0 < p < 1$,

$= 1$ if p=0 or 1.

(b) $E_\omega \eta(R_i) = \{N(k;c,t,n) - \rho M(k;c,t,n)\}/\binom{kn}{t}$ for $0 < p < 1$

$= 1-\rho$ if c=0 and p=0 or 1

$= 1$ if c>0 and p=0 or 1.

Proof:

(a) For $0 < p < 1$, $P_\omega(R_i \le c | \sum_1^k X_j = t) = \dfrac{P(X_i \ge X_{max}-c, \sum_{j=1}^{k} X_j = t)}{P_\omega(\sum_{i=1}^{k} X_i = t)}$

$$= \frac{\Sigma \binom{n}{s_1}\binom{n}{s_2}\cdots\binom{n}{s_k} p^{s_1+s_2+\ldots+s_k}(1-p)^{kn-(s_1+s_2+\ldots+s_k)}}{\binom{kn}{t} p^t (1-p)^{kn-t}}$$

since it is well-known that $T \equiv \sum_{i=1}^{k} X_i$ is binomially distributed with density $b(t;kn,p)$, and where the summation is over all non-negative integer s_i such that $\sum_1^k s_i = t$ and $s_i \geq \max_{j \neq i} s_j - c$. By the definition of (2.2), the result follows. When p=0 or 1, the result is immediate.

(b) For $0 < p < 1$, $E_\omega \eta(R_i) = \dfrac{P(X_i > X_{max} - c | \sum_1^k X_i = t) + (1-\rho) P(X_i = X_{max} - c | \sum_1^k X_i = t)}{P_\omega(\sum_1^k X_i = t)}$

$$= \frac{\Sigma \binom{n}{s_1}\binom{n}{s_2} \cdots \binom{n}{s_k} + (1-\rho)\Sigma \binom{n}{t_1}\binom{n}{t_2} \cdots \binom{n}{t_k}}{\binom{kn}{t}}$$

where the first summation is over all non-negative integer s_j such that $\sum_1^k s_j = t$ and $s_i > \max_{j \neq i} s_j - c$ and the second summation is over all non-negative integer t_j such that $\Sigma t_j = t$ and $t_i = \max_{j \neq i} t_j - c$. By our definitions of (2.2) and (2.3), we have that the first summation equals $\{N(k;c,t,n) - M(k;c,t,n)\}$ and the second summation equals $M(k;c,t,n)$. It follows thus that $E_\omega \eta(R_i) = \{N(k;c,t,n) - \rho\, M(k;c,t,n)\}/\binom{kn}{t}$. When p=0 or 1, the proof is obvious. We note that $P(\max_{1 \leq j \leq k} X_j - X_i \leq c | \sum_{i=1}^{k} X_i = t) = P(\max_{j \neq i} X_j - X_i \leq c | \sum_1^k X_i = t)$ if, and only if, $c \geq 0$. The proof is thus complete.

3. Moments

Let $X_1, X_2, \ldots, X_k$ be k independent random observations from a binomial population with a common density $b(n;p)$. Let $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(k)}$ denote the ordered values of X_i. Gupta and Panchapakesan [3] obtained the explicit forms of the first two moments of $X_{(i)}$. They also derived an explicit form for the covariance of $X_{(1)}$ and $X_{(k)}$. In this section we derive, respectively, the rth moments of conditional distributions of order statistics $X_{(k)}-X_i$ and $X_{(k)}$ given $\sum_1^k X_i=t$ for any fixed $i=1,2,\ldots,k$. These are given in the following

Theorem 3.1. For $r > 0$, $p > 0$,

(a) $$E_p((X_{(k)}-X_i)^r|\sum_1^k X_j=t) = \sum_{j=0}^{n} j^r M(k;j,t,n)/\binom{kn}{t}$$

(b) $$E_p(X_{(k)}^r|\sum_{i=1}^{k} X_i=t) = \sum_{j=0}^{n} j^r A(k,j;t,n)/\binom{kn}{t}$$

where the functions M and A are defined in Lemma 2.1.

Proof:

(a) $$P(X_{(k)}-X_i=j|\sum_1^k X_r=t) = \frac{\Sigma\binom{n}{t_1}\binom{n}{t_2}\cdots\binom{n}{t_k}}{\binom{kn}{t}}$$

where the summation is over all non-negative t_j such that $t_i=\max_r t_r-j$ and $\sum_1^k t_r=t$. From the definition (2.3) we then have

$P(X_{(k)}-X_i=j|\sum_1^k X_r=t)=M(k;j,t,n)/\binom{kn}{t}$ and the result is thus immediate.

(b) We note that $P(X_{(k)}=j|\sum_1^k X_i=t) = \dfrac{\Sigma\binom{n}{r_1}\binom{n}{r_2}\cdots\binom{n}{r_k}}{\binom{kn}{t}}$ where the summation is over all non-negative r_i such that $\max r_i=j$ and $\Sigma r_i=t$. Thus from the definition in (2.1) we have

$P(X_{(k)}=j|\sum_1^k X_i=t) = A(k,j;t,n)/\binom{kn}{t}$. The result thus follows.

We note that both moments are independent of p.

4. Test of Homogeneity; Comparison with Siotani's Test

To test the homogeneity of k experiments, i.e. to test $H: p_1=p_2=\ldots=p_k$ against the alternative H_A: not H, we consider the following test $\phi_1(T)$, which as usual denotes the probability of rejection and which is given by

$$(4.1)\quad \phi_1(T) = \begin{cases} 1 & \text{if } X_i < \max_{1\leq j\leq k} X_j - c(t) \text{ for at least one } i \\ 0 & \text{otherwise} \end{cases}$$

given that $T \equiv \sum_{i=1}^{k} X_i = t$.

It should be noted that the test can also be written as $\phi_1(T)=1$ if $\max_i X_i - \min_j X_j > c(t)$ and $\phi_1(T)=0$ otherwise, given that $T=t$.

Let $Y_i=n-X_i$ for $i=1,2,\ldots,k$. Then it is true that $\phi_1(T)$ is equivalent to

$$(4.2)\quad \phi_1'(T) = \begin{cases} 1 & \text{if } Y_i > \min_{1\leq j\leq k} Y_j + c(kn-t) \text{ for at least one } i \\ 0 & \text{otherwise} \end{cases}$$

given that $T \equiv \sum_{i=1}^{k} X_i = t$.

We note that the constant c in ϕ_1' corresponding to $T=t$ is the same as the c in ϕ_1 corresponding to $T=kn-t$.

The probability of the error of the first kind for ϕ_1 is then given by

$$\begin{aligned} E\phi_1(T|H,t) &= P(\max_{1\leq j\leq k} X_j - X_i > c(t) \text{ for some } i=1,2,\ldots,k \,|\, \sum_{i=1}^{k} X_i = t) \\ &\leq \sum_{i=1}^{k} P(\max_j X_j - X_i > c(t) \,|\, \sum_1^k X_i = t) \\ &= k\, P(\max_j X_j - X_1 > c(t) \,|\, \sum_1^k X_i = t) \\ &= k\{1-[N(k;c,t,n)/\tbinom{kn}{t}]\} \qquad \text{(by Theorem 2.1).} \end{aligned}$$

Hence, for given significance level α, we can find c(t) such that $E\phi_1(T|H,t) \leq \alpha$. The computations of c_1 is discussed in Section 6.

When k=2, the randomized version of test ϕ_1 for H_0: $p_1=p_2$ vs H_1: $p_1 \neq p_2$ is defined by

$$(4.3) \quad \phi_2(T) = \begin{cases} 1 & \text{if } X_1 > \frac{t+c}{2} \text{ or } X_1 < \frac{t-c}{2} \\ \rho & \text{if } X_1 = \frac{t+c}{2} \text{ or } X_1 = \frac{t-c}{2} \\ 0 & \text{otherwise} \end{cases}$$

given that $T \equiv X_1+X_2 = t$ with $E(\phi_2(T)|H_0,t) = \alpha$. It is known that this test is uniformly most powerful unbiased (see, for example, pp. 142-143 in Lehmann [5]). The values of c and ρ in $\phi_2(T)$ can be obtained in Table 1 of Gupta and Nagel [2] for some special values of α and n. For given values of α, the value c is given by t-2c' and ρ is given by $1-\rho'$ where c' and ρ' are the values, corresponding to $P^* = 1-\alpha$ for given k, n and t, given in Table 1 of Gupta and Nagel [2].

Siotani [6] proposed a test for the hypothesis H': $p_1=p_2=\ldots=p_k=p$ where he assumes p to be known. He used $\hat{p} = (\sum_{i=1}^{k} X_i)/kn$ to estimate p when kn is large (Tables of Siotani and Ozawa [7] starts with n=10). According to Siotani's procedure, H' is rejected if, and only if, $X_{max}-X_{min} \geq d$, some constant depending on k, n, p and α. Siotani's test is thus not based on the sufficient statistics $\sum_{i=1}^{k} X_i$ for p. Also, his test is not exact and when kn is small, the test is not available.

In our test, we see that the critical region is independent of p and our test is available for any kn. Furthermore, it is found that the value of d in [7] is monotone increasing with respect to p and thus

$d(p=0.5) = \max_{0.1 \leq p \leq 0.5} d(p)$. Comparing with tables in [7], it is found that $\max_{1 \leq t \leq kn} c(t) \leq d(p=0.5)$ for some k, n and α considering that $d^{+} = d+1$ in [7]. This means that in most cases, our test ϕ_1 has bigger power than Siotani's test.

5. Application to Selection and Ranking

For the problem of selecting a subset containing the population associated with the largest p_i, Gupta and Sobel [4] proposed a procedure. In this paper we utilize the additional non-trivial information, i.e. the sufficient statistics for a common p which is the sum of all observations. Let X_i denote the number of successes in n independent experiments from π_i and let the total sum of all kn observations be t. We study the Gupta (see Gupta [1]) type selection rule conditioned on the total sum of all observations, i.e. select π_i if, and only if, $X_i \geq \max_j X_j - c_1(t)$ given $\Sigma X_i = t$, where c_1 is some non-negative constant depending on t. For the randomized case, we select π_i with probability 1 if, and only if, $X_i > \max_j X_j - c_2(t)$ and select π_i with probability ρ, if and only if $X_i = \max_j X_j - c_2(t)$ given that $\sum_{i=1}^{k} X_i = t$. The probability of correct selection for both rules when the configuration is Ω_0 defined in Section 2 are given in Theorem 2.1. The computations of c_1, c_2 and ρ are discussed in Section 6.

6. Computations of c, c_1, c_2 and ρ_1

To compute the values of the critical region defined by ϕ_1, for given k, n, t and α, we define an auxiliary quantity $P^* = 1 - \frac{\alpha}{k}$. We start from c=0 and compute N(k;c,t,n) using Lemma 2.1. If $N(k;0,t,n) < \binom{kn}{t} P^*$, we increase c by 1 and again compute N(k;c,t,n).

This process continues until for the first time c=a, say, such that $N(k;t,n,a) \geq \binom{kn}{t} P^*$. This value a is the c needed in ϕ_1.

When $t > \frac{kn}{2}$, $\phi_1'(T)$ defined in (4.2) with $c(t')$ is used where $t'=kn-t$.

It should be pointed out that when t=0 we have c=0 and $\rho=0$.

We tabulate the values of c for the test ϕ_1 for the case k=2(1)10, n=1(1)10 and $t=1(1) <\frac{kn}{2}>$.

To compute the values c_1, c_2 and ρ_1 for the selection rules given in Section 5, for given k,n,t and P*, we use the same method used for computing c to compute c_1. We choose c_2 to be the smallest non-negative integer satisfying $N(k;c_2,t,n) > \binom{kn}{t}\cdot P^*$. Then, compute $\rho_1=1-\{[N(k;c_2,t,n) - \binom{kn}{t}\cdot P^*]/M(k;c_2,t,n)\}$.

Acknowledgement

The authors thankfully acknowledge the assistance of Professor S. Panchapakesan for a critical reading of this paper and for many helpful comments. The authors also wish to thank Mr. C. Hinkle for programming and computational assistance.

References

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

[2] Gupta, S. S. and Nagel, K. (1971). On some contributions to multiple decision theory. In Statistical Decision Theory and Related Topics (S. S. Gupta and J. Yackel ed.) Academic Press, New York.

[3] Gupta, S. S. and Panchapakesan, S. (1972). On moments of order statistics from independent binomial populations. Mimeo series #297, Department of Statistics, Purdue University, W. Lafayette, Indiana.

[4] Gupta, S. S. and Sobel, M. (1960). Selecting a subset containing the best of several binomial populations. In Contributions to Probability and Statistics (ed. I. Olkin), 224-248, Stanford, California.

[5] Lehmann, E. L. (1959). Testing Statistical Hypotheses, John Wiley, New York.

[6] Siotani, M. (1956). Order statistics for discrete case with a numerical application to the binomial distribution. Ann. Inst. Stat. Math. 8, 95-104.

[7] Siotani, M. and Ozawa, M. (1958). Tables for testing the homogeneity of k independent binomial experiments on a certain event based on the range. Ann. Inst. Stat. Math. 10, 47-63.

Table of c(t) values for the test of homogeneity $p_1=p_2=\dots=p_k$

k=2

t \ n	1	2	3	4	5	6	7	8	9	10
1	1,1,1,1,1	*	*	*	*	*	*	*	*	*
2		2,2,2,2,2	*	*	*	*	*	*	*	*
3			1,3,3,3,3	3,3,3,3,3	*	*	*	*	*	*
4				2,2,4,4,4	*	2,4,4,4,4	*	*	*	*
5					3,3,3,3,5	3,3,3,5,5	*	3,3,5,5,5	*	*
6						2,4,4,4,6	4,4,4,4,6	*	*	4,4,4,6,6
7							3,3,5,5,5	3,3,5,5,7	*	3,5,5,5,7
8								4,4,4,6,6	*	*
9									3,5,5,5,7	*
10										4,4,4,6,8

k=3

t \ n	1	2	3	4	5	6	7	8	9	10
1	1,1,1,1,1	*	*	*	*	*	*	*	*	*
2	-	2,2,2,2,2	*	*	*	*	*	*	*	*
3		-	2,3,3,3,3	3,3,3,3,3	*	*	*	*	*	*
4			3,3,3,3,3	3,3,3,4,4	*	3,3,4,4,4	*	*	*	*
5			-	3,3,4,4,4	3,4,4,4,5	*	*	*	*	3,4,4,5,5
6				-	-	3,4,4,5,5	4,4,4,5,5	4,4,4,5,6	*	4,4,5,5,6
7					-	4,4,4,5,6	4,4,5,5,6	-	4,4,5,5,6	-
8					-	-	-	-	4,5,5,5,6	4,5,5,6,7
9						-	-	4,5,5,6,7	4,5,5,6,7	-
10							-	-	-	-
11							-	-	-	5,5,6,6,7
12								-	4,5,6,6,7	5,5,6,6,8
13									-	-
14									-	-
15										-

1. The five entries in this table are the c-values for $\phi_1(t)$ defined in (4.1) corresponding to α=0.1, 0.05, 0.025, 0.01 and 0.001.

2. For any t,n the entry * denotes the same c-values as those in the preceding column to the left for the same t. Again the entry - denotes the same c-values as in the preceding row corresponding to t-1 and the same n

Table of c(t) values for the test of homogeneity $p_1=p_2=\ldots=p_k$

k=4

t \ n	1	2	3	4	5	6	7	8	9	10
1	1,1,1,1,1	*	*	*	*	*	*	*	*	*
2	-	2,2,2,2,2	*	*	*	*	*	*	*	*
3		-	2,3,3,3,3	*	3,3,3,3,3	*	*	*	*	*
4		-	3,3,3,3,3	3,3,3,3,4	3,3,3,4,4	*	*	*	3,3,4,4,4	*
5			-	3,3,3,4,4	3,3,4,4,4	3,4,4,4,5	*	*	*	*
6			-	3,3,4,4,4	3,4,4,4,5	*	4,4,4,4,5	4,4,4,5,5	*	*
7				3,4,4,4,4	4,4,4,4,5	*	4,4,4,5,6	4,4,5,5,6	*	*
8				-	4,4,4,5,5	4,4,4,5,6	4,4,5,5,6	-	4,5,5,5,6	*
9					-	4,4,5,5,6	-	4,5,5,5,6	4,5,5,6,6	4,5,5,6,7
10					-	-	4,5,5,5,6	4,5,5,6,6	4,5,5,6,7	5,5,5,6,7
11						-	-	4,5,5,6,7	5,5,5,6,7	5,5,6,6,7
12						-	4,5,5,6,6	5,5,5,6,7	5,5,6,6,7	-
13							-	-	-	-
14							4,5,5,6,7	5,5,6,6,7	-	5,6,6,6,8
15								-	-	5,6,6,7,8
16								-	-	-
17									5,5,6,6,8	-
18									-	-
19										-
20										-

1. The five entries in this table are the c-values for $\phi_1(t)$ defined in (4.1) corresponding to $\alpha=0.1$, 0.05, 0.025, 0.01 and 0.001.

2. For any t,n the entry * denotes the same c-values as those in the preceding column to the left for the same t. Again the entry - denotes the same c-values as in the preceding row corresponding to t-1 and the same n.

Table of c(t) values for the test of homogeneity $p_1=p_2=\ldots=p_k$

k=5

t \ n	1	2	3	4	5	6	7	8	9	10
1	1,1,1,1,1	*	*	*	*	*	*	*	*	*
2	-	2,2,2,2,2	*	*	*	*	*	*	*	*
3	-	-	2,2,3,3,3	2,3,3,3,3	*	*	3,3,3,3,3	*	*	*
4		-	3,3,3,3,3	3,3,3,3,4	*	3,3,3,4,4	*	*	*	*
5		-	-	3,3,3,4,4	3,3,4,4,4	3,3,4,4,4	3,3,4,4,5	3,4,4,4,5	*	*
6			-	3,3,4,4,4	3,4,4,4,5	*	*	4,4,4,4,5	*	*
7			-	3,4,4,4,4	4,4,4,4,5	*	4,4,4,5,5	*	4,4,4,5,6	4,4,4,5,6
8			-	-	-	4,4,4,5,5	4,4,5,5,6	*	4,4,5,5,6	4,4,5,5,6
9				-	4,4,4,5,5	4,4,5,5,6	-	4,5,5,5,6	4,5,5,5,6	*
10				-	-	-	4,5,5,5,6	-	4,5,5,6,6	4,5,5,6,7
11					-	4,5,5,5,6	-	4,5,5,6,6	5,5,5,6,7	*
12					-	-	-	5,5,5,6,7	-	5,5,6,6,7
13					-	-	5,5,5,6,6	5,5,6,6,7	*	-
14						-	5,5,5,6,7	-	5,5,6,6,7	-
15						-	-	-	-	5,6,6,6,7
16							-	-	5,6,6,6,7	5,6,6,7,8
17							-	-	-	-
18							-	-	5,6,6,7,8	-
19								-	-	-
20								-	-	6,6,6,7,8
21									-	-
22									-	6,6,7,7,8
23									-	-
24										-
25										-

1. The five entries in this table are the c-values for $\phi_1(t)$ defined in (4.1) corresponding to $\alpha=0.1$, 0.05, 0.025, 0.01 and 0.001.

2. For any t,n the entry * denotes the same c-values as those in the preceding column to the left for the same t. Again the entry - denotes the same c-values as in the preceding row corresponding to t-1 and the same n.

Table of c(t) values for the test of homogeneity $p_1=p_k=\ldots=p_k$

k=6

t \ n	1	2	3	4	5	6	7	8	9	10
1	1,1,1,1,1	*	*	*	*	*	*	*	*	*
2	-	2,2,2,2,2	*	*	*	*	*	*	*	*
3	-	-	2,2,3,3,3	2,3,3,3,3	*	*	*	*	3,3,3,3,3	*
4		-	3,3,3,3,3	3,3,3,3,4	*	*	*	3,3,3,4,4	3,3,3,4,4	*
5		-	-	3,3,3,4,4	*	3,3,4,4,4	*	-	3,3,4,4,5	*
6		-	-	3,3,4,4,4	3,3,4,4,5	3,4,4,4,5	*	3,4,4,4,5	*	4,4,4,4,5
7			-	-	3,4,4,4,5	4,4,4,4,5	*	4,4,4,5,5	*	*
8			-	3,4,4,4,4	4,4,4,4,5	4,4,4,5,5	*	4,4,4,5,6	4,4,5,5,6	*
9			-	-	4,4,4,5,5	-	4,4,5,5,6	*	-	4,5,5,5,6
10				4,4,4,4,4	-	4,4,5,5,6	-	4,5,5,5,6	*	-
11				-	-	-	4,5,5,5,6	-	5,5,5,6,6	5,5,5,6,7
12				-	4,4,5,5,5	4,5,5,5,6	-	5,5,5,6,6	5,5,5,6,7	-
13					-	-	5,5,5,6,6	5,5,5,6,7	-	5,5,6,6,7
14					-	-	-	-	5,5,6,6,7	-
15					-	-	-	5,5,6,6,7	-	5,6,6,6,7
16						-	5,5,5,6,7	-	5,6,6,6,7	-
17						-	-	-	-	5,6,6,7,8
18						-	5,5,6,6,7	-	-	-
19							-	5,6,6,6,7	5,6,6,7,8	6,6,6,7,8
20							-	-	-	-
21							-	-	-	-
22								-	-	6,6,7,7,8
23								-	6,6,6,7,8	-
24								-	-	-
25									-	-
26									-	-
27									-	-
28										-
29										-
30										-

1. The five entries in this table are the c-values for $\phi_1(t)$ defined in (4.1) corresponding to $\alpha=0.1$, 0.05, 0.025, 0.01 and 0.001.

2. For any t,n the entry * denotes the same c-values as those in the preceding column to the left for the same t. Again the entry - denotes the same c-values as in the preceding row corresponding to t-1 and the same n.

Table of c(t) values for the test of homogeneity $p_1=p_2=\dots=p_k$

k=7

t\n	1	2	3	4	5	6	7	8	9	10
1	1,1,1,1,1	*	*	*	*	*	*	*	*	*
2	-	2,2,2,2,2	*	*	*	*	*	*	*	*
3	-	-	2,2,3,3,3	2,3,3,3,3	*	*	*	*	*	*
4	-	-	3,3,3,3,3	3,3,3,3,4	*	*	*	*	*	*
5		-	-	-	3,3,3,4,4	*	3,3,4,4,4	*	*	*
6		-	-	3,3,3,4,4	3,3,4,4,4	3,4,4,4,5	*	*	*	*
7		-	-	3,3,4,4,4	3,4,4,4,5	4,4,4,4,5	*	*	*	4,4,4,5,5
8			-	3,4,4,4,4	4,4,4,4,5	-	4,4,4,5,5	*	*	4,4,4,5,6
9			-	-	-	4,4,4,5,5	-	4,4,5,5,6	*	*
10			-	4,4,4,4,4	4,4,4,5,5	4,4,5,5,5	4,4,5,5,6	-	4,5,5,5,6	*
11			-	-	-	4,4,5,5,6	4,5,5,5,6	4,5,5,5,6	-	4,5,5,5,6
12				-	-	4,5,5,5,6	-	-	5,5,5,6,6	5,5,5,6,7
13				-	4,4,5,5,5	-	-	5,5,5,6,6	5,5,5,6,7	-
14				-	-	-	5,5,5,5,6	5,5,5,6,7	5,5,6,6,7	*
15					-	-	-	-	-	5,5,6,6,7
16					-	5,5,5,5,6	5,5,5,6,6	5,5,6,6,7	-	5,6,6,6,7
17					-	-	5,5,5,6,7	-	5,6,6,6,7	*
18					-	5,5,5,6,6	5,5,6,6,7	-	-	5,6,6,7,7
19						-	-	5,6,6,6,7	-	5,6,6,7,8
20						-	-	-	5,6,6,7,7	6,6,6,7,8
21						-	-	-	5,6,6,7,8	-
22							-	-	6,6,6,7,8	-
23							-	-	-	6,6,7,7,8
24							-	5,6,6,7,7	-	-
25							-	-	-	-
26								-	-	-
27								5,6,6,6,7	6,6,7,7,8	-
28								-	-	-
29									-	-
30									-	-
31									-	6,7,7,7,8
32									-	-
33										-
34										-
35										-

1. The five entries in this table are the c-values for $\phi_1(t)$ defined in (4.1) corresponding to $\alpha=0.1$, 0.05, 0.025, 0.01 and 0.001.
2. For any t,n the entry * denotes the same c-values as those in the preceding column to the left for the same t. Again the entry - denotes the same c-values as in the preceding row corresponding to t-1 and the same n.

Table of c(t) values for the test of homogeneity $p_1=p_2=\ldots=p_k$

k=8

t \ n	1	2	3	4	5	6	7	8	9	10
1	1,1,1,1,1	*	*	*	*	*	*	*	*	*
2	-	2,2,2,2,2	*	*	*	*	*	*	*	*
3	-	-	2,2,3,3,3	*	2,3,3,3,3	*	*	*	*	*
4	-	-	2,3,3,3,3	3,3,3,3,4	*	*	*	*	*	*
5		-	3,3,3,3,3	-	3,3,3,4,4	*	*	3,3,4,4,4	*	*
6		-	-	3,3,3,4,4	3,3,4,4,4	*	3,4,4,4,5	*	*	*
7		-	-	3,3,4,4,4	3,4,4,4,5	3,4,4,4,5	4,4,4,4,5	*	*	*
8		-	-	3,4,4,4,4	4,4,4,4,5	*	-	4,4,4,5,5	*	*
9			-	-	-	4,4,4,5,5	4,4,4,5,5	4,4,4,5,6	4,4,5,5,6	*
10			-	4,4,4,4,4	-	-	4,4,5,5,6	*	*	4,5,5,5,6
11			-	-	4,4,4,5,5	4,4,5,5,5	-	4,5,5,5,6	*	-
12			-	-	-	4,4,5,5,6	4,5,5,5,6	-	4,5,5,5,6	5,5,5,6,6
13				-	4,4,5,5,5	4,5,5,5,6	-	5,5,5,5,6	5,5,5,6,6	5,5,5,6,7
14				-	-	-	5,5,5,5,6	5,5,5,6,6	5,5,5,6,7	-
15				-	-	-	5,5,5,6,6	5,5,5,6,7	5,5,6,6,7	*
16				-	-	5,5,5,5,6	-	5,5,6,6,7	-	-
17					4,5,5,5,5	-	-	-	-	5,6,6,6,7
18					-	-	5,5,5,6,7	-	-	-
19					-	5,5,5,6,6	5,5,6,6,7	-	-	5,6,6,7,7
20					-	-	-	5,6,6,6,7	-	5,6,6,7,8
21						-	-	-	-	6,6,6,7,8
22						-	-	-	6,6,6,7,8	-
23						-	-	-	-	-
24						-	-	5,6,6,7,7	-	-
25							5,6,6,6,7	-	-	6,6,7,7,8
26							-	-	-	-
27							-	*	6,6,7,7,8	-
28							-	6,6,6,7,7	-	-
29								-	-	-
30								-	-	6,7,7,7,8
31								-	-	-
32								-	-	-
33									-	-
34									-	6,7,7,8,9
35									-	-

(continued)

t \ n	1	2	3	4	5	6	7	8	9	10
36									-	-
37										-
38										-
39										-
40										-

1. The five entries in this table are the c-values for $\phi_1(t)$ defined in (4.1) corresponding to $\alpha=0.1$, 0.05, 0.025, 0.01 and 0.001.

2. For any t,n the entry * denotes the same c-values as those in the preceding column to the left for the same t. Again the entry - denotes the same c-values as in the preceding row corresponding to t-1 and the same n.

Table of c(t) values for the test of homogeneity $p_1=p_2=\dots=p_k$

k=9

t \ n	1	2	3	4	5	6	7	8	9	10
1	1,1,1,1,1	*	*	*	*	*	*	*	*	*
2	-	2,2,2,2,2	*	*	*	*	*	*	*	*
3	-	-	2,2,2,3,3	2,2,3,3,3	2,3,3,3,3	*	*	*	*	*
4	-	-	2,3,3,3,3	3,3,3,3,4	*	*	*	*	*	*
5	-	-	3,3,3,3,3	-	3,3,3,4,4	*	3,3,3,4,4	*	*	*
6		-	-	3,3,3,4,4	3,3,4,4,4	*	3,3,4,4,5	3,4,4,4,5	*	*
7		-	-	3,3,4,4,4	3,4,4,4,4	3,4,4,4,5	*	4,4,4,4,5	*	*
8		-	-	-	3,4,4,4,5	4,4,4,4,5	4,4,4,4,5	-	4,4,4,5,5	*
9		-	-	3,4,4,4,4	4,4,4,4,5	-	4,4,4,5,5	*	-	4,4,5,5,6
10			-	4,4,4,4,4	-	4,4,4,5,5	-	4,4,5,5,6	*	-
11			-	-	4,4,4,5,5	-	4,4,5,5,6	-	4,5,5,5,6	*
12			-	-	-	4,4,5,5,6	4,5,5,5,6	4,5,5,5,6	-	-
13			-	-	-	-	-	-	5,5,5,5,6	5,5,5,6,6
14			-	-	4,4,5,5,5	4,5,5,5,6	-	5,5,5,6,6	*	5,5,5,6,7
15				-	-	-	5,5,5,5,6	-	5,5,5,6,7	5,5,6,6,7
16				-	-	5,5,5,5,6	5,5,5,6,6	5,5,5,6,7	5,5,6,6,7	-
17				-	4,5,5,5,5	-	-	5,5,6,6,7	-	-
18				-	-	-	-	-	-	5,6,6,6,7
19					-	-	5,5,6,6,7	-	5,6,6,6,7	-
20					-	5,5,5,5,6	-	-	-	5,6,6,7,7
21					-	5,5,5,6,6	-	5,6,6,6,7	-	6,6,6,7,8
22					-	-	-	-	5,6,6,7,7	-
23					-	-	-	-	6,6,6,7,7	-
24						-	-	-	6,6,6,7,8	-
25						-	-	5,6,6,7,7	-	6,6,7,7,8
26						-	5,6,6,6,7	6,6,6,7,7	-	-
27						-	-	-	-	-
28							-	-	-	-
29							-	-	-	-
30							-	-	6,6,7,7,8	6,7,7,7,8
31							-	6,6,6,7,8	-	-
32							-	-	-	-
33								-	-	-
34								-	-	-
35								-	-	6,7,7,8,8

(continued)

t\n	1	2	3	4	5	6	7	8	9	10
36								-	-	6,7,7,8,9
37									-	-
38									-	-
39									-	-
40									-	-
41									-	-
42										-
43										-
44										-
45										-

1. The five entries in this table are the c-values for $\phi_1(t)$ defined in (4.1) corresponding to $\alpha=0.1$, 0.05, 0.025, 0.01 and 0.001.

2. For any n,t the entry * denotes the same c-values as those in the preceding column to the left for the same t. Again the entry - denotes the same c-values as in the preceding row corresponding to t-1 and the same n.

Table of c(t) values for the test of homogeneity $p_1=p_2=\ldots=p_k$

k=10

t \ n	1	2	3	4	5	6	7	8	9	10
1	1,1,1,1,1	*	*	*	*	*	*	*	*	*
2	-	2,2,2,2,2	*	*	*	*	*	*	*	*
3	-	-	2,2,2,3,3	2,2,3,3,3	*	2,3,3,3,3	*	*	*	3,3,3,3,3
4	-	-	2,3,3,3,3	3,3,3,3,3	3,3,3,3,4	*	*	*	*	*
5	-	-	3,3,3,3,3	3,3,3,3,4	-	3,3,3,4,4	3,3,3,4,4	*	*	*
6		-	-	3,3,3,4,4	*	3,3,4,4,4	3,3,4,4,4	3,3,4,4,5	3,3,4,4,5	3,4,4,4,5
7		-	-	-	3,3,4,4,4	3,4,4,4,5	*	*	*	4,4,4,4,5
8		-	-	3,3,4,4,4	3,4,4,4,5	4,4,4,4,5	*	*	*	4,4,4,5,5
9		-	-	3,4,4,4,4	4,4,4,4,5	-	4,4,4,5,5	*	*	-
10		-	-	-	-	4,4,4,5,5	-	-	4,4,5,5,6	*
11			-	4,4,4,4,4	-	-	4,4,5,5,6	-	-	4,5,5,5,6
12			-	-	4,4,4,5,5	4,4,5,5,5	-	4,5,5,5,6	-	-
13			-	-	-	4,4,5,5,6	4,5,5,5,6	-	-	5,5,5,5,6
14			-	-	4,4,5,5,5	4,5,5,5,6	-	5,5,5,5,6	5,5,5,6,6	*
15			-	-	-	-	5,5,5,5,6	5,5,5,6,6	-	5,5,5,6,7
16				-	-	-	-	-	5,5,5,6,7	5,5,6,6,7
17				-	4,5,5,5,5	5,5,5,5,6	5,5,5,6,6,	5,5,5,6,7	5,5,6,6,7	-
18				-	-	-	-	5,5,6,6,7	-	-
19				-	-	-	-	-	-	5,6,6,6,7
20				-	-	5,5,5,6,6	5,5,6,6,7	-	5,6,6,6,7	-
21					-	-	-	5,6,6,6,7	-	-
22					-	-	-	-	-	6,6,6,7,7
23					-	-	-	-	5,6,6,7,7	6,6,6,7,8
24					-	-	-	-	6,6,6,7,7	-
25					-	-	5,6,6,6,7	-	6,6,6,7,8	-
26						-	-	6,6,6,6,7	-	6,6,7,7,8
27						-	-	6,6,6,7,7	-	-
28						-	-	-	-	-
29						-	-	-	6,6,7,7,8	-
30						-	-	-	-	-
31							-	-	-	6,7,7,7,8
32							-	6,6,6,7,8	-	-
33							-	-	-	-
34							-	-	-	-
35								-	-	-

(continued)

t\n	1	2	3	4	5	6	7	8	9	10
36								-	-	6,7,7,8,8
37								-	-	6,7,7,8,9
38								-	6,7,7,7,8	-
39								-	-	-
40								-	-	-
41									-	-
42									-	-
43									-	-
44									-	7,7,7,8,9
45									-	-
46										-
47										-
48										-
49										-
50										-

1. The five entries in this table are the c-values for $\phi_1(t)$ defined in (4.1) corresponding to α=0.1, 0.05, 0.025, 0.01 and 0.001.

2. For any n,t the entry * denotes the same c-values as those in the preceding column to the left for the same t. Again the entry - denotes the same c-values as in the preceding row corresponding to t-1 and the same n.

Unclassified
Security Classification

DOCUMENT CONTROL DATA - R&D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) Purdue University	2a. REPORT SECURITY CLASSIFICATION Unclassified
	2b. GROUP

3. REPORT TITLE

On the Conditional Distribution of a Statistic that Arises in Tests of Homogeneity and Selection and Ranking of Binomial Populations

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

Technical Report, August 1973

5. AUTHOR(S) (Last name, first name, initial)

Gupta, Shanti S. and Huang, Wen-Tao

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
August 1973	12	7

8a. CONTRACT OR GRANT NO. N00014-67-A-0226-00014	9a. ORIGINATOR'S REPORT NUMBER(S) Mimeo Series #329
b. PROJECT NO.	
c.	9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)
d.	

10. AVAILABILITY/LIMITATION NOTICES

Distribution of this document is unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
	Office of Naval Research Washington, DC

13. ABSTRACT

Let $\pi_1, \pi_2, \ldots, \pi_k$ be k binomial populations such that X_i, the random observation from π_i, has density $b(n;p_i)$ for $i=1,2,\ldots,k$. In this paper we derive the conditional distribution of the statistic $\max_{1\le j\le k} X_j - X_i$ given $\sum_{i=1}^{k} X_i = t$ assuming $p_1=p_2=\ldots=p_k$. The moments of $\max_{1\le j\le k} X_j - X_i$ and $\max_{1\le j\le k} X_j$ given $\sum_{i=1}^{k} X_i = t$ are also obtained. A critical function for testing the hypothesis $p_1=p_2=\ldots=p_k$ is proposed. Some comparisons with Siotani's test are made. The proposed test does not depend on the common value of p_i. Two selection rules based on $\max_{1\le j\le k} X_j - X_i$ given $\sum_{i=1}^{k} X_i = t$ are proposed and studied. Tables of the critical values for the test of homogeneity are also given. It is shown that for k=2, the exact test which is uniformly most powerful unbiased, can be carried out by using the tables in Gupta and Nage [2].

DD FORM 1 JAN 64 1473

Unclassified
Security Classification