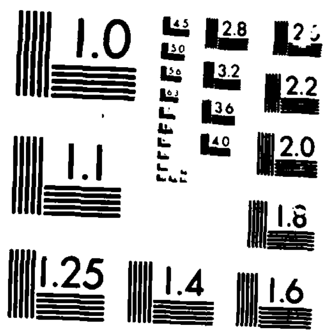
AD-A166 152 SELECTING THE BEST POPULATION: A DECISION THEORETIC APPROACH: THE CASE OF..(U) PITTSBURGH UNIV PA CENTER FOR MULTIVARIATE ANALYSIS N DAILAMI ET AL. NOV 85 1/1

UNCLASSIFIED TR-85-41 AFOSR-TR-86-0061 F49620-85-C-0008 F/G 12/1 NL



MICROCOPY RESOLUTION TEST CHART

SELECTING THE BEST POPULATION: A DECISION THEORETIC APPROACH: THE CASE OF PARETO DISTRIBUTION *

by

N. Dailami
Department of Probability and Statistics
The University of Sheffield
K. Subramanyam
Department of Mathematics and Statistics
University of Pittsburgh
M. Bhaskara Rao
Department of Probability and Statistics
The University of Sheffield

Center for Multivariate Analysis

University of Pittsburgh

2

SELECTING THE BEST POPULATION: A DECISION THEORETIC APPROACH: THE CASE OF PARETO DISTRIBUTION *

by

N. Dailami
Department of Probability and Statistics
The University of Sheffield
K. Subramanyam
Department of Mathematics and Statistics
University of Pittsburgh
M. Bhaskara Rao
Department of Probability and Statistics
The University of Sheffield

November 1985

Technical Report 85-41

Center for Multivariate Analysis
Fifth Floor, Thackeray Hall
University of Pittsburgh
Pittsburgh, PA 15260

*The work of K. Subramanyam and part of the work of M. Bhaskara Rao was done at the Center for Multivariate Analysis, University of Pittsburgh. It was supported partially by Contract F49620-85-C-0008 of the Air Force Office of Scientific Research. The United States Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright notation hereon.

SELECTING THE BEST POPULATION : A DECISION THEORETIC APPROACH : THE CASE OF PARETO DISTRIBUTION

by
N. Dailami
Department of Probability and Statistics
The University, Sheffield
K. Subramanyam
Department of Mathematics and Statistics
University of Pittsburgh
M. Bhaskara Rao
Department of Probability and Statistics
The University, Sheffield

Abstract

A decision theoretic approach is followed in selecting the best of k Pareto populations taking into account the cost of sampling and penalties for wrong decisions. Minimax sample sizes are determined under various types of penalty functions.

AMS 1980 subject classification: primary: 62F07, secondary: 62C20.

Key words and phrases. Pareto distribution, cost of sampling, penalty function, minimax criterion, selection problem.

SELECTING THE BEST POPULATION : A DECISION THEORETIC APPROACH : THE CASE OF PARETO DISTRIBUTION

N. DAILAMI, Department of Probability and Statistics, The University, Sheffield S3 7RH, U.K.,

K. SUBRAMANYAM, Department of Mathematics and Statistics, University of Pittsburgh, Pittsburgh, PA 15260, U.S.A.,

and

M. Bhaskara Rao, Department of Probability and Statistics, The University, Sheffield S3 7RH, U.K.

1. Introduction

The main ideas in selecting the best populations meeting some prescribed optimality criterion have been mooted originally by Bechchofer (1954) and Gupta (1956) and the subject has gone from strength to strength by several contributions by several statisticians over the last three decades. In this paper, the selection problem is tackled from a decision theoretic point of view. In selecting the best population, we take into account the cost of sampling and the penalties for taking a wrong decision. This kind of approach has been promulgated by Somerville (1954) and Ofosu (1972). We are basically interested in selecting the best Pareto population following the lead given by Somerville (1954) and Ofosu (1972). The Pareto probability model has been found useful in a wide variety of contexts. Pareto (1927) proposed this model to study the distribution of incomes in various societies for comparison. In medical circles, this has been used as a model for the remission rate of discharged psyciatric patients (See Fox and Kraemer (1971).)as a survival model for cardiac patients waiting for a heart transplant operation.(See Wingo (1979).). In the context of economics and marketing, this has been used as a model for the distribution of property values, business mortality, migration of workers, size of cities and firms, and consumer prices. See Steindl (1965) and Koutrouvelis (1981). Ofosu (1972) has worked out minimax sample sizes under a certain penalty function in the selection of the best Gamma population. We feel that there is a gap in one of the steps involved in the derivation of minimax sample sizes. In this paper, we consider four different types of penalty functions including

the one considered by Ofosu (1972). Under three of these penalty functions, we derive the minimax sample sizes. The maximum of the resultant loss functions is explicitly derived overcoming the difficulty faced by Ofosu (1972). We will elaborate on this at the appropriate juncture.

2. General formulation

In this section, we present general ideas in selecting the best poulation from a decision theoretic point of view. In the next section, we specialize in the Pareto distribution.

There are k populations $\pi_1, \pi_2, \cdots, \pi_k$ under study. Let X_i denote an observation selected at random from the i-th population π_i, $i = 1,2,\cdots,k$. The probability law governing the generation of X_i's is described by a probability density function $f(\cdot;\theta_i)$ which depends on an unknown parameter π_i, $i = 1,2,\cdots,k$. We assume that the functional form of f is completely specified and it is the same for all the populations. The only unknown quantity that enters the density function is θ_i and the set of all possible configurations $(\theta_1,\theta_2,\cdots,\theta_k)$ is denoted by Θ . We declare a population to be the best if its θ-value is the largest. If two or more -values are equal and largest, we adopt a well defined convention in declaring one of the corresponding populations to be the best. One such commonly adopted convention is the following. If θ_i and θ_j are equal and the largest, the population π_j is declared to be the best if $j > i$. Conventions are needed only in the calculation of probabilities of certain events in the selection problem. In reality, if θ_i and θ_j are equal and the largest, we could regard both the populations π_i and π_j to be the best.

One could also define 'the best population' to be the one whose θ-value is the least. The treatment of the selection problem within the purview of this definition is analogous to the one we are going to develop for the above definition.

A selection problem basically consists of two components.

(i) Draw a random sample of size n_i from population π_i, $i = 1,2,\cdots,k$.

(ii) Develop a statistical procedure R built on $n_1 + n_2 + \cdots + n_k$ observations which, once the data are given, clearly, declares the best population.

Introducing a good statistical procedure R to select the best population is not a difficult job for many probability models. The following is a natural procedure. Let $Y_1, Y_2, \cdots, Y_n$ be n independent identically distributed random variables with common probability density function $f(\cdot;\theta)$, θ unknown. Choose, if possible, a minimum variance unbiased estimator of θ based on $Y_1, Y_2, \cdots, Y_k$. If minimum variance unbiased estimator of θ does not exist, choose some decent estimator of θ based on $Y_1, Y_2, \cdots, Y_k$ which has good asymptotic properties. Let us denote the chosen estimator by $T(Y_1, Y_2, \cdots, Y_k)$. (In the Pareto case we are going to discuss, a natural estimator of θ presents itself.)

<u>Statistical Procedure R</u>

Suppose the data from the k populations are arranged as follows.

<u>Population</u>	$\underline{\pi_1}$	$\underline{\pi_2}$	....	$\underline{\pi_k}$
	X_{11}	X_{21}	....	X_{k1}
<u>Data</u>	X_{12}	X_{22}	....	X_{k2}
	......	......	......	......
	X_{1n_1}	X_{2n_2}	....	X_{kn_k}

Let $T_j = T(X_{j1}, X_{j2}, \cdots, X_{jn_j})$, $j = 1,2,\cdots,k$.

<u>Population π_i is the best if $T_i > T_j$ for every $j \neq i$.</u>

For simplicity, here, we assume that each T_i has a continuous probability density function. If two or more T_i values are equal and the largest, the population with largest suffix i is declared to be the best (our convention). Of course, this event occurs with probability zero.

The heart of the matter in selection problems is the choice of the sample sizes $n_1, n_2, \cdots, n_k$. We need to introduce an optimality criterion so that the sample sizes chosen are optimal according to the criterion proposed. To simplify the problem, we decide in advance that we intend to select samples of the same size form each population. Let $n_1 = n_2 = \cdots = n_k = n$.

Following Somerville (1954) and Ofosu (1972), we proceed as follows. We adopt the statistical procedure R described above to select the best population. No statistical procedure is infallible. It might declare a wrong population to be the best. Let $\theta_1, \theta_2, \cdots, \theta_k$ be a configuration of the populations $\pi_1, \pi_2, \cdots, \pi_k$ respectively. Let $\theta_{(1)} \leq \theta_{(2)} \leq \cdots \leq \theta_{(k)}$ be the arrangement of $\theta_1, \theta_2, \cdots, \theta_k$ in increasing order of magnitude. Let π_{i_j} be the population whose parameter value is $\theta_{(j)}$, $j = 1,2,\cdots,k$. If two or more θ-values coincide, by the convention mentioned earlier, the corresponding i_j's are taken to be in increasing order of magnitude. $i_1, i_2, \cdots, i_k$ is merely a permutation of $1,2,\cdots,k$ which depends on the configuration $(\theta_1,\theta_2,\cdots,\theta_k)$ in Θ. Let the statistic $T_{i_j} = T(X_{i_j1},X_{i_j2},\cdots,X_{i_jn})$ correspond to the population π_{i_j}, $j = 1,2,\cdots,k$. According to the configuration $(\theta_1,\theta_2,\cdots,\theta_k)$ in Θ, π_{i_k} is the best population. Under the given configuration $(\theta_1,\theta_2,\cdots,\theta_k)$ in Θ, the statistical procedure R can commit $k-1$ different types of wrong decisions. The list is given below.

<u>Wrong decision</u> : <u>Type j</u> : <u>π_{i_j} is declared to be the best.</u>

$(j = 1,2,\cdots,k-1)$

<u>Correct decision</u> : <u>π_{i_k} is declared to be the best.</u>

In order to calculate the probabilities of the above events, we assume that the probability density function of the statistic $T = T(Y_1,Y_2,\cdots,Y_n)$, where $Y_1, Y_2, \cdots, Y_k$ are independent identically distributed random variables with common probability density function $f(\cdot;\theta)$, is of the continuous type. Under this assumption, the random variables $T_1, T_2, \cdots, T_k$ are distinct with probability one under every configuration $(\theta_1,\theta_2,\cdots,\theta_k)$ in Θ. Let $g(\cdot;\theta;n)$ be the probability density function of $T = T(Y_1,Y_2,\cdots,Y_k)$ and $G(\cdot;\theta;n)$ its distribution function. Let

$$\begin{aligned} P_j(\theta_1,\theta_2,\cdots,\theta_k;n) &= \text{Prob(R takes a wrong decision of Type j)} \\ &= Pr_{\theta_1,\theta_2,\cdots,\theta_k}(T_{i_j} > T_{i_r} \text{ for every } r \neq j) \\ &= \int_{-\infty}^{\infty} \prod_{\substack{r=1 \\ r\neq j}}^{k} G(x;\theta_{(r)};n)\, g(x;\theta_{(j)};n)\, dx, \\ & \qquad j = 1,2,\cdots,k-1, \end{aligned}$$

and

$$\begin{aligned} P_k(\theta_1,\theta_2,\cdots,\theta_k;n) &= \text{Prob(R takes the correct decision)} \\ &= Pr_{\theta_1,\theta_2,\cdots,\theta_k}(T_{i_k} > T_{i_r} \text{ for } r = 1,2,\cdots,k-1) \\ &= \int_{-\infty}^{\infty} \prod_{r=1}^{k-1} G(x;\theta_{(r)};n)\, g(x;\theta_{(k)};n)\, dx . \end{aligned}$$

<u>Cost of sampling</u>

The cost of selecting a random sample of size n from each population has to be taken into account in the determination of an optimal sample size n. We assume that the cost function to be of the form

$$C(n) = c_o + kc_1 n^d, \; n = 1,2,\cdots ,$$

where c_o, c_1 and d are nonnegative constants. c_o and c_1 are measured in the same units, and c_o represents fixed administrative costs involved in setting up a sampling plan. If $d = 1$, the cost of taking additional samples rises linearly with n. If $d < 1$, the rise in the cost does not increase relatively with increasing sample sizes. If $d > 1$, it will become more and more expensive to take additional samples.

Penalty for wrong decisions

Let $W_j(\theta_1,\theta_2,\cdots,\theta_k)$ denote penalty for taking a wrong decision of Type j which is measured in the same units as those of c_o and c_1 for $j = 1,2,\ldots,k-1$ and for every configuration $(\theta_1,\theta_2,\ldots,\theta_k)$ in Θ .(In the next section, we discuss several choices of penalty functions.)

Loss function

For each configuration $(\theta_1,\theta_2,\ldots,\theta_k)$ in Θ, and for each sample size $n \geq 1$, we assume that the loss function L to have the following structure.

$$\begin{aligned} L(\theta_1,\theta_2,\ldots,\theta_k;n) &= \text{Cost of sampling + Average penalty for wrong decisions} \\ &= C(n) + \sum_{j=1}^{k-1} \text{(Penalty for wrong decision of Type j)} \times \\ &\qquad \text{Pr(R commits wrong decision of Type j)} \\ &= C(n) + \sum_{j=1}^{k-1} W_j(\theta_1,\theta_2,\ldots,\theta_k)\, P_j(\theta_1,\theta_2,\ldots,\theta_k;n) . \end{aligned}$$

Decision theoretic formulation

We now identify the state space to be Θ and the action space to be $N = \{1,2,3, \cdots \}$, the set of all possible sample sizes, and the loss function L is the one described above. It is defined on the cartesian product space $\Theta \times N$. The minimax sample size n minimizes

$$\max_{\tilde{\theta}\varepsilon\Theta} L(\tilde{\theta};m)$$

over all m in N. In the next section, we specialize the case of Pareto distribution.

3. On selecting the best Pareto population

The probability density function of a Pareto distribution is given by

$$f(x;\theta;m) = \theta m^{\theta}/x^{\theta+1} , \quad m \leq x < \infty ,$$

where $m > 0$ and $\theta > 0$ are the parameters of the model. We recall some distribution theory concerning this model. For details, see Johnson and Kotz (1970). Let $X_1,X_2,\cdots,X_n$ be independent identically distributed random variables with common probability density function f given above. Let

$$\hat{m} = \min_{1\leq i\leq n} X_i$$

and

$$\hat{\theta} = (\log((\prod_{i=1}^{n} X_i)^{1/n}/\hat{m}))^{-1}.$$

Then $\hat{m}$ and $\hat{\theta}$ are jointly sufficient for m and θ. Further,

$$2n\theta/\hat{\theta} \sim \chi^2_{2n},$$

where χ^2_{2n} chi-square distribution with $2n$ degrees of freedom.

The problem to which we address ourselves in this section is the following. The j-th population π_j has Pareto distribution with parameters

m_j and θ_j, both unknown, $j = 1,2,\cdots,k$. We want to select that population π_j for which θ_j is the largest. As expostulated earlier, we have to strike at a decent estimator T of θ based on n independent observations from a Pareto distribution with parameters m and θ in order to describe the statistical procedure R in the selection of the best population. From the deliberations carried out above, a clear choice emerges. Let

$$T(X_1,X_2,\cdots,X_n) = \hat{\theta} = \left[\log \frac{\left(\prod_{i=1}^{n} X_i\right)^{1/n}}{\min\limits_{1\le i\le n} X_i}\right]^{-1} \tag{3.1}$$

The parameter space Θ in this case identifies as the positive orthant of the k-dimensional euclidean space, i.e.,

$$\Theta = \{(\theta_1,\theta_2,\cdots,\theta_k) \; ; \; \theta_i > 0 \text{ for all } i\} .$$

<u>Statistical Procedure R</u>

Let $X_{j1},X_{j2},\cdots,X_{jn}$ be a random sample of size n from population π_j, $j = 1,2,\cdots,k$.

<u>Declare population π_i $(i = 1,2,\cdots,k)$ to be the best if</u>

<u>$T(X_{i1},X_{i2},\cdots,X_{in}) = T_i > T_j = T(X_{j1},X_{j2},\cdots,X_{jn})$</u>

<u>for all $j \ne i$</u>,

where the statistic T is given by the formula (3.1).

The error probabilities in this case work out as follows.

$$p_j(\theta_1,\theta_2,\cdots,\theta_k;n) = \int_0^\infty \prod_{\substack{r=1\\ r\ne j}}^{k} \left(1 - G_{2n}\left(\frac{\theta(r)}{\theta(j)} x\right)\right) g_{2n}(x)\, dx, \tag{3.2}$$

for $j = 1,2,\cdots,k-1$, where $g_{2n}(\cdot)$ is the probability density function of a chi-square distribution with $2n$ degrees of freedom and $G_{2n}(\cdot)$ its distribution function.

The probability that R takes the correct decision is given by

$$p_k(\theta_1,\theta_2,\cdots,\theta_k;n) = \int_0^\infty \prod_{r=1}^{k-1} \left(1 - G_{2n}\left(\frac{\theta_{(r)}}{\theta_{(k)}} x\right)\right) g_{2n}(x) \; dx . \qquad (3.3)$$

At this juncture, we state some inequalities concerning the above probability for future use.

<u>Lemma 3.1</u> For every $(\theta_1,\theta_2,\cdots,\theta_k)$ in Θ,

$$p_k(\theta_1,\theta_2,\cdots,\theta_k;n) \geq \int_0^\infty (1 - G_{2n}(x))^{k-1} g_{2n}(x) \; dx,$$

and equality holds in the above if $\theta_1 = \theta_2 = \cdots = \theta_k$.

<u>Proof</u>. Obvious.

<u>Lemma 3.2</u> Let $(\theta_1,\theta_2,\cdots,\theta_k) \in \Theta$. Write $\theta_{(k)} = r\theta_{(k-1)}$. (Obviously, $r \geq 1$.) Then

$$p_k(\theta_1,\theta_2,\cdots,\theta_k;n) \geq \int_0^\infty (1 - G_{2n}(x/r))^{k-1} g_{2n}(x) \; dx,$$

and equality in the above holds if $\theta_1 = \theta_2 = \cdots = \theta_{k-1}$ and $\theta_k = r\theta_{k-1}$.

<u>Proof</u>. Easy to check.

Now, the loss function L takes the following form. For $(\theta_1,\theta_2,\cdots,\theta_k)$ in Θ and $n \geq 1$,

$$L(\theta_1,\theta_2,\cdots,\theta_k;n) = C(n) + \sum_{j=1}^{k-1} W_j(\theta_1,\theta_2,\cdots,\theta_k) \int_0^\infty \prod_{\substack{r=1 \\ r\neq j}}^{k} (1 - G_{2n}(\frac{\theta_{(r)}}{\theta_{(j)}} x)) \; g_{2n}(x) \; ($$

In order to find the minimax sample size, we need to maximize the above loss function over Θ for every fixed $n \geq 1$. We consider four different types of penalty functions W_j's and examine the behaviour of the loss function under each one of these penalty functions.

Penalty function of Type 1 : Constant penalty.

Let $W_j(\theta_1,\theta_2,\cdots,\theta_k) = c$, a constant for every $(\theta_1,\theta_2,\cdots,\theta_k)$ in Θ and $j = 1,2,\cdots,k-1$. (For simplicity, take $c = 1$.) What this means is that whatever may be the type of wrong decision, the same amount of penalty is imposed. In this case, the loss function L simplifies to

$$L(\theta_1,\theta_2,\cdots,\theta_k;n) = C(n) + \sum_{j=1}^{k-1} p_j(\theta_1,\theta_2,\cdots,\theta_k;n)$$

$$= C(n) + 1 - p_k(\theta_1,\theta_2,\cdots,\theta_k;n)$$

for every $(\theta_1,\theta_2,\cdots,\theta_k)$ in Θ and $n \geq 1$. The maximization of L over Θ can easily be worked out for every fixed $n \geq 1$. For this, let us introduce the following subset of Θ.

$$\Theta_0 = \{(\theta_1,\theta_2,\cdots,\theta_k) \in \Theta \; ; \; \theta_1 = \theta_2 = \cdots = \theta_k\}$$

Proposition 3.3 For every $n \geq 1$,

$$\max_{\tilde{\theta}\in\Theta} L(\tilde{\theta};n) = \max_{\tilde{\theta}\in\Theta_0} L(\tilde{\theta};n) = (k-1)/k.$$

Proof. First, we note that if $(\theta_1,\theta_2,\cdots,\theta_k) \in \Theta_0$, then $T_1, T_2, \cdots, T_k$ are independently identically distributed. Consequently, $p_j(\theta_1,\theta_2,\cdots,\theta_k;n) = 1/k$ for every $j = 1,2,\cdots,k$. Therefore,

$$\sum_{j=1}^{k-1} p_j(\theta_1,\theta_2,\cdots,\theta_k;n) = (k-1)/k$$

for every $(\theta_1,\theta_2,\cdots,\theta_k)$ in Θ_0. Also,

$$\begin{aligned}\max_{\tilde{\theta}\varepsilon\Theta} \sum_{j=1}^{k-1} p_j(\tilde{\theta};n) &= \max_{\tilde{\theta}\varepsilon\Theta} (1 - p_k(\tilde{\theta};n)) \\ &= 1 - \min_{\tilde{\theta}\varepsilon\Theta} p_k(\tilde{\theta};n) \\ &= 1 - \int_0^\infty (1 - G_{2n}(x))^{k-1} g_{2n}(x)\,dx, \\ &\quad \text{by Lemma 3.1,} \\ &= 1 - p_k(\theta_1,\theta_1,\cdots,\theta_1;n) \\ &= 1 - (1/k) = (k-1)/k.\end{aligned}$$

This completes the proof.

As a consequence of the above proposition, we have the following results.

<u>Theorem 3.4</u> For every fixed $n \geq 1$,

$$\max_{\tilde{\theta}\varepsilon\Theta} L(\tilde{\theta};n) = C(n) + (k-1)/k.$$

<u>Corollary 3.5</u> Under the Penalty function of Type 1, the minimax sample size is $n = 1$.

<u>Penalty function of Type 2</u> : Penalty function which takes into account differences between the best and the second best populations.

It is natural to impose penalties for wrong decisions based on the magnitude of the parameters involved. A critical point in the selection problem is the times at which the second best population passes as the best population. If $(\theta_1, \theta_2, \cdots, \theta_k)$ in Θ is a configuration of the parameters of the k populations, the penalty for declaring the second best population as the best population should naturally depend on $\theta_{(k-1)}$ and $\theta_{(k)}$. Some natural penalty functions are

$$W_{k-1}(\theta_1, \theta_2, \cdots, \theta_k) = c_2 \frac{\theta_{(k)}}{\theta_{(k-1)}} \text{ or } = c_2 \log \frac{\theta_{(k)}}{\theta_{(k-1)}},$$

where $c_2 > 0$ is measured in the same units as those of c_o and c_1. There are other types of wrong decisions too. We could insist that the penalty for other types of wrong decisions should be atleast $W_{k-1}(\theta_1, \theta_2, \cdots, \theta_k)$. Let

$$W_j(\theta_1, \theta_2, \cdots, \theta_k) = d_j W_{k-1}(\theta_1, \theta_2, \cdots, \theta_k),$$

for $j = 1, 2, \cdots, k-2$, where $d_1, d_2, \cdots, d_{k-2}$ are constants exceeding unity. It is natural to take $d_1 \geq d_2 \geq \cdots \geq d_{k-2}$ embodying the principle that the more extreme the type of wrong decision the more severe the penalty is. In what follows, we take

$$W_{k-1}(\theta_1, \theta_2, \cdots, \theta_k) = c_2 \log \frac{\theta_{(k)}}{\theta_{(k-1)}}.$$

The loss function then becomes

$$L(\theta_1, \theta_2, \cdots, \theta_k; n) = C(n) + \sum_{j=1}^{k-1} c_2 d_j (\log \frac{\theta_{(k)}}{\theta_{(k-1)}}) p_j(\theta_1, \theta_2, \cdots, \theta_k; n)$$

for all $(\theta_1, \theta_2, \cdots, \theta_k)$ in Θ and $n \geq 1$, where $d_{k-1} = 1$.

Our next objective is to maximize the loss function L over Θ for every fixed $n \geq 1$. We were unable to maximize the loss function explicitly in its full generality stated above. The following results give explicitly the maximum of this loss function over Θ in the case $d_1 = d_2 = \cdots = d_{k-2}$.

Before stating the relevant results, we introduce the following subset of Θ.

$$\Theta_1 = \{(\theta_1,\theta_2,\cdots,\theta_k) \in \Theta \; ; \; \theta_1 = \theta_2 = \cdots = \theta_{k-1} \text{ and } \theta_k = r\theta_{k-1} \text{ for some } r \geq 1\}$$

<u>Lemma 3.6</u> For every fixed $n \geq 1$, let

$$p = \max \sum_{j=1}^{k-1} \left(\log \frac{\theta_{(k)}}{\theta_{(k-1)}}\right) p_j(\theta_1,\theta_2,\cdots,\theta_k;n) \, ,$$

where the maximum is taken over all $(\theta_1,\theta_2,\cdots,\theta_k)$ in Θ,

and

$$q = \max \sum_{j=1}^{k-1} \left(\log \frac{\theta^*_{(k)}}{\theta^*_{(k-1)}}\right) p_j(\theta^*_1,\theta^*_2,\cdots,\theta^*_k;n) \, ,$$

where the maximum is taken over all $(\theta^*_1,\theta^*_2,\cdots,\theta^*_k)$ in Θ_1.

Then

$$p = q = \max_{r \geq 1} (\log r)\,(k-1) \int_0^\infty (1 - G_{2n}(x))^{k-2}\,(1 - G_{2n}(xr))\, g_{2n}(x)\, dx$$

$$= \max_{r \geq 1} (\log r)\left(1 - \int_0^\infty (1 - G_{2n}(x/r))^{k-1}\, g_{2n}(x)\, dx\right).$$

<u>Proof</u>. Since $\Theta_1 \subset \Theta$, $q \leq p$. We now prove $p \leq q$. For this, it suffices

to show the following. Given $(\theta_1,\theta_2,\cdots,\theta_k)$ in Θ, there exists $(\theta_1^*,\theta_2^*,\cdots,\theta_k^*)$ in Θ_1 such that

$$\sum_{j=1}^{k-1} (\log \frac{\theta_{(k)}}{\theta_{(k-1)}})\ p_j(\theta_1,\theta_2,\cdots,\theta_k;n) \leq \sum_{j=1}^{k-1} (\log \frac{\theta^*_{(k)}}{\theta^*_{(k-1)}})\ p_j(\theta_1^*,\theta_2^*,\cdots,\theta_k^*;n).$$

Let $\theta_1^* = \theta_2^* = \cdots = \theta_{k-1}^* = \theta_{(k-1)}$ and $\theta_k^* = \theta_{(k)}$. Obviously, $\theta_k^* = r\theta_{k-1}^*$ for some $r \geq 1$ and $(\theta_1^*,\theta_2^*,\cdots,\theta_k^*) \in \Theta_1$. Therefore,

$$\sum_{j=1}^{k-1} (\log \frac{\theta_{(k)}}{\theta_{(k-1)}})\ p_j(\theta_1,\theta_2,\cdots,\theta_k;n)$$

$$= \sum_{j=1}^{k-1} (\log r)\ p_j(\theta_1,\theta_2,\cdots,\theta_k;n)$$

$$= (\log r)\ (1 - p_k(\theta_1,\theta_2,\cdots,\theta_k;n))$$

$$\leq (\log r)\ (1 - \int_0^\infty (1 - G_{2n}(x/r))^{k-1}\ g_{2n}(x)\ dx), \text{ by Lemma 3.2}$$

$$\leq (\log r)\ (1 - p_k(\theta_1^*,\theta_2^*,\cdots,\theta_k^*;n))$$

$$= (\log r) \sum_{j=1}^{k-1} p_j(\theta_1^*,\theta_2^*,\cdots,\theta_k^*;n)$$

$$= \sum_{j=1}^{k-1} (\log \frac{\theta^*_{(k)}}{\theta^*_{(k-1)}})\ p_j(\theta_1^*,\theta_2^*,\cdots,\theta_k^*;n)\ .$$

This proves that $p \leq q$ and hence $p = q$. Incidentally, if $\theta_1^* = \theta_2^* = \cdots = \theta_{k-1}^*$ and $\theta_k^* = r\theta_{k-1}^*$ for some $r \geq 1$, then for every $j = 1,2,\cdots,k-1$, we have

$$p_j(\theta_1^*,\theta_2^*,\cdots,\theta_k^*;n) = \int_0^\infty \prod_{\substack{p=1 \\ p\neq j}}^{k} (1 - G_{2n}(\frac{\theta^*_{(p)}}{\theta^*_{(j)}}\ x))\ \ g_{2n}(x)\ \ dx$$

$$= \int_0^\infty (1 - G_{2n}(x))^{k-2}\ (1 - G_{2n}(xr))\ g_{2n}(x)\ dx.$$

Consequently,

$$\max \sum_{j=1}^{k-1} \left(\log \frac{\theta_k^*}{\theta_{k-1}^*}\right) p_j(\theta_1^*,\theta_2^*,\cdots,\theta_k^*;n) ,$$

where the maximum is taken over all $(\theta_1^*,\theta_2^*,\cdots,\theta_k^*)$ in Θ_1,

$$= \max_{r\geq 1} (\log r)\ (k-1) \int_0^\infty (1 - G_{2n}(x))^{k-2} (1 - G_{2n}(xr))\ g_{2n}(x)\ dx$$

$$= \max_{r\geq 1} (\log r)\ (1 - P_k(\theta,\theta,\cdots,\theta,r\theta;n))$$

$$= \max_{r\geq 1} (\log r)\ (1 - \int_0^\infty (1 - G_{2n}(x/r))^{k-1}\ g_{2n}(x)\ dx).$$

This completes the proof.

As a consequence of this lemma, we obtain the following result.

<u>Theorem 3.7</u> For every $n \geq 1$,

$$\max_{\tilde{\theta}\varepsilon\Theta} L(\tilde{\theta};n) = \max_{\tilde{\theta}\varepsilon\Theta_1} L(\tilde{\theta};n)$$

$$= C(n) + c_2 \max_{r\geq 1} (\log r)(k-1) \int_0^\infty (1 - G_{2n}(x))^{k-2}(1 - G_{2n}(xr))\ g_{2n}(x)\ dx$$

$$= C(n) + c_2 \max_{r\geq 1} (\log r)\ (1 - \int_0^\infty (1 - G_{2n}(x/r))^{k-1}\ g_{2n}(x)\ dx).$$

Theorem 3.7 reduces the problem of maximizing L over Θ to the problem of maximizing the following function

$$f(r) = (\log r)\ (1 - \int_0^\infty (1 - G_{2n}(x/r))^{k-1}\ g_{2n}(x)\ dx)$$

over all $r \geq 1$. In view of Theorem 3.7, we modify the notation of the loss function from $L(\theta_1, \theta_2, \cdots, \theta_k; n)$ to

$$L(r;n) = C(n) + c_2 f(r),$$
$$r \geq 1 \text{ and } n \geq 1.$$

There are two stages involved in finding the minimax sample size.

Stage 1. For every fixed $n \geq 1$, maximize $L(r;n)$ over all $r \geq 1$. This is equivalent to maximizing $f(r)$ over all $r \geq 1$. Let $r_{n,k}$ be the value at which $f(r)$ is maximum.

Stage 2 Then find the minimum of $L(r_{n,k};n)$ over all $n \geq 1$. The value of n at which $L(r_{n,k};n)$ is minimum is the required minimax sample size.

Maximization of $f(r)$ over all $r \geq 1$ is not easy analytically. We observe that $f(\cdot)$ has the following properties. (i) $f(1) = 0$ (obvious). (ii) $\lim_{r \to \infty} f(r) = 0$. In order to understand the function $f(\cdot)$, we used a numerical quadrature formula to evaluate $f(r)$ for an extensive range of values of r and k. A sample of these findings is reproduced below.

Table 1 : Tabulation of values of $r_{n,k}$, the value at which $f(\cdot)$ is maximum

Sample size	No. of populations, k			
n	2	3	4	5
2	2.3	2.5	2.8	2.9
3	1.9	2.1	2.2	2.3
4	1.7	1.9	2.0	2.0
5	1.6	1.7	1.8	1.9
6	1.6	1.7	1.7	1.8
7	1.5	1.6	1.6	1.7
8	1.5	1.5	1.6	1.6
9	1.4	1.5	1.5	1.6
10	1.4	1.5	1.5	1.5

The following information emerges from these studies. (iii) For every fixed n and k, $f(\cdot)$ seems to be unimodal. (iv) For every fixed k, $r_{n,k}$ decreases with increasing n, i.e., $r_{m,k} \leq r_{n,k}$ if $m \geq n$. (This property useful when determining minimax sample sizes.)

Ofosu (1972), in his study of selecting the best of k Gamma populations, suggested two methods of determining minimax sample sizes.

Method 1 Choose a certain range of plausible values of sample size in which we hope the minimax sample size lies. Find $r_{n,k}$ for every n in the range selected. Evaluate $L(r_{n,k};n)$ for every n in the range. Then that value of n for which $L(r_{n,k};n)$ is minimum is the desired minimax sample size.

This method involved a lot of computer time. Determination of plausible minimax values of n involves an extensive tabulation of $L(r;n)$'s. Further, locating $r_{n,k}$'s is a time consuming process. The second method we are going to describe now is a slight modification of another method suggested by Ofosu (1972). This method uses the property (iv) mentioned above.

Method 2 k is fixed. Find $r_{2,k}$, the point at which $f(\cdot)$ is maximum when the sample size is 2. Evaluate $L(r;n)$ for $r = 1.0, 1.1, 1.2, \cdots, r_{2,k}$ and $n = 2,3,4, \cdots, 20$. These values are tabulated in the two-way grid, with rows corresponding to r and columns corresponding to n. For each column in the two-way grid, locate the maximum. Under each column representing a particular sample size n, we know that the maximum of $L(r;n)$ occurs at some value of r between 1.0 and $r_{2,k}$. See property (iv) above. These maxima exhibit a decreasing trend to start with and then rise steadily. We locate the column for which the maximum entry is minimum. The corresponding sample size is taken to be the minimax sample size. If these column maximums exhibit a downward trend over the entire sample range from 2 to 20 considered, then the minimax sample size is ≥ 20. In this case, we need to extend the range of values of n. We may need to use asymptotic expressions for $f(\cdot)$ for large values of n. This aspect is discussed in the last section.

We found that, in practice, Method 2 works faster than Method 1. By way of illustration, we have adopted Method 2 to find the minimax sample size for

$$k = 2,3,4,5;$$

$$c_0 = 0, \quad c_1 = 1 \text{ and } d = 1; \text{ and}$$

$$c_2 = 1000.$$

The loss function L is given by

$$L(r;n) = kn + 1000\,(\log r)\left(1 - \int_0^\infty (1 - G_{2n}(x/r))^{k-1}\, g_{2n}(x)\, dx\right),$$

$r \geq 1$ and $n \geq 1$. A sample two-way grid is given on the following page for $k = 4$. From the computations performed using Method 2, the following information is obtained.

Table 2 : Minimax sample sizes under Penalty function of Type 2

No. of populations k	Minimax sample size n	Minimax loss
2	15	92.54
3	16	142.53
4	18	183.83
5	14	218.97

Table 3: Pareto Distribution: Number of Populations = 4:

LOSS TABLE

$$\text{Loss} = L(r,n) = 4n + 1000\,(\log r)\,\{1 - \int_0^{\infty} (1 - G_{2n}(\tfrac{x}{r}))^3\, g_{2n}(x)\, dx\}$$

r \ n	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1.0	8.00	12.00	16.00	20.00	24.00	28.00	32.00	36.00	40.00	44.00	48.00	52.00	56.00	60.00	64.00	68.00	72.00	76.00	80.00
1.1	76.87	79.85	83.15	86.54	89.97	93.45	96.97	100.51	104.07	107.66	111.26	114.87	118.50	122.14	125.80	129.46	133.13	136.80	140.48
1.2	134.17	134.89	136.28	137.87	139.67	141.63	143.73	145.92	148.21	150.57	153.00	155.49	158.03	160.62	163.26	165.93	168.65	171.39	174.17
1.3	182.34	179.50	177.64	176.43	175.68	175.29	175.19	175.33	175.67	176.20	176.88	177.70	178.66	179.73	180.90	182.18	183.55	185. 0	186.54
1.4	222.69	215.09	209.18	204.38	200.41	197.08	194.27	191.91	189.93	188.28	186.94	185.87	185.04	184.44	184.05	183.85	183.83	183.97	184.28
1.5	258.43	243.21	232.57	223.66	216.04	209.44	203.69	198.66	194.25	190.40	187.05	184.14	181.65	179.52	177.74	176.27	175.10	174.21	173.57
1.6	284.58	265.10	249.26	235.92	224.46	214.51	205.82	198.20	191.52	185.66	180.54	176.09	172.24	168.93	166.13	163.78	161.86	160.33	159.16
1.7	307.97	281.77	260.48	242.60	227.31	214.13	202.69	192.74	184.09	176.58	170.08	164.48	159.70	155.65	152.25	149.46	147.22	145.48	144.19
1.8	327.29	294.08	267.24	244.88	225.95	209.78	195.92	184.02	173.81	165.08	157.64	151.35	146.07	141.70	138.14	135.30	133.12	131.52	130.44
1.9	343.14	302.78	270.44	243.77	221.47	202.69	186.81	173.39	162.07	152.55	144.61	138.04	132.67	128.35	124.96	122.39	120.55	119.35	118.71
2.0	356.02	308.48	270.77	240.09	214.79	193.79	176.34	161.84	149.63	139.95	131.90	125.40	120.26	116.28	113.32	111.23	109.91	109.26	109.19
2.1	366.33	311.66	268.86	234.50	206.60	183.63	165.23	150.07	137.78	127.89	120.04	113.90	109.22	105.79	103.41	101.93	101.23	101.18	102.31
2.2	374.45	312.81	265.20	227.54	197.48	173.34	154.01	138.57	126.33	116.73	109.32	103.74	99.68	96.90	95.19	94.37	94.31	94.88	95.13
2.3	380.68	312.28	260.19	219.64	187.81	162.74	143.04	127.64	115.72	106.63	99.85	94.95	91.61	89.54	88.51	88.36	88.91	90.05	91.12
2.4	385.28	310.37	254.17	211.14	177.95	152.30	132.56	117.47	106.09	97.67	91.62	87.48	84.88	83.52	83.17	83.63	84.75	86.41	88.50
2.5	388.48	307.35	247.41	202.30	168.13	142.24	122.72	108.15	97.46	89.61	84.55	81.10	79.33	78.66	78.93	79.96	81.58	83.68	86.16
2.6	390.49	303.44	240.15	193.33	158.52	132.68	113.59	99.71	89.81	82.99	78.55	75.95	74.60	74.77	75.60	77.13	79.18	81.68	84.46
2.7	391.45	298.81	232.55	184.38	149.23	123.66	105.21	92.12	83.06	77.11	73.48	71.62	71.12	71.67	73.01	74.96	77.38	80.16	83.22
2.8	391.51	293.63	224.75	175.57	140.36	115.26	97.58	85.38	77.21	72.09	69.22	68.05	68.15	69.21	70.98	73.30	76.03	79.06	82.33
2.9	390.82	288.02	216.88	166.99	131.95	107.49	90.66	79.37	72.11	67.80	65.87	65.13	65.76	67.27	69.42	72.04	75.01	78.25	81.68
3.0	389.46	282.10	209.02	158.70	124.03	100.34	84.43	74.07	67.68	64.16	62.70	62.73	63.85	65.74	68.20	71.08	74.26	77.68	81.22
Col max.	391.51	312.81	270.77	244.88	227.31	214.51	205.82	198.66	194.25	190.40	187.05	185.87	185.04	184.44	184.05	183.85	183.83	185.00	186.54

Penalty function of Type 3 : Penalties which take into account differences between the best and the rest of the populations.

Suppose the statistical procedure R described above commits the mistake of type j $(j = 1,2,\cdots,k-1)$. It is natural to insist that the penalty for this wrong decision should depend on the magnitudes of $\theta_{(j)}$ and $\theta_{(k)}$. Ofosu (1972), accordingly introduced the following penalty functions.

$$W_j(\theta_1,\theta_2,\cdots,\theta_k) = a_j \log \theta_{(k)}/\theta_{(j)}$$

for all $(\theta_1,\theta_2,\cdots,\theta_k)$ in Θ and $j = 1,2,\cdots,k-1$, where a_j's are positive constants measured in the same units as those of c_0 and c_1. The loss function L then works out to be

$$L(\theta_1,\theta_2,\cdots,\theta_k;n) = C(n) + \sum_{j=1}^{k-1} a_j(\log \theta_{(k)}/\theta_{(j)})\, p_j(\theta_1,\theta_2,\cdots,\theta_k;n)$$

for all $(\theta_1,\theta_2,\cdots,\theta_k)$ in Θ and $n \geq 1$. Ofosu (1972) asserts that

$$\max_{\tilde{\theta}\in\Theta} L(\tilde{\theta};n) = \max_{\tilde{\theta}\in\Theta_1} L(\tilde{\theta};n)$$

for every $n \geq 1$ and that this equality can be verified numerically only! He has no analytical proof of this equality. Then he went on to obtain the minimax sample size for selecting the best of k Gamma populations. We also do not have any analytical proof of this equality.We abandon the project of working with penalty functions of this type.

Penalty function of Type 4 : Penalties are imposed only when the ratio of the parameter values of the best population and the rest exceed a prescribed number.

In this part, we are going to introduce a new, natural and special penalty function. Let $0 < \delta < 1$ be given. Let $\theta_1, \theta_2, \cdots, \theta_k$ be a configuration of the parameters of the k populations $\pi_1, \pi_2, \cdots, \pi_k$ respectively. Suppose the statistical procedure R declares π_{i_j} to be the best for some $j = 1,2,\cdots,k-1$. The parameter values associated with the best population π_{i_k} and π_{i_j} are $\theta_{(k)}$ and $\theta_{(j)}$ respectively. If the ratio $\theta_{(j)}/\theta_{(k)}$ (≤ 1) of $\theta_{(j)}$ and $\theta_{(k)}$ is close to unity, we would not like to be penalized for taking the wrong decision of accepting π_{i_j} to be the best population. On the other hand, if the ratio $\theta_{(j)}/\theta_{(k)}$ is small, we certainly wish to be penalized for accepting π_{i_j} to be the best. A line has to be drawn somewhere between the statements that the ratio $\theta_{(j)}/\theta_{(k)}$ being close to unity and that it is being small. The number δ distinguishes these two statements and the choice of δ is subjective. The discussion carried out above can be embodied mathematically in the following way.

$$
\begin{aligned}
W_j(\theta_1,\theta_2,\cdots,\theta_k) &= a_j \quad \text{if} \quad \theta_{(j)}/\theta_{(k)} \leq \delta, \\
&= 0 \quad \text{if} \quad \theta_{(j)}/\theta_{(k)} > \delta,
\end{aligned}
$$

for all $(\theta_1,\theta_2,\cdots,\theta_k)$ in Θ and $j = 1,2,\cdots,k-1$, where a_j's are positive numbers measured in the same units as those of c_0 and c_1. To emphasize the gravity of the type of wrong decision taken, one may wish to have $a_1 \geq a_2 \geq \cdots \geq a_{k-1}$. This chain of inequalities indicates that the more extreme the type of wrong decision the more severe the penalty is. We deal with only the case $a_1 = a_2 = \cdots = a_{k-1} = a$, say. We were unable to obtain concrete results in the general case of these constants. Under this case, the loss function L works out as follows.

Let $(\theta_1, \theta_2, \cdots, \theta_k)$ in Θ be given. Let r be the largest index, if it exists, in $\{1,2,\cdots,k-1\}$ such that $\theta_{(r)}/\theta_{(k)} \leq \delta$. This implies that $\theta_{(j)}/\theta_{(k)} \leq \delta$ for $j = 1,2,\cdots,r$ and $\theta_{(j)}/\theta_{(k)} > \delta$ for $j = r+1, r+2,\cdots,k$. Then

$$L(\theta_1,\theta_2,\cdots,\theta_k;n) = C(n) + a \sum_{j=1}^{r} p_j(\theta_1,\theta_2,\cdots,\theta_k;n).$$

If there is no index r satisfying the above, then

$$L(\theta_1,\theta_2,\cdots,\theta_k;n) = C(n).$$

The index r, of course, depends on the configuration $(\theta_1,\theta_2,\cdots,\theta_k)$ in Θ.

We now proceed to obtain the minimax sample size n under the loss function given above. Towards this goal, we consider the following subset of Θ.

$$\Theta_2 = \{(\theta_1,\theta_2,\cdots,\theta_k) \varepsilon \Theta ; \theta_1 = \theta_2 = \cdots = \theta_r \text{ and } \theta_{r+1} = \theta_{r+2} = \cdots = \theta_k = (1/\delta)\theta_r \text{ for some } r \text{ in } \{1,2,\cdots,k-1\}\}.$$

The following results help us to maximize the loss function $L(\tilde{\theta};n)$ over all $\tilde{\theta} \varepsilon \Theta$ for every fixed $n \geq 1$.

<u>Lemma 3.8</u> Let $(\theta_1,\theta_2,\cdots,\theta_k) \varepsilon \Theta$. Suppose there exists a largest index r in $\{1,2,\cdots,k-1\}$ such that $\theta_{(r)}/\theta_{(k)} \leq \delta$. Let

$$\theta_1^* = \theta_2^* = \cdots = \theta_r^* \text{ and } \theta_{r+1}^* = \theta_{r+2}^* = \cdots = \theta_k^* = (1/\delta)\theta_r^*,$$

where $\theta_r^* = \theta_{(r)}$. Then $(\theta_1^*,\theta_2^*,\cdots,\theta_k^*) \varepsilon \Theta_2$ and

$$p_k(\theta_1,\theta_2,\cdots,\theta_k;n) \geq p_k(\theta_1^*,\theta_2^*,\cdots,\theta_k^*;n)$$

for every $n \geq 1$.

Proof. From the definition of Θ_2 , it is clear that $(\theta_1^*,\theta_2^*,\cdots,\theta_k^*) \varepsilon \Theta_2$.

Observe that

$$p_k(\theta_1^*,\theta_2^*,\cdots,\theta_k^*;n) = \int_0^\infty \prod_{j=1}^{k-1} (1 - G_{2n}(\frac{\theta_{(j)}^*}{\theta_{(k)}^*} x)) \; g_{2n}(x) \; dx$$

$$= \int_0^\infty (1 - G_{2n}(\delta x))^r (1 - G_{2n}(x))^{k-1-r} \; g_{2n}(x) \; dx.$$

Note that $\frac{\theta_{(i)}}{\theta_{(k)}} \leq \delta$ for every $i = 1,2,\cdots,r$ and $\frac{\theta_{(i)}}{\theta_{(k)}} \leq 1$ for every $i = r+1, r+2,\cdots, k$. Consequently,

$$p_k(\theta_1,\theta_2,\cdots,\theta_k;n) = \int_0^\infty \prod_{j=1}^{k-1} (1 - G_{2n}(\frac{\theta_{(j)}}{\theta_{(k)}} x)) \; g_{2n}(x) \; dx$$

$$= \int_0^\infty \prod_{i=1}^{r} (1 - G_{2n}(\frac{\theta_{(i)}}{\theta_{(k)}} x)) \prod_{i=r+1}^{k} (1 - G_{2n}(\frac{\theta_{(i)}}{\theta_{(k)}} x)) \; g_{2n}(x) \; dx$$

$$\geq \int_0^\infty (1 - G_{2n}(\delta x))^r (1 - G_{2n}(x))^{k-1-r} \; g_{2n}(x) \; dx$$

$$= p_k(\theta_1^*,\theta_2^*,\cdots,\theta_k^*;n).$$

This completes the proof.

Lemma 3.9 The following equality is true for every $n \geq 1$.

$$\max_{\tilde{\theta}\varepsilon\Theta} L(\tilde{\theta};n) = \max_{\tilde{\theta}\varepsilon\Theta_2} L(\tilde{\theta};n).$$

Proof. Let $p = \max_{\tilde{\theta}\in\Theta} L(\tilde{\theta};n)$ and $q = \max_{\tilde{\theta}\in\Theta_2} L(\tilde{\theta};n)$. Since $\Theta_2 \subset \Theta$, $q \leq p$. To prove the reverse inequality $p \leq q$, it suffices to prove the following. Given $(\theta_1,\theta_2,\cdots,\theta_k)$ in Θ, there exists $(\theta_1^*,\theta_2^*,\cdots,\theta_k^*)$ in Θ_2 such that

$$L(\theta_1,\theta_2,\cdots,\theta_k;n) \leq L(\theta_1^*,\theta_2^*,\cdots,\theta_k^*;n).$$

Let $(\theta_1,\theta_2,\cdots,\theta_k)$ in Θ be given. Case (i). $\frac{\theta_{(j)}}{\theta_{(k)}} > \delta$ for all $j = 1,2,\cdots,k-1$. Then $L(\theta_1,\theta_2,\cdots,\theta_k;n) = C(n) + 0 = C(n) \leq L(\theta_1^*,\theta_2^*,\cdots,\theta_k^*;n)$ for any $(\theta_1^*,\theta_2^*,\cdots,\theta_k^*)$ in Θ_2. Case (ii). There exists j in $\{1,2,\cdots,k-1\}$ such that $\frac{\theta_{(j)}}{\theta_{(k)}} \leq \delta$. Let r be the largest index in $\{1,2,\cdots,k-1\}$ such that $\frac{\theta_{(r)}}{\theta_{(k)}} \leq \delta$. Let $\theta_1^* = \theta_2^* = \cdots = \theta_r^* = \theta_{(r)}$ and $\theta_{r+1}^* = \theta_{r+2}^* = \cdots = \theta_k^* = (1/\delta)\theta_{(r)}$. Obviously, $(\theta_1^*,\theta_2^*,\cdots,\theta_k^*) \in \Theta_2$. Also,

$$\begin{aligned}
L(\theta_1,\theta_2,\cdots,\theta_k;n) &= C(n) + a\sum_{j=1}^{r} p_j(\theta_1,\theta_2,\cdots,\theta_k;n)\\
&\leq C(n) + a\sum_{j=1}^{k-1} p_j(\theta_1,\theta_2,\cdots,\theta_k;n)\\
&= C(n) + a(1 - p_k(\theta_1,\theta_2,\cdots,\theta_k;n))\\
&\leq C(n) + a(1 - p_k(\theta_1^*,\theta_2^*,\cdots,\theta_k^*;n)), \text{ by Lemma 3.8}\\
&= C(n) + a\sum_{j=1}^{k-1} p_j(\theta_1^*,\theta_2^*,\cdots,\theta_k^*;n)\\
&= L(\theta_1^*,\theta_2^*,\cdots,\theta_k^*;n).
\end{aligned}$$

This completes the proof.

The above lemma simplifies the problem of maximizing $L(\cdot;n)$ over Θ

to that of maximization of $L(\cdot;n)$ over Θ_2. We now solve the problem of maximization of $L(\cdot;n)$ over Θ_2. We partition Θ_2 as follows. Let

$$\Theta_{2j} = \{(\theta_1,\theta_2,\cdots,\theta_k) \in \Theta_2;\ \theta_1 = \theta_2 = \cdots = \theta_j \text{ and } \theta_{j+1} = \theta_{j+2} = \cdots = \theta_k = (1/\delta)\theta_1\},$$

for $j = 1,2,\cdots,k-1$.

Lemma 3.10 The following statement is true. If $(\theta_1,\theta_2,\cdots,\theta_k) \in \Theta_{2j}$, then

$$p_i(\theta_1,\theta_2,\cdots,\theta_k;n) = \int_0^\infty (1- G_{2n}(x/\delta))^{k-j} (1 - G_{2n}(x))^{j-1} g_{2n}(x)\, dx$$

for $i = 1,2,\cdots,j$ and $j = 1,2,\cdots,k-1$.

Proof. The above assertion follows directly from the definition of the probabilities p_i's and Θ_{2j}'s.

Theorem 3.11 For every fixed $n \geq 1$,

$$\max_{\tilde{\theta}\in\Theta} L(\tilde{\theta};n) = C(n) + a\left(1 - \int_0^\infty (1 - G_{2n}(\delta x))^{k-1} g_{2n}(x)\, dx\right)$$

$$= C(n) + a(k-1) \int_0^\infty (1 - G_{2n}(x/\delta))(1 - G_{2n}(x))^{k-2} g_{2n}(x)\, dx.$$

Proof. The proof is carried out in the following two major steps.

1°. In view of Lemma 3.9, we maximize $L(\tilde{\theta};n)$ over Θ_2. As a first step, we maximize $L(\cdot;n)$ over Θ_{2i} for every $i = 1,2,\cdots,k-1$. We observe that

$$\max_{\tilde{\theta}\in\Theta_{2i}} L(\tilde{\theta};n) = C(n) + \max_{\tilde{\theta}\in\Theta_{2i}} a \sum_{j=1}^{i} p_j(\tilde{\theta};n)$$

$$= C(n) + ia \int_0^\infty (1 - G_{2n}(x/\delta))^{k-i} (1 - G_{2n}(x))^{i-1} g_{2n}(x)\, dx,$$

by Lemma 3.10.

The maximum of $L(\tilde{\theta};n)$ over $\Theta_{2,k-1}$ has another equivalent expression. It is given by

$$\max_{\tilde{\theta}\varepsilon\Theta_{2,k-1}} L(\tilde{\theta};n) = C(n) + a \max_{\theta\varepsilon\Theta_{2,k-1}} \sum_{j=1}^{k-1} p_j(\tilde{\theta};n)$$

$$= C(n) + a \max_{\tilde{\theta}\varepsilon\Theta_{2,k-1}} (1 - p_k(\tilde{\theta};n))$$

$$= C(n) + a(1 - \int_0^\infty (1 - G_{2n}(\delta x))^{k-1} g_{2n}(x)\, dx)$$

2^o. The following chain of inequalities can easily be checked to be true.

$$\int_0^\infty (1 - G_{2n}(x/\delta))^{k-1} g_{2n}(x)\, dx \leq \int_0^\infty (1 - G_{2n}(x/\delta))^{k-2} (1 - G_{2n}(x))\, g_{2n}(x)\, dx$$

$$\leq \int_0^\infty (1 - G_{2n}(x/\delta))^{k-3} (1 - G_{2n}(x))^2 g_{2n}(x)\, dx \leq \cdots\cdots$$

$$\leq \int_0^\infty (1 - G_{2n}(x/\delta))\, (1 - G_{2n}(x))^{k-2} g_{2n}(x)\, dx.$$

Since Θ_{2i}, $i = 1,2,\cdots,k-1$ is a partition of Θ_2, we have

$$\max_{\tilde{\theta}\varepsilon\Theta_2} L(\tilde{\theta};n) = \max_{1\leq i\leq k-1} \{C(n) + ia \int_0^\infty (1 - G_{2n}(x/\delta))^{k-i} (1 - G_{2n}(x))^{i-1} g_{2n}(x)\, d$$

$$= C(n) + a(k-1) \int_0^\infty (1 - G_{2n}(x/\delta))(1 - G_{2n}(x))^{k-2} g_{2n}(x)\, dx$$

$$= C(n) + a(1 - \int_0^\infty (1 - G_{2n}(\delta x))^{k-1} g_{2n}(x)\, dx) \text{ , by Step } 1^o.$$

This completes the proof.

The above theorem explicitly determines the maximum of $L(\tilde{\theta};n)$ over all $\tilde{\theta}$ in Θ. This maximum is the same as the maximum of $L(\tilde{\theta};n)$ over all $\tilde{\theta}$ in $\Theta_{2,k-1}$, the so called "Least Favourable Choice Set". In order to find the minimax sample size, we have to solve the following problem.

Objective. For a given $0 < \delta < 1$, minimize

$$C(n) + a(1 - \int_0^\infty (1 - G_{2n}(\delta x))^{k-1} g_{2n}(x)\, dx)$$

over all $n \geq 1$.

We obtain the minimax sample sizes for $\delta = 0.5, 0.6, 0.7, 0.8, 0.9$; $k = 2,3,4,5$; $a = 200$; $c_o = 0$, $c_1 = 1$ and $d = 1$ by solving the above problem. These sample sizes are tabulated below.

Table 4 : Minimax sample sizes under Penalty function of Type 4

No. of populations k	δ = 0.5 Minimax sample size	0.5 Minimax loss	0.6 Minimax sample size	0.6 Minimax loss	0.7 Minimax sample size	0.7 Minimax loss	0.8 Minimax sample size	0.8 Minimax loss
2	10	32.95	12	45.80	12	62.84	8	82.07
3	11	53.14	12	73.34	11	98.20	6	120.55
4	11	71.68	11	97.23	8	124.99	3	143.26
5	10	88.89	10	118.10	6	144.74	2	157.26

k	δ = 0.9 Minimax sample size	Minimax loss
2	2	96.01
3	2	131.87
4	2	151.46
5	2	167.50

4. Asymptotic minimax value of n

In Section 3, we obtained minimax sample sizes under penalty function of type 2 and the relevant loss function involved is

$$L(r;n) = kn + 1000 (\log r) (1 - \int_0^\infty (1 - G_{2n}(x/r))^{k-1} g_{2n}(x)\,dx)$$
$$\text{for } r \geq 1 \text{ and } n \geq 1. \qquad (4.1)$$

Using Mehtod 2, in order to find the minimax sample size for a given k, we have tabulated the values of $L(r;n)$ for $r = 1.0, 1.1, \cdots, r_{2,k}$ and $n = 1,2,\cdots,20$. Fortunately, the minimax sample size n was one of the numbers $2,3,\cdots,20$. It could happen that the minimax sample size n is a number beyond 20. The constant 1000 appearing in the above loss function played a crucial role in keeping minimax sample sizes to a moderate level. If we replace 1000 by a larger number, minimax sample sizes do increase. To illustrate this point, we worked out the minimax sample sizes under the loss function

$$L(r;n) = kn + 1200 (\log r) (1 - \int_0^\infty (1 - G_{2n}(x/r))^{k-1} g_{2n}(x)\,dx),$$
$$\text{for } r \geq 1 \text{ and } n \geq 1, \qquad (4.2)$$

and these are tabulated below.

Table 5 : Minimax sample sizes under the loss functions (4.1) and (4.2)

No. of populations	Minimax sample size under (4.1)	Minimax sample size under (4.2)
2	15	18
3	16	16
4	18	18
5	14	19

If we increase the value of c_2, we may need to compute the value of $L(r;n)$ for $n > 20$. In this section, we approximate the integral involved in $L(r;n)$ by an integral involving the standard normal distribution for large values of n. The integral in focus is

$$\int_0^\infty (1 - G_{2n}(x/r))^{k-1} g_{2n}(x)\, dx\ . \tag{4.3}$$

The above integral can be obtained probabilistically as follows. Let Y_1, Y_2, $\cdots$, Y_k be k independent identically distributed random variables each having chi-square distribution with $2n$ degrees of freedom. Then

$$\begin{aligned} P_k(r;n) &= \Pr(Y_k < rY_j,\ j = 1,2,\cdots,k-1) \\ &= \int_0^\infty (1 - G_{2n}(x/r))^{k-1} g_{2n}(x)\, dx. \end{aligned}$$

Employing Wilson-Hilferty's transformation, we define

$$Z_i = \frac{(Y_i/2n)^{1/3} - (1 - 1/9n)}{(1/9n)^{\frac{1}{2}}}\ , \quad i = 1,2,\cdots,k.$$

See Kendall and Stuart (1977, p.398-399). If n is large, each Z_i is normally distributed with mean zero and variance unity. We note that

$$Y_k < rY_j \quad \text{if and only if} \quad Z_k < r^{1/3}Z_j + a_{n,r}\ , \quad \text{where}$$

$$a_{n,r} = \frac{(1 - 1/9n)(r^{1/3} - 1)}{(1/9n)^{\frac{1}{2}}}\ .$$

Consequently,

$$\begin{aligned} P_k(r;n) &= \Pr(Y_k < rY_j,\ j = 1,2,\cdots,k-1) = \\ &= \Pr(Z_k < r^{1/3}Z_j + a_{n,r},\ j = 1,2,\cdots,k-1) \end{aligned}$$

$$\cong \int_{-\infty}^{\infty} \left(1 - \Phi\left(\frac{x - a_{n,r}}{r^{1/3}}\right)\right)^{k-1} (2\pi)^{-\frac{1}{2}} \exp(-x^2/2)\, dx,$$

where $\Phi(\cdot)$ is the distribution function of the standard normal probability model. Thus for large n, the loss function works out to be

$$L(r;n) \cong kn + c_2(\log n)\left(1 - \int_{-\infty}^{\infty}\left(1-\Phi\left(\frac{x - a_{n,r}}{r^{1/3}}\right)\right)^{k-1} (2\pi)^{-\frac{1}{2}} \exp(-x^2/2)\, dx\right)$$

for $r \geq 1$ and $n \geq 1$.

The integral involved in the loss function under penalty function of type 4 is given by

$$\int_{0}^{\infty} (1 - G_{2n}(\delta x))^{k-1} g_{2n}(x)\, dx.$$

As before, if n is large, this integral is approximately equal to

$$\int_{-\infty}^{\infty} (1 - \Phi(\delta^{1/3}(x - b_{n,\delta}))^{k-1} (2\pi)^{-\frac{1}{2}} \exp(-x^2/2)\, dx,$$

where $b_{n,\delta} = \dfrac{(1 - 1/9n)((1/\delta)^{1/3} - 1)}{(1/9n)^{\frac{1}{2}}}$. The loss function, if n is large, under penalty function of type 4 is given by

$$L(r;n) = C(n) + a\left(1 - \int_{-\infty}^{\infty} (1 - \Phi(\delta^{1/3}(x - b_{n,\delta})))^{k-1} (2\pi)^{-\frac{1}{2}} \exp(-x^2/2)\, dx\right),$$

$r \geq 1$ and $n \geq 1$.

***

REFERENCES

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

Fox, P.D. and Kraemer, H.C. (1971). A probability model for the remission rate of discharged psychiatric patients. Management Sci., 17, B694-B699.

Gupta, S.S. (1956). On a decision rule for a problem in ranking means, Ph.D. Thesis (Mimeo. Ser. No. 150), Inst. of Statist., University of North Carolina, Chapel Hill.

Kendall, M.G. and Stuart, A, (1977). The Advanced Theory of Statistics, Vol. I, Charles Griffin and Company, Ltd., London, fourth edition.

Koutrouvelis, I.A. (1981). Large sample quantile estimation in Pareto laws, Comm. Statis. Theory Methods, A10(2), 189-201.

Ofosu, J.B. (1972). A minimax procedure for selecting the population with the largest (smallest) scale parameter, Calcutta Statist. Assoc. Bull., 21, 143-154. Errata 22 (1973), 129.

Pareto, V. (1927). Manuel d'economie politique, Giard, Paris.

Somerville, P.N. (1954). Some problems of optimum sampling, Biometrika, 41, 420-429.

Steindl, J. (1965). Random processes and the growth of firms, Hafner, New York.

Wingo, D.R. (1979). Estimation in a Pareto distribution: theory and computation. IEEE Trans. Rel. R28, 35-37.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER AFOSR-TR- 86-0061 85-41	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Selecting the best population: A decision theoretic approach: The case of Pareto distribution		5. TYPE OF REPORT & PERIOD COVERED Technical - November
		6. PERFORMING ORG. REPORT NUMBER 85-41
7. AUTHOR(s) N. Dailami, K. Subramanyam and M. Bhaskara Rao		8. CONTRACT OR GRANT NUMBER(s) F49620-85-C-0008
9. PERFORMING ORGANIZATION NAME AND ADDRESS Center for Multivariate Analysis 515 Thackeray Hall University of Pittsburgh, Pittsburgh, PA		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 6110 2F 2304 A5
11. CONTROLLING OFFICE NAME AND ADDRESS Air Force Office of Scientific Research Department of the Air Force Bolling Air Force DC 20332		12. REPORT DATE November 1985
		13. NUMBER OF PAGES 32
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)

Cost of sampling, minimax criterion, Pareto distribution, penalty function, selection problem.

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

A decision theoretic approach is followed in selecting the best of k Pareto populations taking into account the cost of sampling and penalties for wrong decisions. Minimax sample sizes are determined under various types of penalty functions.

DD FORM 1 JAN 73 1473

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

END
FILMED

5-86

DTIC