ON SOME INEQUALITIES AND MONOTONICITY PROPERTIES WITH SPECIAL REFERENCE TO SELECTION AND RANKING PROBLEMS*

by

Shanti S. Gupta, Purdue University
Deng-Yuan Huang, National Taiwan Normal University
S. Panchapakesan, Southern Illinois University

Technical Report #82-37

Department of Statistics
Purdue University

October 1982

*This research was supported by the Office of Naval Research Contract N00014-75-C-0455 at Purdue University. Reproduction in whole or in part is permitted for any purpose of the United States Government.

ON SOME INEQUALITIES AND MONOTONICITY RESULTS IN SELECTION AND RANKING THEORY*

by

Shanti S. Gupta, Purdue University
Deng-Yuan Huang, National Taiwan Normal University
S. Panchapakesan, Southern Illinois University

1. INTRODUCTION

Inequalities play a fundamental role in nearly all branches of mathematics -- especially so in probability and statistics. The impact of basic inequalities such as those that carry the names of Cauchy-Schwarz, Chebyshev, Cramér-Rao, and Bonferroni in statistics is well known. Inequalities have been profitably used to obtain bounds for probabilities that are more tedious to compute or analytically impossible to handle. Especially in reliability problems, the limited assumptions that could be made about the nature of the life distributions of the components of a system as well as the structure of the system itself render inequalities not merely useful and desirable but essential. Since interest in inequalities pervades through nearly all branches of mathematics, significant contributions have been made by a very large number of researchers whose efforts span well over a century. From time to time, books and monographs have been written which are completely devoted to inequalities. The classic book of Hardy, Littlewood and Pólya [35], first published in 1934, is a remarkable collection of mathematical inequalities. Some important works that followed are Beckenbach and Bellman [12], Godwin [20], Kazarinoff [40], Marshall and Olkin [47], Mitrinović [49], [50], Pólya and Szegö [54], Shisha [57], and Tong [59]. Of these, the monographs of Marshall and Olkin [47] and Tong [61] contain the recent developments in the area of

*This research was supported by the Office of Naval Research Contract N00014-75-C-0455 at Purdue University. Reproduction in whole or in part is permitted for any purpose of the United States Government.

multivariate probability inequalities; this topic has seen a major growth in the last ten or fifteen years. In this connection we also refer to a recent review paper by Eaton [19].

In selection and ranking problem, inequalities and monotonicity properties have a vital role to play. Consider the classical formulations of these problems in which one proposes a procedure which will guarantee a minimum probability of correct selection (PCS). This amounts to evaluating the PCS, determining the parametric configuration for which the PCS is minimum, and then determine the constants defining the procedure so that this minimum is at least a specified level P*. Determining this configuration, known as a least favorable configuration (LFC), is a vital part of the analysis. Obviously, this involves establishing an inequality that the PCS for a certain parametric configuration does not exceed the PCS for any other configuration. In some situations, this can be established by demonstrating a monotonic behavior of the PCS. There are a number of problems in which the LFC cannot be analytically established; in such cases, recourse has been taken to obtain a good lower bound for the PCS first and then seek the LFC for this lower bound. Even when the LFC for the PCS can be analytically established, inequalities are further useful in obtaining conservative but easier-to-compute values for the constants of the procedure. Similar situations arise when we consider the worst configuration for any suitable performance characteristic such as the expected number of nonbest populations included in the selected subset. Additional uses of inequalities arise due to specific assumptions regarding the families of distributions under consideration; for example, distributions having an increasing failure rate (IFR) and increasing failure rate average (IFRA). For a general view of selection and ranking problems and the various formulations and goals that have been studied, we refer to Gupta and Panchapakesan [31].

In this paper, we restrict our attention mainly to some inequlities and monotonicity properties that have typically arisen in the development of the selection and ranking theory. Basic to the setup of these problems is the assumption regarding some order relations such as stochastic ordering and the monotone likelihood property. These and other related ideas, along with some basic inequalities that arise under these assumptions are discussed in Section 2. In reliability models, partial order relations such as convex ordering, star ordering and tail ordering play an important role. Section 3 deals with restricted families of distributions defined by such partial order relations and some important inequalities obtained in the investigation of selection problems for such families. Interesting inequalities appear in the study of selection rules for normal, multinomial and gamma distributions. These are discussed in Section 4.

2. ORDERED FAMILIES OF DISTRIBUTIONS

Inherent to a selection and ranking problem is the choice of a ranking parameter, say, θ. The natural setup consists of k populations that are described by their associated probability distributions P_{θ_i}, $i = 1, \ldots, k$, where $\theta_i \in \Omega$, a subset of the real line. In other words, these populations belong to a family $\mathcal{P} = \{P_\theta\}$ indexed by $\theta \in \Omega$. A reasonable procedure can be proposed if we have some knowledge of the structural properties of this family. For example, if $X_1, \ldots, X_k$ are observations from the k populations, we would like to say that large values of X generally go with large values of θ. Such statements bring in order relations for distributions belonging to the family. We will now formalize such concepts and state some monotonicity results.

2.1. Stochastic Ordering and Monotone Likelihood Ratio Property.

Let X be a real valued random variable with distribution P_θ, $\theta \in \Omega$. Then the family $\mathcal{P} = \{P_\theta\}$, $\theta \in \Omega$, is said to be stochastically increasing (SI) in θ if for $\theta_1 < \theta_2$, the distributions P_{θ_1} and P_{θ_2} are distinct, and for any real number a, (2.1) $$P_{\theta_1}[X \in (a,\infty)] \le P_{\theta_2}[X \in (a,\infty)].$$

It is well known that a stronger property is that of monotone likelihood ratio (MLR) introduced by Karlin and Rubin [39] and this is equivalent to the frequency function having total positivity of order 2 (TP_2). The concept of total positivity is, however, more general and is not restricted to frequency functions (see Karlin [38]).

A basic result of Lehmann ([44], p. 112, Problem 11) can be stated as follows.

Theorem 2.1. Let $\{P_\theta\}$, $\theta \in \Omega$, be an SI family of distributions and let $\psi(x)$ be a real valued function nondecreasing in x. Then $E_\theta[\psi(X)]$ is nondreasing in θ.

A straight forward generalization of this theorem independently obtained by Alam and Rizvi [4] and Mahamunulu [46] is given below.

Theorem 2.2. Let $\{P_\theta\}$, $\theta \in \Omega$, be an SI family of distributions. Let $X_1, \ldots, X_k$ be independent ramdom variables, X_i having the distribution P_{θ_i}, $\theta_i \in \Omega$, $i = 1, \ldots, k$. Then $E_{\underline{\theta}}\psi(X_1, \ldots, X_k)$ is nondecreasing in each component of $\underline{\theta} = (\theta_1, \ldots, \theta_k)$ if $\psi(x_1, \ldots, x_k)$ is nondecreasing in each of its arguments.

Theorem 2.2 has been successfully applied to many selection problems. For suitably chosen $\psi(x_1, \ldots, x_k)$, the expectation $E_{\underline{\theta}}\psi(X_1, \ldots, X_k)$ becomes the PCS. The monotonicity property of the expectation enables one to obtain the LFC.

Another generalization of Theorem 2.1 in a different direction is due to Gupta and Panchapakesan [28] who considered a class of subset selection rules defined through a class of functions h. For evaluating the infimum of the PCS, we need to minimize over θ the expectation $E_\theta[\psi(X,\theta)]$. The following theorem of Gupta and Panchapakesan [28] gives a sufficient condition for the monotonicity of $E_\theta[\psi(X,\theta)]$.

Theorem 2.3. Let $F(\cdot;\theta)$, $\theta \in \Omega$, be a family of absolutely continuous distributions on the real line $\mathbb{R}$ with continuous densities $f(\cdot;\theta)$ and let $\psi(x,\theta)$ be a bounded real valued function possessing first partial derivatives ψ_x and ψ_θ with respect to x and θ, respectively, and satisfying certain regularity conditions C. Then $E_\theta[\psi(X,\theta)]$ is nondecreasing in θ provided that for all $\theta \in \Omega$,

$$f(x;\theta)\psi_\theta(x,\theta) - \frac{\partial F(x;\theta)}{\partial\theta}\psi_x(x,\theta) \geq 0 \qquad \text{a.e.x,} \tag{2.2}$$

where the regularity conditions C are:

(i) for all $\theta \in \Omega$, $\psi_x(x,\theta)$ is Lebesgue integrable on $\mathbb{R}$; and

(ii) for every $[\theta_1,\theta_2] \subset \Omega$ and $\theta_3 \in \Omega$, there exists g(x) depending only on θ_1, θ_2, θ_3 such that

$$\left| \psi_\theta(x,\theta)f(x;\theta_3) - \frac{\partial F(x;\theta)}{\partial\theta}\psi_x(x,\theta_3) \right| \leq g(x)$$

for all $\theta \in [\theta_1,\theta_2]$ and g(x) is Lebesgue integrable on $\mathbb{R}$.

Remark 2.4 (1) If $\psi(x,\theta) = \psi(x)$ for all $\theta \in \Omega$, the sufficient condition (2.2) reduces to $\frac{\partial F(x,\theta)}{\partial\theta}\psi_x(x) \leq 0$, which is satisfied by the hypotheses of Theorem 2.1 since $\{F_\theta\}$ is SI and $\psi(x)$ is nondecreasing in x.

(2) For the class of procedures defined by Gupta and Panchapakesan [28], $\psi(x,\theta) = F(h(x);\theta)$ and (2.2) becomes

$$f(x;\theta)\frac{\partial F(h(x);\theta)}{\partial\theta} - h'(x)\, f(h(x);\theta)\frac{\partial F(x;\theta)}{\partial\theta} \geq 0 \tag{2.3}$$

where $h'(x) = (d/dx)\, h(x)$.

(3) This condition has been specialized to the cases of (i) location parameter, (ii) scale parameter, and (iii) convex mixtures of distributions by Gupta

and Panchapakesan for the purposes of specific applications.

(4) An analogue of this theorem for discrete distributions is given by Panchapakesan [52], who has given in another paper [53] sufficient conditions for monotonicity when Ω is a countable set.

(5) The monotonicity of $E_\theta[\psi(x,\theta)]$ in θ is strict if strict inequality holds in (2.3) on a set of positive Lebesgue measure.

(6) Obvious modifications in Theorems 2.1 through 2.3 give monotonicity in the opposite direction.

For subset selection rules the expected subset size has been used as a performance characteristic. We naturally want to know the worst configuration in the sense that it maximizes the expected subset size. The following theorem (discussed and proved without a formal statement) of Gupta and Panchapakesan [28] gives a sufficient condition for the expected subset size to be maximized at an equi-parameter configuration.

Theorem 2.5. Let $X_1, \ldots, X_k$ be independent random variables, X_i having an absolutely continuous distribution $F(\cdot,\theta_i)$, $\theta_i \in \Omega$, with continuous densities $f(\cdot,\theta_i)$. Let $\psi(x,\theta)$ be a bounded function possessing the first partial derivatives ψ_x and ψ_θ with respect to x and θ, respectively, and satisfying the regularity conditions of Theorem 2.3. Define

$$B(\theta_1, \ldots, \theta_k) = \sum_{i=1}^{k} E_{\theta_i}\Big[\prod_{\substack{r=1\\ r\neq i}}^{k} \psi(X,\theta_r)\Big].$$ Then

$$B(\underline{\theta} \mid \theta_1 \le \ldots \le \theta_k) \le B(\underline{\theta} \mid \theta_1 = \ldots = \theta_k) \tag{2.4}$$

provided that, for all $\theta_i \le \theta_j$ and a.e.x, the following holds:

$$\frac{\partial\psi(x,\theta_i)}{\partial\theta_i} f(x;\theta_j) - \frac{\partial\psi(x,\theta_j)}{\partial x}\frac{\partial F(x;\theta_i)}{\partial\theta_i} \ge 0. \tag{2.5}$$

Remarks 2.6. As in the case of Theorem 2.3, Gupta and Panchapakesan [28] have specialized this for (i) location parameter, (ii) scale parameter, and (iii) convex mixtures. For their class of procedures, $\psi(x,\theta_i) = F(h(x);\theta_i)$, $i=1, \ldots, k$. For location and scale parameter cases, the usual choices are $h(x) = x+b$, $b \geq 0$, and $h(x) = ax$, $a \geq 1$, respectively. In these cases, the left-hand side of (2.3) is zero for all x; thereby showing that $E_\theta[\psi(X,\theta)]$ is independent of θ. Further, the condition (2.5) in these cases reduces to the monotone likelihood ratio property, a result directly proved by Gupta [22]. □

Now, we note that Theorem 2.2 is a simple generalization of Theorem 2.1 to $\mathbb{R}^k$, the k-dimenional Euclidean space. We now consider various generalizations of the concepts of stochastic ordering and monotone likelihood ratio to distributions in higher dimensions. To this end, we introduce the following definitions.

Definition 2.7. A function ψ defined on $\mathbb{R}^k$ is said to be increasing with respect to a partial order relation "$\prec$" if $\underline{x}_1 \prec \underline{x}_2$ implies $\psi(\underline{x}_1) \leq \psi(\underline{x}_2)$ for all $\underline{x}_1, \underline{x}_2 \in \mathbb{R}^k$.

Definition 2.8. A set S in $\mathbb{R}^k$ is said to be an increasing set if its indicator function is increasing; that is, if $\underline{x}_1 \in S$ and $\underline{x}_1 \prec \underline{x}_2$, then $\underline{x}_2 \in S$.

Let $\underline{X}$ be a k-dimensional random vector with distribution $P_{\underline{\theta}}$ in $\mathbb{R}^k$, where $\underline{\theta} = (\theta_1, \ldots, \theta_k)$. Let $P_{\underline{\theta}}(S) = P_{\underline{\theta}}(\underline{X} \in S)$ for any measurable set S.

Definition 2.9. A distribution $P_{\underline{\theta}}$ is said to have stochastically increasing property (SIP) in $\underline{\theta}$ if $P_{\underline{\theta}_1}(S) \leq P_{\underline{\theta}_2}(S)$ for every monotone increasing measurable set S and for every $\underline{\theta}_1 \prec \underline{\theta}_2$.

The following lemma is due to Lehmann [43].

<u>Lemma 2.10.</u> A family of distributions $P_{\underline{\theta}}$ has SIP in $\underline{\theta}$ if and only if $E_{\underline{\theta}_1}\psi(\underline{X}) \leq E_{\underline{\theta}_2}\psi(\underline{X})$ for all increasing integrable functions $\psi(\underline{X})$ and $\underline{\theta}_1 \prec \underline{\theta}_2$.

The following theorem follows easily from Lemma 2.10.

<u>Theorem 2.11.</u> Let the distribution of $\underline{X}$ have SIP in $\underline{\theta}$ and let $\psi(\underline{x},\underline{\theta})$ be increasing in $\underline{x}$ and $\underline{\theta}$. Then $E_{\underline{\theta}}\psi(\underline{X},\underline{\theta})$ is increasing in $\underline{\theta}$.

When we have independence, it is easily verified that the MLR property implies SIP (Lehmann [43]). When we deal with correlated random variables $X_1, \ldots, X_n$, it is natural to look for a generalized concept of MLR in higher dimensions. For a density $f(x;\theta)$ in the one-dimensional case, the MLR property says that

(2.6) $$f(x_1;\theta_1)\, f(x_2;\theta_2) - f(x_1;\theta_2)\, f(x_2;\theta_1) \geq 0.$$

for every $x_1 \leq x_2$ and $\theta_1 \leq \theta_2$. We can rewrite (2.6) in the form

(2.7) $$f(\underline{x};\underline{\theta}) \geq f(\underline{x};\, (1,2)\underline{\theta})$$

where $f(\underline{x};\underline{\theta}) = \prod_{i=1}^{2} f(x_i;\theta_i)$, $\underline{\theta} = (\theta_1,\theta_2)$, and $(1,2)\underline{\theta}$ is the vector obtained from $\underline{\theta}$ by interchanging θ_1 and θ_2. This provides the motivation for the following definition of Property M by Eaton [18].

<u>Definition 2.12.</u> A family of real valued density functions $\{f_\alpha(\underline{x};\underline{\theta})\}$, $\alpha \in \mathscr{A}$, is said to have <u>Property</u> <u>M</u> if, for each $\alpha \in \mathscr{A}$ and for each pair (i,j), $1 \leq i \neq j \leq k$, the following holds:

(2.8). $$x_i \geq x_j \text{ and } \theta_i \geq \theta_j \Rightarrow f_\alpha(\underline{x};\underline{\theta}) \geq f_\alpha(\underline{x};\, (i,j)\underline{\theta}).$$

Eaton [18] has given a necessary and sufficient condition for a class of densities to possess Property M. Bechhofer, Kiefer and Sobel ([11], p. 41) in their monograph on sequential identification and selection rules define a rankability condition which is same as Property M. Hollander, Proschan and Sethuraman [36] have defined a concept of decreasing in transposition (DT) which is also same as Property M; however, their motivation comes from finding classes of functions which share certain properties of Schur functions. In fact, when $g(\underline{x},\underline{\theta}) = h(\underline{x}-\underline{\theta})$, g is DT on $\mathbb{R}^{2k}$ if and only if h is Schur-concave on $\mathbb{R}^k$. Finally, Marshall and Olkin ([47], p. 160) have also used DT functions but they call them arrangement increasing (AI) functions.

It is important to note that, unlike in the case of one-dimensional distributions, Property M does not imply SIP. The following simple example of Hsu [37] illustrates this point.

Example 2.13. $\underline{X} = (X_1,X_2)$ has the following distribution for four permissible values of $\underline{\theta} = (\theta_1,\theta_2)$.

$\underline{\theta}$ \ $\underline{x}$	(5,6)	(6,5)
(1,2)	0.9	0.1
(2,1)	0.1	0.9
(3,4)	0.6	0.4
(4,3)	0.4	0.6

Further, we can have SIP without Property M; this is true in one-dimension also. Finally, it is possible to have both SIP and Property M as it is the case with the multinomial distribution.

Another generalization of MLR is given by Gupta and Huang [25] who obtained for a family of densities having this generalized MLR property an essentially complete class of multiple decision rules.

Definition 2.14. A probability density $f(\underline{x};\underline{\theta})$ is said to have a generalized monotone likelihood ratio (GMLR) in $\underline{x}$, if for every i and all fixed x_j, $j = 1, \ldots, k$, $j \neq i$, $f(\underline{x};\underline{\theta}_1)/f(\underline{x};\underline{\theta}_2)$ is nondecreasing in x_i, where $\underline{\theta}_\ell = (\theta_{\ell 1}, \ldots, \theta_{\ell k})$, $\ell = 1, 2$; $\theta_{1j} = \theta_{2j}$ for all $j \neq i$, and $\theta_{1i} > \theta_{2i}$.

What we have discussed so far are some basic assumptions that are usually made regarding the underlying family, and the monotonicity behavior of the expectations of certain functions. Also of relevance here is the concept of stochastic majorization and inequalities obtained by majorization. One definition of stochastic majorization is to say that $\underline{X}$ is stochastically majorized by $\underline{Y}$ if $E(\psi(\underline{X})) \leq E(\psi(\underline{Y}))$ for all Schur-convex functions ψ; of course, there are other possible definitions (see Marshall and Olkin [47], chapter 11). Majorization techniques can be used to show that $E[\psi(\underline{X})] \leq E[\psi(\underline{Y})]$ for several other families of functions ψ. The relevance of these results to selection problems is obvious, when $\psi(\underline{X})$ is the indicator function of the event "a correct selection is made." For several useful inequalities in this direction, we refer to Chapters 12 and 13 of Marshall and Olkin [47].

3. RESTRICTED FAMILIES OF DISTRIBUTIONS

By restricted families of distributions, we mean a family of distributions $\mathfrak{F}$ each member of which is partially ordered in a sense with respect to a given distribution G. Such families do arise naturally in reliability studies. More commonly known families of this type are those with increasing failure rate (IFR) and increasing failure rate on the average (IFRA) and naturally those with corresponding decreasing properties. In dealing with such classes we do not know the exact forms of the distributions that belong to $\mathfrak{F}$, but we do know the nature of the partial order relation and the distribution G. Precisely this knowledge enables one to find bounds for quantities of interest such as the probability of survival and mean life in terms of G. Inequalities are thus very important in reliability studies. As a matter of no surprise, significant contributions to inequalities for restricted families have been made by researchers in mathematical reliability -- Barlow, Marshall and Proschan, to mention a few. Typical of these problems is the use of order statistics. Many important order statistics inequalities that arise in inference problems of reliability are reviewed by Gupta and Panchapakesan [29].

Selection procedures for restricted families of distributions were first studied by Barlow and Gupta [7]. When we have k populations from $\mathfrak{F}$, we can generally evaluate (under some additional assumptions) the infimum of the PCS in terms of the known G by establishing appropriate inequalities. We describe in this section such inequalities and explain the contexts of the selection problems. For purpose of describing these results, we need to introduce some definitions.

Assuming that all our distributions are absolutely continuous, we now define some of the special order relations of interest to us. F and G denote distribution functions.

<u>Definitions 3.1.</u> (i) F is said to be convex with respect to (w.r.t.) G (written $F \underset{c}{<} G$) if and only if $G^{-1}F(x)$ is convex on the support of F. (ii) F is <u>star shaped</u> w.r.t. G($F \underset{*}{<} G$) if and only if $F(0) = G(0) = 0$ and $G^{-1}F(x)/x$ is increasing in $x \geq 0$ on the support of F. (iii) F is <u>tail ordered</u> w.r.t. G($F \underset{t}{<} G$) if and only if $F(0) = G(0) = 1/2$, and $G^{-1}F(x) - x$ is nondecreasing on the support of F.

If $G(x) = 1-e^{-x}$, $x \geq 0$, then (i) defines the class of IFR distributions studied by Barlow, Marshall and Proschan [9] while (ii) defines the class of IFRA distributions studied by Birnbaum, Esary and Marshall [14]. Convex ordering was studied by van Zwet [62]. Doksum [17] has used the tail ordering. It is easy to verify that the above order relations are all partial order relations. One can also easily see that convex ordering implies star ordering. Without the assumption of the common median zero, the definition (iii) has been used by Bickel and Lehmann [13] to define an <u>ordering by spread</u> with the germinal concept attributed to Brown and Tukey [15] by them. This kind of ordering has also been perceived by Saunders and Moran [56] in the context of a neurobiological problem and is called <u>ordering by dispersion</u> by them. We now give a formal definition below.

<u>Definition 3.2.</u> G is <u>more dispersed</u> than F ($F \underset{d}{\prec} G$) if

(3.1) $G^{-1}(\beta) - G^{-1}(\alpha) \geq F^{-1}(\beta) - F^{-1}(\alpha)$ for all $0 < \alpha < \beta < 1$.

By setting $x = F^{-1}(\beta)$ and $y = F^{-1}(\alpha)$, it is easy to see that (3.1) is equivalent to saying that $G^{-1}F(t) - t$ is increasing in t. However, (3.1) presents the idea more clearly, that is, any two percentage points of G are at least as far apart as the corresponding percentage points of F.

Finally, we define a general partial order relation through a class of real functions introduced by Gupta and Panchapakesan [29] The star and tail orderings can be obtained as special cases.

<u>Definition 3.3.</u> Let $\mathcal{H} = \{h(x)\}$ be a class of real valued functions h(x) defined on the real line. Let F and G be distributions on the real line such that F(0) = G(0). We say that F is $\mathcal{H}$-<u>ordered</u> w.r.t. G ($F \underset{\mathcal{H}}{\prec} G$) if

$$G^{-1}F(h(x)) \geq h(G^{-1}F(x)) \tag{3.2}$$

for all $h \in \mathcal{H}$ and all x on the support of F.

All the order relations we have defined so far can easily be verified to be partial order relations in that they satisfy only reflexivity and transitivity. It can be seen immediately from the above definition that, if $\mathcal{H} = \{ax, a \geq 1\}$ and $F(0) = G(0) = 0$, we get the star ordering and that the tail ordering is obtained by taking $\mathcal{H} = \{x+b, b \geq 0\}$ and $F(0) = G(0) = 1/2$. Also, if we do not include $F(0) = G(0)$ in the definition, then the dispersion ordering becomes a special case.

The next theorem gives the basic inequality of Gupta and Panchapakesan [29] and some related inequalities.

Theorem 3.4. Let $X_0, X_1, \ldots, X_p (Y_0, Y_1, \ldots, Y_p)$ be independent and identically distributed, each with distribution function F (G), and let $F \underset{\mathcal{H}}{<} G$. Then the following inequalities hold.

(a) $\Pr\{h(X_0) \geq X_i, i=1, \ldots, p\} \geq \Pr\{h(Y_0) \geq Y_i, i=1, \ldots, p\}$,

(b) $\Pr\{X_0 \geq h(X_i), i=1, \ldots, p\} \leq \Pr\{Y_0 \geq h(Y_i), i=1, \ldots, p\}$,

(c) $\Pr\{h(X_0) \leq X_i, i=1, \ldots, p\} \leq \Pr\{h(Y_0) \leq Y_i, i=1, \ldots, p\}$,

(d) $\Pr\{X_0 \leq h(X_i), i=1, \ldots, p\} \geq \Pr\{Y_0 \leq h(Y_i), i=1, \ldots, p\}$,

Proof. We will prove (a). The other inequalities can be established similarly. Let $\varphi = G^{-1}F$. Then

$\Pr\{h(X_0) \geq X_i, i=1, \ldots, p\}$

$= \Pr\{\varphi(h(X_0)) \geq \varphi(X_i), i=1, \ldots, p\}$, since φ is nondecreasing

$\geq \Pr\{h(\varphi(X_0)) \geq \varphi(X_i), i=1, \ldots, p\}$, since $F \underset{\mathcal{H}}{<} G$

$= \Pr\{h(Y_0) \geq Y_i, i=1, \ldots, p\}$, since $\varphi(X_i)$ is stochastically equal to Y_i, $i=0, 1, \ldots, p$. □

The inequalities (a) through (d) of the above theorem can be rewritten respectively as

$$\int F^p(h(x))\, dF(x) \geq \int G^p(h(x))\, dG(x), \tag{3.3}$$

$$(3.4) \qquad \int F^p(h^{-1}(x))\, dF(x) \le \int G^p(h^{-1}(x))\, dG(x)$$

$$(3.5) \qquad \int [1-F(h(x))]^p dF(x) \le \int [1-G(h(x))]^p dG(x),$$

and

$$(3.6) \qquad \int [1-F(h^{-1}(x))]^p\, dF(x) \ge \int [1-G(h^{-1}(x))]^p\, dG(x),$$

where h^{-1} is assumed to exist and the integrals extend over the supports of the relevant distributions. Gupta [23] obtained essentially these inequalities for any $p > 0$ under a set of hypotheses which amounts to $\mathcal{H}$-ordering. Also, in selection and ranking problems, we typically get the probabilities, $\Pr\{h(X_0) \ge X_i,\ i=0, 1, \ldots, p\}$ and $\Pr\{X_0 \le h(X_i),\ i=0, 1, \ldots, p\}$.

These are same as the left-hand side probabilities in (a) and (d) of Theorem 3.4 if we assume that $h(x) \ge x$. This is satisfied for natural choices of $h(x)$ in the procedures. It should be noted that $h(x) \ge x$ in the special classes of $\mathcal{H}$ yielding star and tail ordering.

Interesting special inequalities are obtained by considering special pairs of F and G in Theorem 3.4. We mention here a few of them relevant to selection rules, thus generally applying inequalities (a) and (d) of Theorem 3.4.

Suppose $X_1, \ldots, X_n$ are i.i.d. with distribution F and $Y_1, \ldots, Y_n$ are i.i.d. with distribution G. Let $F \underset{\mathcal{H}}{<} G$. Let $F_{[j]}$ and $G_{[j]}$ denote the cdf's of the jth order statistic of the X_i and the Y_i respectively. Define

$$B_{j,n}(x) = [n!/(j-1)!(n-j)!] \int_0^x u^{j-1}(1-u)^{n-j} du$$

so that

$$(3.7) \qquad F_{[j]}(x) = B_{j,n}(F(x)) \equiv B_{j,n}F(x).$$

Since

$$(3.8) \qquad G^{-1}_{[j]}\, F_{[j]}\, (x) = [B_{j,n}G]^{-1} B_{j,n}F(x) = G^{-1}F(x),$$

we see that order statistics preserve $\mathcal{H}$-ordering. So we get

$$\int F^p_{[j]}(h(x))\, dF_{[j]}(x) \geq \int G^p_{[j]}(h(x))\, dG_{[j]}(x) \tag{3.9}$$

and

$$\int [1-F_{[j]}(h^{-1}(x))]^p\, dF_{[j]}(x) \geq \int [1-G_{[j]}(h^{-1}(x))]^p\, dG_{[j]}(x). \tag{3.10}$$

Barlow and Gupta [7] studied subset selection procedures for selecting the distribution with the largest (smallest) α-quantile from $k = p+1$ distributions that are star ordered w.r.t. G. In their procedures, $h(x) = ax$, $a \geq 1$. With this choice of $h(x)$, the right-hand sides of (3.9) and (3.10) become the infimum of PCS in these two cases. Specializing these inequalities further to the case of IFRA distributions, we get the following corollary.

<u>Corollary 3.5.</u> Let $F_{[j]}$ denote the cdf of the jth order statistic in a random sample of n observations from an IFRA distribution F. Then

$$\int_0^\infty F^p_{[j]}(ax)\, dF_{[j]}(x) \geq \int_0^\infty G^p_{[j]}(ax)\, dG_{[j]}(x) \tag{3.11}$$

and

$$\int_0^\infty [1-F_{[j]}(\tfrac{x}{a})]^p\, dF_{[j]}(x) \geq \int_0^\infty [1-G_{[j]}(\tfrac{x}{a})]^p\, dG_{[j]}(x). \tag{3.12}$$

where

$$G_{[j]}(x) = \sum_{t=j}^{n} \binom{n}{t} [1-e^{-x}]^t\, e^{-(n-t)x} = B_{j,n}(1-e^{-x}). \tag{3.13}$$

Barlow, Gupta and Panchapakesan [8] have tabulated the values of a^{-1} for which the right-hand sides of (3.11) and (3.12) are equal to P* (the guaranteed minimum PCS) for selected values of p, n, j and P*. Gupta and Panchapakesan [30] studied a similar quantile selection procedure for selecting the largest quantile for distributions that are star ordered w.r.t. the standard normal distribution folded at the origin. In this case, the inequality (3.11) holds with $G_{[j]}(x) = B_{j,n}(2\Phi(x)-1)$, where $\Phi(x)$ is the standard normal cdf. The values of a^{-1} for which the right-hand side of (3.11) is equal to P* are tabulated by Gupta and Panchapakesan [30] for selected values of p, n, j and P*.

It is easy to verify that the folded normal distribution is an IFR and therefore an IFRA distribution. So we can obtain further inequalities by taking $F_{[j]}(x) = B_{j,n}(2\Phi(x)-1)$ in the above corollary.

We can get similar inequalities for F and G such that $F \underset{d}{<} G$. We have to take $h(x) = x+b$, $b > 0$, in (3.5) and (3.6). More inequalities can be obtained by considering $F_{[j]}$ and $G_{[j]}$ with special choices of G. These inequalities occur in selection procedures of Barlow and Gupta [7] for selection in terms of medians for a class of distributions (not defined in this paper) and the procedures of Gupta and Panchapakesan [29] who have used the logistic distribution for G.

<u>Remarks 3.6</u> Suppose we take $\mathcal{H} = \{ax, a \geq 1\}$ in Theorem 3.4. Then, letting $Z_1 = \max\{\frac{X_1}{X_0}, \ldots, \frac{X_p}{X_0}\}$, $Z_2 = \min\{\frac{X_1}{X_0}, \ldots, \frac{X_p}{X_0}\}$, $W_1 = \max\{\frac{Y_1}{Y_0}, \ldots, \frac{Y_p}{Y_0}\}$ and $W_2 = \min\{\frac{Y_1}{Y_0}, \ldots, \frac{Y_p}{Y_0}\}$, we get

$$\begin{aligned} &\Pr\{Z_1 \leq a\} \geq \Pr\{W_1 \leq a\}, \\ (3.14) \qquad &\Pr\{Z_1 \leq \tfrac{1}{a}\} \leq \Pr\{W_1 \leq \tfrac{1}{a}\}, \\ &\Pr\{Z_2 \geq a\} \leq \Pr\{W_2 \geq a\}, \\ &\Pr\{Z_2 \geq \tfrac{1}{a}\} \geq \Pr\{W_2 \geq \tfrac{1}{a}\}. \end{aligned}$$

In other words, we have inequalities for the distribution functions (and hence for quantiles) of the maximum and the minimum of certain correlated ratios of variables with distributions F and G.

In the case of $\mathcal{H} = \{x+b,\ b \geq 0\}$, we let $Z_1' = \max\{X_1-X_0, \ldots, X_p-X_0\}$, $Z_2' = \min\{X_1-X_0, \ldots, X_p-X_0\}$, $W_1' = \max\{Y_1-Y_0, \ldots, Y_p-Y_0\}$ and $W_2' = \min\{Y_1-Y_0, \ldots, Y_p-Y_0\}$. Then, we get

$$\text{(3.15)} \quad \begin{cases} \Pr\{Z_1' \leq b\} \geq \Pr\{W_1' \leq b\}, \\ \Pr\{Z_1' \leq -b\} \leq \Pr\{W_1' \leq -b\}, \\ \Pr\{Z_2' \geq b\} \leq \Pr\{W_2' \geq b\}, \\ \Pr\{Z_2' \geq -b\} \geq \Pr\{W_2' \geq -b\}. \end{cases}$$

We will come back to these inequalities in Section 4.3. □

4. INEQUALITIES FOR SPECIFIC DISTRIBUTIONS

We are mainly interested in certain inequalities relating to multivariate normal, multinomial and gamma distributions that occur in ranking and selection problems. Of course, these are of interest otherwise too.

4.1 Inequalities for Multivariate Normal Distribution. A probability expression that occurs frequently in selection problems is $\Pr[X_1 \leq a_1, \ldots, X_k \leq a_k]$ where $X_1, X_2, \ldots, X_k$ are identically distributed but correlated. Most familiar of these and perhaps most often used in practice are the cases where $X_1, \ldots, X_k$ have a joint k-variate normal and t distributions. Evaluation of these probability integrals are difficult to accomplish as k gets large when there is no special pattern of the associated covariance matrix Σ. In such cases, inequalities which give good bounds become more attractive. There are numerous results in the literature in this direction. We will mention here only two results, namely, those of Anderson [6] and Slepian [58]. For a detailed account of these and other related inequalities and references, the reader is referred to the book of Tong [61] and the recent survey paper of Eaton [19]. To state Anderson's theorem, let us define a partial ordering $\underset{\sim}{<}$ for covariance matrices of the same order by $\Psi \underset{\sim}{<} \Sigma$ if $\Sigma - \Psi$ is positive semidefinite.

Theorem 4.1 (Anderson [6]). Let $\underline{X} = (X_1, \ldots, X_k)$ and $\underline{Y} = (Y_1, \ldots, Y_k)$ be k-variate normally distributed random vectors with common mean vector zero and covariance matrices Σ and Ψ respectively and let E be a convex set symmetric about the origin. Then $\Psi \underset{\sim}{<} \Sigma$ implies $\Pr[\underline{Y} \in E] \geq \Pr[\underline{X} \in E]$.

As we have pointed out earlier, inequalities have been used in selection problems typically to obtain the infimum of the PCS or a lower bound for it. One result that has been used very often at some stage of the problem is the Slepian inequality stated below.

Theorem 4.2 (Slepian Inequality). If $\underline{X} = (X_1, \ldots, X_k)$ has the k-variate normal distribution with nonsingular covariance matrix $\Sigma = (\sigma_{ij})$, with $\sigma_{ii} = 1$, $i=1,\ldots,k$, then for any constants $c_1,\ldots,c_k$, the probability $\Pr\{X_1 \leq c_1,\ldots,X_k \leq c_k\}$ is strictly increasing as a function of each σ_{ij} for $i \neq j$. In particular, if $\sigma_{ij} > 0$, $i, j = 1,\ldots,k$, then
$$\Pr[X_i \leq c_i,\ i=1,\ldots,k] > \prod_{i=1}^{k} \Pr[X_i \leq c_i].$$

Motivated by a design problem with a selection and ranking goal, Rinott and Santner [55] obtained an inequality that combines the aspects of the results of Anderson and Slepian; namely, for $d > 0$

$$(4.1) \qquad \int\int \Phi^n(d+x+\alpha y)\, \Phi^m(d+x)\, d\Phi(x)\, d\Phi(y) \leq \int \Phi^{n+m}(d+x)\, d\Phi(x)$$

where $\Phi(x)$ is the standard normal cdf, m and n are integers such that $m+1 \geq n \geq 1$, and all integrals are from $-\infty$ to ∞. It can also be shown that the left-hand side of (2.8) is decreasing in $|\alpha|$ for any $d \geq 0$.

4.2 Inequalities for Multinomial Distributions.

Let $\underline{X} = (X_1, \ldots, X_k)$ have the multinomial distribution given by

(4.2) $$\Pr\{\underline{X} = \underline{x}\} = n! \prod_{i=1}^{k} (\theta_i^{x_i} / x_i!)$$

where $\underline{x} = (x_1, \ldots, x_k), \sum_{i=1}^{k} x_i = n$ and $\sum_{i=1}^{k} \theta_i = 1$.

Define

(4.3) $$C(\theta_1, \ldots, \theta_m) = \Pr\{X_i \geq c_i, i=1, \ldots, m\}$$

where $\sum_{i=1}^{m} c_i \leq n$ and $m \leq \min(k-1, n)$. The results of Alam [1] are summarized in the following theorem.

Theorem 4.3 $C(\theta_1, \ldots, \theta_m)$ is nondecreasing in θ_i, $i=1, 2, \ldots, m$. Further, for $c_i = c_j$,

(4.4) $$C_{ijt}(\theta_1, \ldots, \theta_m) \leq C(\theta_1, \ldots, \theta_m) \leq C_{ij}(\theta_1, \ldots, \theta_m)$$

where $C_{ij}(\theta_1, \ldots, \theta_m)$ is obtained from $C(\theta_1, \ldots, \theta_m)$ by replacing θ_i and θ_j with their average, and $C_{ijt}(\theta_1, \ldots, \theta_m)$ is obtained from $C(\theta_1, \ldots, \theta_m)$ by substituting t for θ_i and $\theta_i+\theta_j-t$ for θ_j where $0 \leq t \leq \min(\theta_i, \theta_j)$.

Let us assume here and in what follows on multinomial distribution that $\theta_1 \leq \theta_2 \leq \ldots \leq \theta_k$. From Theorem 4.3, we have

(4.5) $$\begin{aligned} &\Pr\{X_1 \geq c, \ldots, X_k \geq c \mid \theta_1, \ldots, \theta_1, \theta^*\} \\ &\leq \Pr\{X_1 \geq c, \ldots, X_k \geq c \mid \theta_1, \ldots, \theta_k\} \\ &\leq \Pr\{X_1 \geq c, \ldots, X_k \geq c \mid \bar{\theta}, \ldots, \bar{\theta}\} \end{aligned}$$

where $c \leq n/k$, $\theta^* = 1-(k-1)\theta_1$ and $\bar{\theta} = \Sigma\theta_i/k$.

Using a representation of $\Pr\{X_1 \geq c, \ldots, X_k \geq c \mid \theta_1, \ldots, \theta_k\}$ in terms of the Dirichlet integral, the inequalities in (4.5) can be obtained as a special case of Theorem 1 of Olkin [51] which shows the Dirichlet integral to be a Schur function. More general results are available in Marshall and Olkin ([47], p. 306).

Bechhofer, Elmaghrabi and Morse [10] considered a single sample selection procedure to select the most probable cell with a minimum guaranteed probability P* that the selected cell will be the one associated with θ_k whenever $\theta_k/\theta_{k-1} \geq \delta > 1$. The rule R proposed by Bechhofer, Elmaghrabi and Morse takes a sample of N observations and selects the cell that yields the largest number of observations using randomization to break ties. The PCS is given by

$$\text{(4.6)} \quad \begin{aligned} \text{PCS} &= \Pr\{X_k > X_j,\ j \neq k\} + 1/2 \sum_{i \neq k} \Pr\{X_k = X_i,\ X_k > X_j,\ j \neq i\} \\ &\quad + \ldots + 1/k \Pr\{X_k = X_{k-1} = \ldots = X_1\} \\ &= \Psi(\theta_1, \theta_2, \ldots, \theta_k), \text{ say.} \end{aligned}$$

The following result of Kesten and Morse [41] gives the LFC.

<u>Theorem 4.4</u> With the above assumptions and notations,

$$\text{(4.7)} \qquad \Psi(\theta_1, \ldots, \theta_k \mid \theta_k/\theta_{k-1} \geq \delta > 1) \geq \Psi(\theta_1^*, \ldots, \theta_k^*)$$

where $\theta_1^* = \ldots = \theta_{k-1}^* = (\delta+k-1)^{-1}$ and $\theta_k^* = \delta(\delta+k-1)^{-1}$.

Cacoullos and Sobel [16] used an inverse sampling rule for the same selection problem. Observations are obtained sequentially until one of the k cells has a prespecified count N. This particular cell is then identified as the most probable cell. In this case, the PCS can be written as a Dirichlet integral and the LFC is the same as that of the single sample procedure of Bechhofer, Elmaghrabi and Morse [10]. Alam [3] considered a different stopping rule, namely, the observations are taken sequentially until the difference between the highest and the next highest cell count is equal to r. For k=2,

$$\text{PCS} = \lambda^r/(1+\lambda^r) \tag{4.8}$$

where $\lambda = \theta_2/\theta_1$. For $k > 2$, there is no exact result. Alam [3] gives a lower bound, namely,

$$\text{PCS} \geq 1 - \sum_{i=1}^{k-1} \lambda_i^r / (1+\lambda_i^r) \tag{4.9}$$

where $\lambda_i = \theta_i/\theta_k$, i=1, ..., k-1. An improved bound, namely, $\theta_k^r / \sum_1^k \theta_i^r$, is recently given by Levin and Robbins [45].

Going back to the single sample procedure of Bechhofer, Elmaghraby and Morse [10] for selecting the most probable cell, the LFC is sought subject to $\theta_k/\theta_{k-1} \geq \delta > 1$. If we are interested in selecting the least probable cell, then the analogous problem will be to get the LFC whenever $\theta_2/\theta_1 \geq \delta > 1$. The analogous procedure will select the cell with the least count using randomization to break ties. In this case, a minimum P* for the PCS cannot be guaranteed for all P*. This is shown by Alam and Thompson [5] who proposed a modified indifference-zone. Their rule is still to select the cell with the least count. Let $\Psi'(\theta_1, \ldots, \theta_k)$ denote the PCS for this rule. Then their LFC result can be stated as follows:

$$\Psi'(\theta_1, \ldots, \theta_k | \theta_2 - \theta_1 \geq c) \geq \Psi'(\theta_1^*, \ldots, \theta_k^*) \tag{4.8}$$

where $0 < c < (k-1)^{-1}$, $\theta_1^* = [1-(k-1)c]/k$, and $\theta_2^* = \ldots = \theta_k^* = (1+c)/k$.

We get additional probability inequalities via subset selection rules. Gupta and Nagel [27] discussed single sample subset slection rules for selecting the most (least) probable cell. If we denote the cell counts by $X_1, \ldots, X_k$, their rules R_1 and R_2 for the most and the least probable cell, respectively, are as follows:

Select the cell with count X_i if and only if

$$R_1: \qquad X_i \geq \max(X_1, \ldots, X_k) - d$$
$$R_2: \qquad X_i \leq \min(X_1, \ldots, X_k) + c$$

where c and d are nonnegative integers chosen suitably to guarantee the specified minimum PCS.

The PCS for R_1 is given by

(4.9) $$P(CS \mid R_1) = F(k,n,d;\ \theta_1, \ldots, \theta_k) = \Sigma(\nu_1 \overset{n}{\ldots} \nu_k)\ \theta_1^{\nu_1} \ldots \theta_k^{\nu_k}$$

where the summation is over all k- tuples $(\nu_1, \ldots, \nu_k)$ such that the ν_i are nonnegative, $\Sigma\nu_i = n$ and $\nu_i \leq \nu_k+d$, $i=1, \ldots, k-1$. In the case of R_2, $P(CS|R_2) = G(k,n,c;\theta_1, \ldots, \theta_k)$ is given by the summation in (4.9) extending over k- tuples $(\nu_1, \ldots, \nu_k)$ such that the ν_i are nonnegative, $\Sigma\nu_i = n$ and $\nu_i \geq \nu_1 - c$, $i=2, \ldots, k$.

We now summarize the inequality results of Gupta and Nagel [27] in the following lemmas and theorems.

<u>Lemma 4.5</u> $F(k,n,d;\ \theta_1, \ldots, \theta_k)$ satisfies the following inequalities:

(1) For $1 \leq i < j < k$, and $0 < \epsilon \leq \theta_i$,

$$F(k,n,d;\ \theta_1, \ldots, \theta_k) \geq F(k,n,d;\ \theta_1, \ldots, \theta_i-\epsilon, \ldots, \theta_j+\epsilon, \ldots, \theta_k).$$

(2) For $1 \leq i < k$, and $0 < \epsilon \leq \theta_k$,

$$F(k,n,d;\ \theta_1, \ldots, \theta_k) \geq F(k,n,d;\ \theta_1, \ldots, \theta_i+\epsilon, \ldots, \theta_k-\epsilon).$$

It should be noted that Lemma 4.5 is true even if the order is disturbed in the configurations on the right hand side of the inequalities. The next theorem on the LFC is a consequence of Lemma 4.5.

Theorem 4.6 Let r be the smallest integer for which $\theta_i > 0$ and let s be the largest integer such that $\theta_j < \theta_k$. For a configuration minimizing $F(k,n,d;\ \theta_1, \ldots, \theta_k)$, we have $r \geq s$. Furthermore, if $r = k-1$, then $r > s$.

In other words, Theorem 4.6 says that the worst configuration is of the type $(0, \ldots, 0, \alpha, \beta, \ldots, \beta)$, $\alpha \leq \beta$.

Lemma 4.7 $G(k,n,c;\ \theta_1, \ldots, \theta_k)$ satisfies the following inequalities:

(1) For $1 < i < j \leq k$ and $0 < \epsilon \leq \theta_i$,

$G(k,n,c;\ \theta_1, \ldots \theta_k) \geq G(k,n,c; \theta_1, \ldots, \theta_i - \epsilon, \ldots, \theta_j + \epsilon, \ldots, \theta_k)$.

(2) For $1 < j \leq k$ and $0 < \epsilon \leq \theta_j$,

$G(k,n,c;\ \theta_1, \ldots, \theta_k) \geq G(k,n,c;\ \theta_1 + \epsilon, \ldots, \theta_j - \epsilon, \ldots, \theta_k)$.

As in the case of Lemma 4.5, here also the statements are true even if the order is disturbed in the configuration. The following theorem is a consequence of Lemma 4.7.

Theorem 4.8 $G(k,n,c;\ \theta_1, \ldots, \theta_k)$ is minimized at a configuration of the type $\theta_1 = \ldots = \theta_{k-1} \leq \theta_k$.

Now, let us consider ℓ independent multinomial distributions each with k cells. Let $\underline{\theta}_i = (\theta_{i1}, \ldots, \theta_{ik})$ be the vector of the cell probabilities of Π_i, the ith distribution,, i=1, ..., m. We also assume that, for each i, $\theta_{i1} \leq \cdots \leq \theta_{ik}$.

Definition 4.9. We say that $\underline{\theta}_i$ majorizes $\underline{\theta}_j (\underline{\theta}_i \underset{m}{\geq} \underline{\theta}_j)$ if

$\sum_{\alpha=r}^{k} \theta_{i\alpha} \geq \sum_{\alpha=r}^{k} \theta_{j\alpha}$ for r = 1, ..., k with equality holding for r = 1.

Definition 4.10 If a function φ satisfies the property that $\varphi(\underline{x}) \le \varphi(\underline{y})$ $(\varphi(\underline{x}) \ge \varphi(\underline{y}))$ whenever $\underline{x} \underset{m}{>} \underline{y}$, then φ is called a Schur-concave (Schur-convex) function.

If $\underline{\theta}_i \underset{m}{>} \underline{\theta}_j$, it implies that $H(\underline{\theta}_i) \le H(\underline{\theta}_j)$, where $H(\underline{\theta}_i) = -\sum_{\alpha=1}^{k} \theta_{i\alpha} \log \theta_{i\alpha}$ is the Shannon entropy function associated with Π_i.

Suppose we take n independent observations from each multinomial distribution. Let $x_{i\alpha}$ denote the number of outcomes in the cell with probability $\theta_{i\alpha}$ in Π_i, $\alpha = 1, \ldots, k$; $i=1, \ldots, \ell$. Define

$$(4.10) \qquad Q_j(n,k,\ell;\ \underline{\theta}_1, \ldots, \underline{\theta}_\ell)$$

$$= \Pr\left\{\varphi\left(\frac{X_{j1}}{n}, \ldots, \frac{X_{jk}}{n}\right) \ge \max_{1\le\alpha\le\ell} \varphi\left(\frac{X_{\alpha 1}}{n}, \ldots, \frac{X_{\alpha k}}{n}\right) - d\right\},\ j = 1, \ldots, \ell,$$

where φ is a Schur-concave function and $d > 0$.

Gupta and Wong [34] investigated a subset selection rule for selecting the population whose cell probability vector majorizes that of any other, assuming that one such exists. The special case of $k = 2$ multinomial distributions with the Shannon entropy function as a particular choice of φ was earlier considered by Gupta and Huang [24]. The following theorem relates to the properties of the procedure of Gupta and Wong [34].

Theorem 4.11. If $\underline{\theta}_i \underset{m}{>} \underline{\theta}_j$, then $Q_i(n,k,\ell;\ \underline{\theta}_1, \ldots, \underline{\theta}_\ell) \le Q_j(n,k,\ell;\ \underline{\theta}_1, \ldots, \underline{\theta}_\ell)$. Further, if $\underline{\theta}_i \underset{m}{>} \underline{\theta}_j$ for all $j=1, \ldots, \ell$, then $Q_i(n,k,\ell;\ \underline{\theta}_1 \ldots \underline{\theta}_\ell) \ge Q_i(n,k,\ell;\ \underline{\theta}_1 = \ldots = \underline{\theta}_\ell)$.

4.3 Inequalities for the Gamma Distribution

Let

$$\gamma(m,x) = \int_0^x t^{m-1} e^{-t}\, dt \tag{4.11}$$

and

$$\Gamma(m,x) = \Gamma(m) - \gamma(m,x), \quad m > 0. \tag{4.12}$$

Of course,

$$f(x;m) = \frac{e^{-t} t^{m-1}}{\Gamma(m)}, \quad x \geq 0, \; m > 0, \tag{4.13}$$

is the gamma density where m is the shape parameter. For $0 < m < 1$, continued fraction expansions can be obtained (see, for example, Khovanskii [42]) for $x^{-m} e^{x} \gamma(m,x)$ and $x^{-m} e^{x} \Gamma(m,x)$. Let $P_n(m,x)/Q_n(m,x)$ and $P_n'(m,x)/Q_n'(m,x)$ be the nth convergents of these two expansions respectively.

In the case of $\gamma(m,x)$, Gupta and Waknis [33] obtained the system of inequalities:

$$\frac{P_n(m,x)}{Q_n(m,x)} < e^{x} x^{-m} \gamma(m,x) < \frac{P_n(m,x)}{Q_n(m,x)} + \frac{x^n (n+1+m)}{(n+m)_{n+1} (n+1+m-x)}, \quad n = 1, 2, \ldots, \tag{4.14}$$

where $x < n+m+1$ is a necessary restriction only on the inequalities on the right-hand side of (4.14) and where $(n)_r = n(n-1) \ldots (n-r+1)$, $r \geq 1$, and

$$\frac{P_n(m,x)}{Q_n(m,x)} = \frac{1}{m} \left[1 + \frac{x}{1+m} + \frac{x^2}{(1+m)\,(2+m)} + \ldots + \frac{x^{n-1}}{(1+m)\,\ldots\,(n-1+m)} \right]. \tag{4.15}$$

In the case of $\Gamma(a,x)$, the even order convergents form a monotonic increasing sequence and the odd order convergents form a monotonic decreasing sequence, both converging to $e^{x} x^{-m} \Gamma(m,x)$. So a system of inequalities can be generated by bounding $e^{x} x^{-m} \Gamma(m,x)$ by successive convergents. These bounds are discussed in Gupta and Waknis [33]. These bounds in turn can be used to get bounds on the integrals

$$(4.16) \qquad \int_0^\infty F^p(cx;m)\ f(x;m)dx$$

and

$$(4.17) \qquad \int_0^\infty [1-F(bx;m)]^p\ f(x;m)dx$$

where $F(x;m)$ is the cdf of the gamma distribution. The integrals (4.16) and (4.17) with $c \geq 1$ and $0 < b \leq 1$ are the infima of the PCS for the subset selection rules of Gupta [21] and Gupta and Sobel [32].

Now, let $X_0, X_1, \ldots, X_p$ be independent identically distributed each having a gamma distribution with density $f(x;m)$ given by (4.13). Let

$$(4.18) \qquad \begin{cases} Z_1 = \max(\frac{X_1}{X_0}, \ldots, \frac{X_p}{X_0}), \\ Z_2 = \min(\frac{X_1}{X_0}, \ldots, \frac{X_p}{X_0}). \end{cases}$$

Let $G_m(y)$ and $H_m(y)$ denote the cdf's of Z_1 and Z_2, respectively. We note that the integrals in (4.16) and (4.17) are $G_m(c)$ and $1-H_m(b)$, respectively. Alam [2] proved that, for $m > 1$, $H_m(y)$ is increasing in m for $y > 1$ and is decreasing in m for $y < 1$. Alam's proof involves a fair amount of analytical details. Further, Alam has no comment on the behavior of $G_m(y)$. The following theorem provides validity of Alam's result for $m > 0$ and establishes the monotonicity behavior of G_m and H_m for a larger class of distributions.

<u>Theorem 4.12.</u> Let $X_0, X_1, \ldots, X_p$ be i.i.d. nonnegative random variables each having the distribution F_λ, where $\{F_\lambda\}$ is a star-preceding family in $\lambda \in \Lambda$ [i.e., $F_{\lambda_2} \underset{*}{<} F_{\lambda_1}$ for $\lambda_1 < \lambda_2$]. Let G_λ and H_λ be the cdf's of Z_1 and Z_2 defined in (4.18). Then $G_\lambda(y)$ and $H_\lambda(y)$ are both increasing in λ for $y > 1$ and decreasing in λ for $y < 1$.

Proof. Since $F_{\lambda_2} \underset{*}{<} F_{\lambda_1}$ for $\lambda_1 < \lambda_2$, the conclusions of the theorem follow immediately from the inequalities (3.14) of Remarks 3.6. □

Remarks 4.13. In the case of the gamma family $\{F_m\}$, it is known that F_m convex precedes in $m > 0$; see van Zwet [60], p. 60. Since the convex ordering implies the star ordering, Alam's result readily follows from Theorem 4.12. As we pointed out earlier, in subset selection procedures, we typically encounter $G_m(y)$ for $y < 1$ and $H_m(y)$ for $y > 1$. That the monotonicity properties of $G_m(y)$ and $H_m(y)$ in these cases can be established by the star-ordering property of the gamma distribution was known though not formally demonstrated; see McDonald [48] and Panchapakesan [53] who have given different alternative proofs in the case of integral m for $p = 1$ and $p \geq 1$ respectively. Finally, the monotonicity property of $H_m(y)$ is applied to evaluate the infimum of the PCS for the inverse sampling procedure of Cacoullos and Sobel [16] for selecting the most probable multinomial cell. □

For the Gamma distribution with density in (4.13), let $\xi_m(\alpha)$ and $\xi_m(\beta)$ denote the αth and the βth quantiles, where $0 < \alpha < \beta < 1$. For $m_1 < m_2$, as pointed out earlier, $F_{m_2} \underset{*}{<} F_{m_1}$. This is equivalent to

$$(4.19) \qquad \frac{F_{m_1}^{-1}(\beta)}{F_{m_1}^{-1}(\alpha)} \geq \frac{F_{m_2}^{-1}(\beta)}{F_{m_2}^{-1}(\alpha)} ;$$

in other words, $\xi_m(\beta) / \xi_m(\alpha)$ decreases in m, a result obtained by Saunders and Moran [56] using a fairly long direct method. They have also shown that, for $m_1 < m_2$, F_{m_2} is more dispersed than F_{m_1}; in other words, $\xi_m(\beta) - \xi_m(\alpha)$ increases in m. Also, we can now apply the inequalities in (3.15) to obtain new inequalities for the distribution functions of the maximum and the minimum of certain correlated differences.

4.4 Inequalities Arising From A Two Stage Selection Procedure.

Gupta and Miescke [26] studied sequential selection procedures with elimination which are based on vector-at-a-time sampling. They showed that the 'natural' terminal decisions are optimum in a fairly decision-theoretic sense. To decribe the inequalities that are obtained, let $\Pi_1, \ldots, \Pi_k$ be k independent populations with densities f_{θ_i}, $\theta_i \in \Omega$, with respect to the Lebesgue measure on the real line $\mathbb{R}$ or any counting measure on a lattice in $\mathbb{R}$, where $\mathfrak{F} = \{f_\theta\}$, $\theta \in \Omega$, is a one-parameter exponential family. Let $X_{i1}, X_{i2}, \ldots$ be independent observations from Π_i, i=1, ..., k. For fixed $n < m$, let $U_i = X_{i1} + \ldots + X_{in}$, $V_i = X_{i,n+1} + \ldots + X_{i,m}$, and $W_i = U_i + V_i$, i=1, ..., k. Further, for fixed $s \subseteq \{1, \ldots, k\}$, permutation symmetric Borel set $A \subseteq \mathbb{R}^k$, and $i \in s$, define

$$(4.20)\qquad \begin{cases} q_i = P_{\underline{\theta}}\{V_i = \max\limits_{j \in s} V_j\}, \\[2ex] r_i = P_{\underline{\theta}}\{W_i = \max\limits_{j \in s} W_j \mid (U_1, \ldots, U_k) \in A\}. \end{cases}$$

Theorem 4.13 For $s = \{i_1, \ldots, i_m\}$

(1) $\theta_{i_j} \le \theta_{i_\ell}$ implies that $r_{i_j} \le r_{i_\ell}$ and $q_{i_j} \le q_{i_\ell}$, $j, \ell = 1, \ldots, m;\ j \neq \ell$, and

(2) the vector $\underline{r} = (r_{i_1}, \ldots, r_{i_m})$ majorizes the vector $\underline{q} = (q_{i_1}, \ldots, q_{i_m})$.

4.5 An Ordering Theorem and Its Specific Applications

Let $X_1, \ldots, X_p$ be conditionally independent and identically distributed random variables, that is, their joint distribution F is a mixture of the form

$$(4.21)\qquad F(x_1,x_2,\ldots,x_p) = \int \prod_{i=1}^{p} F_1(x_i,z)dF_2(z)$$

$$= E[\prod_{i=1}^{p} F_1(x_i,Z)],$$

where F_1 (for given z) and F_2 are distribution functions. The following theorem is due to Tong [59].

Theorem 4.14. Let $\underset{\sim}{a} = (a_1,a_2,\ldots,a_r)$ and $\underset{\sim}{b} = (b_1,b_2,\ldots,b_r)$ be vectors of nonnegative integers such that $a_1 \geq a_2 \geq \ldots \geq a_r$ and $b_1 \geq b_2 \geq \ldots \geq b_r$ with $\sum_{j=1}^{r} a_j = \sum_{j=1}^{r} b_j = p$. If $X_1,\ldots,X_p$ are conditionally i.i.d. random variables and if $\underset{\sim}{a} \underset{m}{>} \underset{\sim}{b}$, then

$$(4.22)\qquad \prod_{j=1}^{r} \Pr\{X_i \in A,\ i = 1,\ldots,a_j\} \geq \prod_{j=1}^{r} \Pr\{X_i \in A,\ i = 1,\ldots,b_j\}$$

holds for every Borel measurable set A.

Now, if $Y_1,Y_2,\ldots,Y_p$ are i.i.d. random variables and Z is independent of the Y_i, then it is known (see Tong [60], Theorem 2) that $X_i = \phi(Y_i,Z)$, $i = 1,2,\ldots,p$, are conditionally i.i.d. for any Borel measurable function ϕ. This fact together with Theorem 4.14 can be used to obtain bounds on the PCS under the indifference zone formulation and the subset selection approach in view of the fact that the PCS for many classical rules (see Gupta and Panchapakesan [28]) is a cumulative probability of conditionally i.i.d. random variables.

Tong [59]) has also discussed a special form of Theorem 4.14 and its applications to several specific multivariate distributions. Applications to multiple decision situations besides selection and ranking are discussed by Tong [60].

[1] Alam, K. (1970 a). Monotonicity properties of the multinomial distribution. Ann. Math. Statist. 41, 315-317.

[2] Alam, K. (1970 b). A monotonicity property of the distribution of the studentized smallest chi-square. Ann. Math. Statist. 41, 318-320.

[3] Alam, K. (1971). On selecting the most probable category. Technometrics 13, 843-850.

[4] Alam, K. and Rizvi, M. H. (1966). Selection from multivariate populations. Ann. Inst. Statist. Math. 18, 307-318.

[5] Alam, K. and Thompson, J. R. (1972). On selecting the least probable multinomial event. Ann. Math. Statist. 43, 1981-1990.

[6] Anderson, T. W. (1955). The integral of a symmetric unimodal function over a symmetric convex set and some probability inequalities. Proc. Amer. Math. Soc. 6, 170-176.

[7] Barlow, R. E. and Gupta, S. S. (1969). Selection procedures for restricted families of distributions. Ann. Math. Statist. 40, 905-917.

[8] Barlow, R. E., Gupta, S. S. and Panchapakesan, S. (1969). On the distribution of the maximum and minimum of ratios of order statistics. Ann. Math. Statist. 40, 918-934.

[9] Barlow, R. E., Marshall, A. W. and Proschan, F. (1963). Properties of probability distributions with monotone hazard rate. Ann. Math. Statist. 34, 375-389.

[10] Bechhofer, R. E., Elmaghrabi, S. and Morse, N. (1959). A single-sample multiple-decision procedure for selecting the multinomial event which has the highest probability. Ann. Math. Statist. 30, 102-119.

[11] Bechhofer, R. E., Kiefer, J. and Sobel, M. (1968). Sequential Identification and Ranking Procedures. The University of Chicago Press, Chicago.

[12] Beckenbach, E. F. and Bellman, R. (1961,1965). Inequalities, 1st ed., 2nd ed. Springer-Verlag, Berlin and New York.

[13] Bickel, P. J. and Lehmann, E. L. (1979). Descriptive statistics for nonparametric models IV. Spread. Contributions to Statistics: Jaroslav Hajek Memorial Volume (Jana Jureckova, ed.), D. Reidel Publishing Co., Boston, 33-40.

[14] Birnbaum, Z. W., Esary, J. D. and Marshall, A. W. (1966). A stochastic characterization of wear-out for components and systems. Ann. Math. Statist. 37, 816-825.

[15] Brown, G. and Tukey, J. W. (1946). Some distributions of sample means. Ann. Math. Statist. 7, 1-12.

[16] Cacoullos, T. and Sobel, M. (1966). An inverse-sampling procedure for selecting the most probable event in a multinomial distribution. Multivariate Analysis (P. R. Krishnaiah, ed.), Academic Press, New York, 423-455.

[17] Doksum, K. (1969). Starshaped transformations and the power of rank tests. Ann. Math. Statist. 40, 1167-1176.

[18] Eaton, M. L. (1967). Some optimum properties of ranking procedures. Ann. Math. Statist. 38, 124-137.

[19] Eaton, M. L. (1982). A review of selected topics in multivariate probability inequalities. Ann. Statist. 10, 11-43.

[20] Godwin, H. A. (1964). Inequalities on Distribution Functions. Griffin's Statistical Monographs and Courses, No. 16. Hafner Publishing Company, New York.

[21] Gupta, S. S. (1963). On a selection and ranking procedure for gamma populations. Ann. Inst. Statist. Math. 14, 199-216.

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

[23] Gupta, S. S. (1966). On some selection and ranking procedures for multivariate normal populations using distance functions. Multivariate Analysis (P. R. Krishnaiah, ed.), Academic Press, New York, 457-475.

[24] Gupta, S. S. and Huang, D-Y. (1976). On subset selection procedures for the entropy function associated with the binomial populations. Sankhyā Ser. A 38, 153-173.

[25] Gupta, S. S. and Huang, D-Y. (1980). An essentially complete class of multiple decision procedures. J. Statist. Planning and Inference 4, 115-121.

[26] Gupta, S. S. and Miescke, K-J. (1982). On the problem of finding a best population with respect to a control in two stages. Statistical Decision Theory and Related Topics III, Vol. 2 (S. S. Gupta and J. Berger, eds.), Academic Press, New York, 473-496.

[27] Gupta, S. S. and Nagel, K. (1967). On selection and ranking procedures and order statistics from the multinomial distribution. Sankhyā Ser. B 29, 1-34.

[28] Gupta, S. S. and Panchapakesan, S. (1972). On a class of subset selection procedures. Ann. Math. Statist. 43, 814-822.

[29] Gupta, S. S. and Panchapakesan, S. (1974). Inference for restricted families: (a) multiple decision procedures; (b) order statistics inequalities. Reliability and Biometry: Statistical Analysis of Lifelength (F. Proschan and R. J. Serfling, eds.), SIAM, Philadelphia, 503-596.

[30] Gupta, S. S. and Panchapakesan, S. (1975). On a quantile selection procedure and associated distribution of ratios of order statistics from a restricted family of probability distributions. Reliability and Fault Tree Analysis: Theoretical and Applied Aspects of System Reliability and Safety Assessment (R. E. Barlow, J. B. Fussell and N. D. Singpurwalla, eds.), SIAM, Philadelphia, 557-576.

[31] Gupta, S. S. and Panchapakesan, S. (1979). Multiple Decision Procedures: Theory and Methodology of Selecting and Ranking Populations. Wiley, New York.

[32] Gupta, S. S. and Sobel, M. (1962). On selecting a subset containing the population with the smallest variance. Biometrika 49, 495-507.

[33] Gupta, S. S. and Waknis, M. N. (1965). A system of inequalities for the incomplete gamma function and the normal integral. Ann. Math. Statist. 36, 139-149.

[34] Gupta, S. S. and Wong, W-Y. (1976). Subset selection procedures for finite schemes in information theory. Colloquia Mathematica Societatis János Bolyai: 16. Topics in Information Theory, Conference held at Keszthely, August 1975, 279-291.

[35] Hardy, G. H., Littlewood, J. E. and Pólya, G. (1934,1952). Inequalities, 1st ed., 2nd ed. Cambridge University Press, London and New York.

[36] Hollander, M., Proschan, F. and Sethuraman, J. (1977). Functions decreasing in transposition and their applications in ranking problems. Ann. Statist. 5, 722-733.

[37] Hsu, J. C. (1977). On Some Decision-Theoretic Contributions to the Problem of Subset Selection. Ph.D. Thesis. Department of Statistics, Purdue University, West Lafayette, Indiana.

[38] Karlin, S. (1968). Total Positivity, Vol. I. Stanford University Press, Stanford.

[39] Karlin, S. and Rubin, H. (1956). Distributions possessing a monotone likelihood ratio. J. Amer. Statist. Assoc. 51, 637-643.

[40] Kazarinoff, N. D. (1961). Geometric Inequalities. Random House, New Mathematical Library, 4, New York.

[41] Kesten, H. and Morse, N. (1959). A property of the multinomial distribution. Ann. Math. Statist. 30, 120-127.

[42] Khovanskii, A. N. (1956). The Application of Continued Fractions. (Translated by P. Wynn (1962)). P. Noordhoff Ltd., Groningen.

[43] Lehmann, E. L. (1955). Ordered families of distributions. Ann. Math. Statist. 26, 399-419.

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

[45] Levin, B. and Robbins, H. (1981). Selecting the highest probability in binomial or multinomial trials. Private Communication.

[46] Mahamunulu, D. M. (1967). Some fixed-sample ranking and selection problems. Ann. Math. Statist. 38, 1079-1091.

[47] Marshall, A. W. and Olkin, I. (1979). Inequalities: Theory of Majorization and Its Applications. Academic Press, New York.

[48] McDonald, G. C. (1969). A note on monotonicity in the gamma distributions. Res. Publ. GMR-930. General Motors Research Laboratories, Warren, Michigan.

[49] Mitrinović, D. S. (1964). Elementary Inequalities. P. Noordhoff, Groningen.

[50] Mitrinović, D. S. (1970). Analytic Inequalities. Springer-Verlag, Berlin and New York.

[51] Olkin, I. (1972). Monotonicity properties of Dirichlet integrals with applications to the multinomial distribution and the analysis of variance. Biometrika 59, 303-307.

[52] Panchapakesan, S. (1969). Some Contributions to Multiple Decision (Selection and Ranking) Procedures. Ph.D. Thesis. (Mimeo. Ser. No. 192). Dept. of Statistics, Purdue University, West Lafayette.

[53] Panchapakesan, S. (1978). On a monotonicity property relating to the gamma distribution. J. Chinese Statist. Assoc. 18, 6003-6005.

[54] Pólya, G. and Szegö, G. (1951). Isoperimetric Inequalities in Mathematical Physics. Ann. of Math. Studies, No. 27, Princeton University Press, Princeton.

[55] Rinott, Y. and Santner, T. J. (1977). An inequality for multivariate normal probabilities with application to a design problem. Ann. Statist. 5, 1228-1234.

[56] Saunders, I. W. and Moran, P.A.P. (1978). On the quantiles of the gamma and F distributions. J. Appl. Prob. 15, 426-432.

[57] Shisha, O. (Ed.) (1967). Inequalities. Academic Press, New York.

[58] Slepian, D. (1962). The one-sided barrier problem for Gaussian noise. Bell System Tech. J. 41, 463-501.

[59] Tong, Y. L. (1977a). An ordering theorem for conditionally independent and identically distributed random variables. Ann. Statist. 5, 274-277.

[60] Tong, Y. L. (1977b). Applications of a probability inequality to ranking and selection and other related problems. Commun. Statist.-Theor. Meth. A6, 1105-1120.

[61] Tong, Y. L. (1980). Probability Inequalities in Multivariate Distributions. Academic Press, New York.

[62] Van Zwet, W. R. (1964). Convex Transformations of Random Variables. Mathematical Center, Amsterdam.

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: Technical Report #82-37	2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle): ON SOME INEQUALITIES AND MONOTONICITY PROPERTIES WITH SPECIAL REFERENCE TO SELECTION AND RANKING PROBLEMS	5. TYPE OF REPORT & PERIOD COVERED: Technical
	6. PERFORMING ORG. REPORT NUMBER: Technical Report #82-37
7. AUTHOR(s): Shanti S. Gupta, Deng-Yuan Huang, S. Panchapakesan	8. CONTRACT OR GRANT NUMBER(s): N00014-75-C-0455
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Purdue University, Department of Statistics, West Lafayette, IN 47907	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS: Office of Naval Research, Washington, DC	12. REPORT DATE: October 1982
	13. NUMBER OF PAGES: 33
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	15. SECURITY CLASS. (of this report): UNCLASSIFIED
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release, distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Selection and ranking, stochastic ordering, monotone likelihood ratio, generalizations, probability of correct selection, expected subset size, sufficient conditions for monotonicity, restricted families, partial ordering, $\mathcal{H}$-ordering inequalities, multivariate normal, multinomial, gamma, exponential family, reliability theory.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

In this paper, we restrict our attention mainly to some inequalities and monotonicity properties that have typically arisen in the development of the selection and ranking theory. Basic to the setup of these problems is the assumption regarding some order relations such as stochastic ordering and the monotone likelihood property. These and other related ideas, along with some basic inequalities that arise under these assumptions are discussed in Section 2. In reliability models, partial order relations such as convex ordering, star ordering and tail ordering play an important role. Section 3 deals with restricted families of

DD FORM 1 JAN 73 1473

(OVER)

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered) distributions defined by such partial order relations and some important inequalities obtained in the investigation of selection problems for such families. Interesting inequalities appear in the study of selection rules for normal, multinomial and gamma distributions. These are discussed in Section 4.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)