# SELECTION OF THE BEST POPULATION: AN INFORMATION THEORETIC APPROACH

**M. L. Menéndez[1], L. Pardo[2], Ch. Tsairidis[3] and K. Zografos[3]**

[1]Department of Applied Mathematics, Technical Univ. of Madrid, Madrid, SPAIN

[2]Department of Statistics and O.R., Complutense Univ. of Madrid, Madrid, SPAIN

[3]Department of Mathematics, University of Ioannina, 45110 Ioannina, GREECE

**Abstract**

This paper is devoted to the statistical problem of ranking and selection populations by using the subset selection formulation. The interest is focused (i) on the selection of the best population among $k$ independent populations and (ii) on the selection of the best population, which is closest to an additional standard or control population. With respect to the first problem the populations are ranked in terms of entropies of their distributions and the population whose distribution has maximum entropy is selected. For the second problem the populations are ranked in terms of divergences of their distributions and the distribution of the standard or control population. The population with the minimum divergence is selected. In each case the populations are assumed to have general parametric densities satisfying the classical regularity conditions of asymptotic statistic. Large sample properties of the estimators of entropies and divergences of the populations will be studied and used in order to determine the probabilities of correct selection of the proposed asymptotic selection rules. Illustrative examples, including a numerical example using real medical data, will be given for multivariate homoscedastic normal populations and populations described by the regular exponential family of distributions.

**Keywords and phrases:** Divergence; Entropy; Exponential family; Mahalanobis distance; Multivariate normal distribution; Ordering populations; Selection of populations; Subset selection approach.

# 1 Introduction

This paper deals with statistical problems of ranking and selection. It is of great theoretical and practical importance to build statistical methodology that helps to analyze data from $k$ independent experiments that have been subject to different (medical, agricultural, manufactoring etc.) treatments. The classical test procedure known as the analysis of variance was introduced by Fisher as a method dealing with the question of whether the $k$ treatment effects are different or not. In many experiments it seems obvious that the different treatments must have produced some difference, however small in effect. Then the

main problem is to find which of the $k$ treatments effects is the best or the largest in some sense. Experiments will be called populations, in the sequel, while treatment effects will be the associated parameters. For the solution of this problem two classical approaches to ranking and selection have been developed during the last fifty years by several authors: namely, the indifference zone approach of Bechhofer (1954) and the subset selection formulation of Gupta (1965). These general approaches are well known and discussions on them can be found in the books by Gibbons et al. (1977) and Gupta and Panchapakesan (1979). We also refer, among other, the following papers devoted to the subject selection approach which will be of our interest in the sequel: Gupta (1963, 1966), Gupta and Studden (1970), Gupta and Panchapakesan (1972), Panchapakesan and Santner (1977), Chattopadhyay (1981). Chen and Sobel (1987), and Jeyaratnam and Panchapakesan (1997) considered an integrated formulation of the problem combining features of the two classical approaches, while some other papers with slightly modified goal are referred by them. Another methodological approach is given by Liese and Miescke (1999) who study the problem of selecting the best of $k$ populations with methods from asymptotic decision theory. Finally we refer to the paper by Thabane and Safiul Haq (1999) where the same problem is studied by using the Bayesian approach.

Typical situations encountered in these and other treatments are as follows: We have $k \geq 2$ independent parametric populations and on the basis of random samples from these populations we wish to select the best population or the $t$, $2 \leq t \leq k$, best populations or to select the best population closest to an additional standard or control population. The term best population means the largest (or, in some cases, smallest) population that results from a ranking procedure of the said populations. For most univariate problems the selection and ranking has been defined in terms of location or scale parameters. Relatively limited is the literature on selection procedures for multivariate populations. In this case the interest was mainly focused on multivariate normal populations and one is usually interested in the ranking and selection problems in terms of suitably defined functions of the several parameters. These functions are usually some scalar quantities and the Mahalanobis distance from the origin is the most commonly used function for selecting and ranking the best from $k$ multivariate normal populations.

In this paper we are concerned with subset selection procedures for multivariate populations. We will discuss selection of the best of $k$ populations and selection of the best population, which is closest to an additional standard or control population. The populations considered will be members of a general multiparameter family of distributions. The interest will also be focused on the multiparameter exponential family of distributions as well as the multivariate normal distribution. Ranking and selection procedures from several populations will be based on generalized distance measures defined in the field of statistical information. Entropy and divergence measures will serve as functions of the parameters of the populations considered. In this context, in the next section, ranking and selection rules will be stated in terms of entropy or divergence functions of the populations. Section 3 will be devoted to the large sample behaviour of the estimators of entropy and divergence functions, which will be obtained if the unknown parameters of the respective populations are replaced by their maximum likelihood estimators. The

asymptotic results of Section 3 will help in order to determine, in the next section, the probabilities of correct selection of the selection rules. In the final Section 5 the ranking and selection procedures will be illustrated for multivariate homoscedastic normal populations, for populations described by the regular exponential family of distributions as well as the gamma distribution.

## 2 Statement of the problem and notations

Suppose that $\{f_\theta : \theta \in \Theta\}$ is a family of probability densities with respect to a measure $\mu$, defined on the space $\mathfrak{X}$, and the parameter space $\Theta \subset R^M$ is an open set. Suppose also that the classical regularity assumptions $(R1) - (R3)$ appearing in pages 144-145 of Serfling (1980) are satisfied:

$(R1)$ For each $\theta \in \Theta$, $f_\theta$ is positive and the derivatives

$$\frac{\partial \log f_\theta}{\partial \theta_s}, \quad \frac{\partial^2 \log f_\theta}{\partial \theta_s \partial \theta_r}, \quad \frac{\partial^3 \log f_\theta}{\partial \theta_s \partial \theta_r \partial \theta_t},$$

exist everywhere on $\mathfrak{X}$, for all $1 \leq s, r, t \leq M$.

$(R2)$ For each $\theta_0 \in \Theta$, there exist measurable functions $\alpha, \beta, \gamma : \mathfrak{X} \to [0, \infty)$, possibly depending on $\theta_0$, such that, for all $\theta$ in a neighborhood $N(\theta_0)$ and all above considered $s, r, t$, the relations

$$\left|\frac{\partial f_\theta}{\partial \theta_s}\right| \leq \alpha, \quad \left|\frac{\partial^2 f_\theta}{\partial \theta_s \partial \theta_r}\right| \leq \beta, \quad \left|\frac{\partial^3 \log f_\theta}{\partial \theta_s \partial \theta_r \partial \theta_t}\right| \leq \gamma,$$

hold on $\mathfrak{X}$, and

$$\int_{\mathfrak{X}} \alpha d\mu < \infty, \quad \int_{\mathfrak{X}} \beta d\mu < \infty, \quad \int_{\mathfrak{X}} \gamma f_\theta d\mu < \infty.$$

$(R3)$ For each $\theta \in \Theta$, the Fisher information matrix

$$I(\theta) = \left[\int_{\mathfrak{X}} \frac{\partial \log f_\theta}{\partial \theta_s} \frac{\partial \log f_\theta}{\partial \theta_r} f_\theta d\mu\right]_{M \times M},$$

exists and is positive definite, with all elements continuous in the variable $\theta$.

Further assume that $\theta_i = (\theta_{i1}, ..., \theta_{iM})^t$ denotes the parameter of the $i$th population $\Pi_i$, with density $f_{\theta_i}$, $i = 1, ..., k$, and that the following identifiability condition is satisfied:

$(R4)$ If $f_{\theta_i} = f_{\theta_j}$, *a.e.* $\mu$, then $\theta_i = \theta_j$, $i, j = 1, ..., k$.

Here, and in the sequel, "$t$" denotes the vector or matrix transpose.

There is an extensive literature and many tests for testing the homogeneity hypothesis $H_0 : \theta_1 = \theta_2 = ... = \theta_k$. However, in practice, a test of homogeneity is of limited interest

and usually alternative methods of statistical analysis are needed. For example, the statistician may be interested to determine (i) which of the $k$ populations differ from which others or (ii) to see which populations can be considered the best in some well defined sense of the term "best". Multiple comparisons methods are useful procedures for solving the first problem, while ranking and selection procedures are more appropriate for the solution of the second problem. We will focus our interest on ranking and selection procedures of the best of $k$ populations and selection of the best population, which is closest to an additional standard or control population $\Pi_0$ with density $f_{\theta_0}$, for $\theta_0 \in \Theta$. The ranking and selection procedures will be based on suitably defined functions of the parameters $\theta_i = (\theta_{i1}, ..., \theta_{iM})^t$, $i = 0, 1, ..., k$.

Denote by $H(\theta)$ the well-known Shannon entropy of the parametric density $f_\theta$, $\theta \in \Theta$, defined by

$$H(\theta) = -\int_{\mathfrak{X}} f_\theta(x) \log f_\theta(x) d\mu. \tag{1}$$

The entropy of the family of densities $f_\theta$ is a measure of variation of the uncertainty for the distributions belonging to the family $\{f_\theta : \theta \in \Theta\}$. A more general measure of entropy with the same meaning and similar properties has been defined by Burbea and Rao (1982). This measure is known as $\varphi-$entropy and it is defined by the relation

$$H^\varphi(\theta) = \int_{\mathfrak{X}} \varphi\left(f_\theta(x)\right) d\mu, \tag{2}$$

where $\varphi$ is a real concave function defined on $[0, \infty)$. In what follows we will suppose that $H^\varphi(\theta) \geq 0$. Shannon's entropy $H(\theta)$, defined by (1), can be immediately obtained from $H^\varphi(\theta)$ for $\varphi(u) = -u \log u$, $u > 0$. Due to the meaning of $H(\theta)$ or $H^\varphi(\theta)$, the ranking of populations $\Pi_i$, in terms of their Shannon or $\varphi-$entropies, can be intuitively interpreted as a ranking of the variation of the diversity of the respective populations. Let $H_i^\varphi = H^\varphi(\theta_i)$ be the entropy of the $i$th population $\Pi_i$, with $i = 1, ..., k$, and let $H^{\varphi\prime}$s be ranked as

$$H_{[1]}^\varphi \leq H_{[2]}^\varphi \leq ... \leq H_{[k]}^\varphi.$$

Suppose that there is not any apriori information regarding the correct pairing of the ordered $H_{[i]}^\varphi$ with the $k$ given populations. In this context, any population associated with $H_{[k]}^\varphi$ will be called the best population. Dudewicz and van der Meulen (1983) have developed methods for the selection of the best entropy population for continuous distributions on (0,1), while Gupta and Panchapakesan (1987) have used Shannon entropy for the selection from several multinomial populations.

Consider now the selection of the best population, among $\Pi_1, ..., \Pi_k$, which is closest to the population $\Pi_0$ with density $f_{\theta_0}$, for $\theta_0 \in \Theta$. The parameter $\theta_0$ may be known or unknown. If $\theta_0$ is known, $\Pi_0$ is referred to as a standard population, while if $\theta_0$ is unknown, $\Pi_0$ is called a control population. The ranking and selection formulation in this case is much more intuitively clear. Let us denote by $D_i$ a distance measure between the population $\Pi_i$ with density $f_{\theta_i}$ and the control or standard population $\Pi_0$ with density

$f_{\theta_0}$, for $\theta_0$, $\theta_i \in \Theta$ and $i = 1, ..., k$. A general class of distance measures, in a broad sense, between two distributions is the well-known $\phi$-divergence introduced independently by Csiszar (1963) and Ali and Silvey (1966). The $\phi$-divergence between population $\Pi_i$ and the population $\Pi_0$ is defined by

$$D_i^{\phi} = D^{\phi}(f_{\theta_i}, f_{\theta_0}) = \int_{\mathfrak{X}} f_{\theta_0}(x)\phi\left(\frac{f_{\theta_i}(x)}{f_{\theta_0}(x)}\right) d\mu, \tag{3}$$

and it is intuitively intended as a measure of the distance, from a statistical point of view, between the respective populations. The subscript $st$ or $co$ will be used in the notation $D_i^{\phi}$, i.e. $D_{i,st}^{\phi}$ or $D_{i,co}^{\phi}$, to declare that the population $\Pi_0$ is standard or control, respectively. Let $D^{\phi\prime}$s be ranked as

$$D_{[1],\zeta}^{\phi} \leq D_{[2],\zeta}^{\phi} \leq ... \leq D_{[k],\zeta}^{\phi},$$

with $\zeta$ being $st$ if $\Pi_0$ is standard and $\zeta$ being $co$ if $\Pi_0$ is a control population. Assume that there is not any apriori information regarding the correct pairing of the ordered $D_{[i],\zeta}^{\phi}$ with the $k$ given populations. In this context, any population associated with $D_{[1],co}^{\phi}$or $D_{[1],st}^{\phi}$ will be called the best population, closest to the control or standard population $\Pi_0$.

The $\varphi-$entropies $H_i^{\varphi}$ and the $\phi$-divergences $D_i^{\phi}$ depend on the populations parameters $\theta_0$ and $\theta_i$, $i = 1, ..., k$. Since the parameters $\theta_i$, $i = 1, ..., k$, are unknown it seems reasonable to compute an estimator $\widehat{\theta}_i$ of each $\theta_i$ from a corresponding random sample. In a similar context we consider an estimator $\widehat{\theta}_0$ of the unknown parameter $\theta_0$ for the case of a control population $\Pi_0$. Based on them the unknown $\varphi-$entropies and $\phi$-divergences $H_i^{\varphi}$ and $D_{i,\zeta}^{\phi}$, $i = 1, ..., k$, respectively, can be estimated by the quantities $\widehat{H}_i^{\varphi}$ and $\widehat{D}_{i,\zeta}^{\phi}$, $i = 1, ..., k$, if the unknown parameters $\theta_i$ and maybe $\theta_0$ are replaced by their estimators $\widehat{\theta}_i$, $i = 1, ..., k$, and $\widehat{\theta}_0$, in the expressions of the respective measures defined by relations (2) and (3). In this direction, consider independent random samples $X_{i1}, X_{i2}, ..., X_{in_i}$ from the population described by $f_{\theta_i}$, , $i = 1, ..., k$, and also an independent random sample $X_{01}, X_{02}, ..., X_{0n_0}$ from the control population $f_{\theta_0}$. Denote by $\widehat{\theta}_0$ and $\widehat{\theta}_i$ the maximum likelihood estimators of the parameters $\theta_0$ and $\theta_i$, $i = 1, ..., k$, respectively. Based on them, the estimators $\widehat{H}_i^{\varphi}$, $\widehat{D}_{i,st}^{\phi}$ and $\widehat{D}_{i,co}^{\phi}$ of $H_i^{\varphi}$, $D_{i,st}^{\phi}$ and $D_{i,co}^{\phi}$, $i = 1, ..., k$, are defined by

$$\widehat{H}_i^{\varphi} = H^{\varphi}(\widehat{\theta}_i) = \int_{\mathfrak{X}} \varphi\left(f_{\widehat{\theta}_i}(x)\right) d\mu, \tag{4}$$

$$\widehat{D}_{i,st}^{\phi} = D^{\phi}(f_{\widehat{\theta}_i}, f_{\theta_0}) = \int_{\mathfrak{X}} f_{\theta_0}(x)\phi\left(\frac{f_{\widehat{\theta}_i}(x)}{f_{\theta_0}(x)}\right) d\mu, \tag{5}$$

and

$$\widehat{D}_{i,co}^{\phi} = D^{\phi}(f_{\widehat{\theta}_i}, f_{\widehat{\theta}_0}) = \int_{\mathfrak{X}} f_{\widehat{\theta}_0}(x)\phi\left(\frac{f_{\widehat{\theta}_i}(x)}{f_{\widehat{\theta}_0}(x)}\right) d\mu. \tag{6}$$

In relation with the function $\phi$ considered in the general $\phi-$divergence statistics, we consider the following assumptions:

($\phi 1$) The function $\phi : [0, \infty) \to (-\infty, \infty]$ is convex and continuous. Its restriction on $(0, \infty)$ is finite, twice continuously differentable, with $\phi(1) = \phi^{'}(1) = 0$ and $\phi^{''}(1) = 1$.

($\phi 2$) For every $\theta^* \in \Theta$ there exists an open neighborhood $N(\theta^*)$, such that, for all $\theta_i \in N(\theta^*)$, $i = 1, ..., k$, it holds

$$\frac{\partial}{\partial \theta_{ir}} \int_{\mathfrak{X}} f_{\theta^*}(x) \phi\left(\frac{f_{\theta_i}(x)}{f_{\theta^*}(x)}\right) d\mu = \int_{\mathfrak{X}} \frac{\partial}{\partial \theta_{ir}} \left[ f_{\theta^*}(x) \phi\left(\frac{f_{\theta_i}(x)}{f_{\theta^*}(x)}\right) d\mu \right], \quad i = 1, ..., k, \; r = 1, ..., M.$$

Condition ($\phi 1$) deals with properties of $\phi-$divergences (cf. Liese and Vajda (1987)). Condition ($\phi 2$) is needed to apply delta method for obtaining asymptotic distributions of $\phi-$divergence statistics. Sufficient conditions for the validity of ($\phi 2$) are presented in Morales et al. (1997).

In this context, it seems reasonable to select the population which gives rise to $\widehat{H}^{\varphi}_{[k]} = \max\{\widehat{H}^{\varphi}_1, ..., \widehat{H}^{\varphi}_k\}$ as the best one or to select the population which gives rise to $\widehat{D}^{\phi}_{[1],\zeta} = \min\{\widehat{D}^{\phi}_{1,\zeta}, ..., \widehat{D}^{\phi}_{k,\zeta}\}$, with $\zeta$ being *st* or *co* for the problem of selection of the best population, which is closest to a standard or control population, respectively. However, since the sample from the population with $H^{\varphi}_{[k]}$ and $D^{\phi}_{[1],\zeta}$ do not always produce respectively $\widehat{H}^{\varphi}_{[k]}$ and $\widehat{D}^{\phi}_{[1],\zeta}$, due to sampling variation, we need a probabilistic formulation giving a specified confidence that the proposed procedures lead to a correct selection.

Following now the subset selection formulation of the problem (cf. for example, Gupta (1966), Gupta and Panchapakesan (1972), Chattopadhyay (1981)), a correct selection ($CS_1$) means selecting a nonempty subset of the $k$ populations so that the best population associated with $H^{\varphi}_{[k]}$ is included in this subset. In a similar manner, a correct selection ($CS_{2,\zeta}$) also means selecting a nonempty subset of the $k$ populations so that the best population, which is closest to a standard or control population, associated with $D^{\phi}_{[1],\zeta}$ is included in this subset, with $\zeta$ being *st* or *co*, respectively. Our aim is to develop a selection rule so that the subsets of the $k$ populations contain the best populations with probabilities at least $p^*_1$ $(1/k < p^*_1 < 1)$ and $p^*_{2,\zeta}$ $(1/k < p^*_{2,\zeta} < 1)$, respectively. If we denote by $R_1$ and $R_{2,\zeta}$ the respective selection rules leading to the selection of the best of $k$ populations and selection of the best population, which is closest to a standard or control population, respectively, then the problem mathematically is stated as follows:

$$\inf_{\Omega_1} \Pr(CS_1 | R_1) \geq p^*_1, \tag{7}$$

where $\Omega_1 = \{(H^{\varphi}_1, ..., H^{\varphi}_k) : H^{\varphi}_i \geq 0, i = 1, ..., k\}$, and

$$\inf_{\Omega_{2,\zeta}} \Pr(CS_{2,\zeta} | R_{2,\zeta}) \geq p^*_{2,\zeta}, \tag{8}$$

where $\Omega_{2,\zeta} = \{(D^{\phi}_{1,\zeta}, ..., D^{\phi}_{k,\zeta}) : D^{\phi}_{i,\zeta} \geq 0, i = 1, ..., k\}$ and $\zeta$ being *st* or *co* depending on whether $\Pi_0$ is a standard or a control population.

Then the natural selection rules are respectively:

$$\begin{array}{ll} & \text{The procedure for selecting the population } \Pi_{(k)} \\ & \text{associated with } H^{\varphi}_{[k]} \text{ is} \\ R_1 \text{ Rule:} & \text{Select } \Pi_i \text{ if and only if } b\widehat{H}^{\varphi}_i \geq \widehat{H}_{\max},\ b > 1, \end{array} \tag{9}$$

where $\widehat{H}_{\max} = \max\{\widehat{H}^{\varphi}_1, ..., \widehat{H}^{\varphi}_k\}$ and $b = b(k, p^*_1)$ a suitably defined constant so that (7) is satisfied. In a similar manner,

$$\begin{array}{ll} & \text{The procedure for selecting the population } \Pi_{(1)} \\ & \text{associated with } D^{\phi}_{[1],\zeta} \text{ is} \\ R_{2,\zeta} \text{ Rule:} & \text{Select } \Pi_i \text{ if and only if } c_\zeta \widehat{D}^{\phi}_{i,\zeta} \leq \widehat{D}_{\min,\zeta},\ c_\zeta < 1, \end{array} \tag{10}$$

where $\widehat{D}_{\min,\zeta} = \min\{\widehat{D}^{\phi}_{1,\zeta}, ..., \widehat{D}^{\phi}_{k,\zeta}\}$ and $c_\zeta = c_\zeta(k, p^*_{2,\zeta})$, with $\zeta$ being *st* or *co*, is again a suitably defined constant so that (8) is satisfied. In order to derive the expression for the probabilities of correct selection and their infimum, we need the distributional properties of the random quantities appeared on them, namely, the asymptotic distributions of the estimated $\varphi-$entropies and $\phi-$divergences.

# 3 Asymptotical results of $\widehat{H}^{\varphi}_i$, $\widehat{D}^{\phi}_{i,st}$ and $\widehat{D}^{\phi}_{i,co}$

In the propositions that follow, the asymptotic distributions of the estimated $\widehat{H}^{\varphi}_i$, $\widehat{D}^{\phi}_{i,st}$ and $\widehat{D}^{\phi}_{i,co}$, $i = 1, ..., k$, are derived. To state the results, the following notation is necessary. Denote by $W^{(i)}_0$ and $W^{(i)}_\zeta$, respectively, the vectors $W^{(i)}_0 = (W^{(i)}_{01}, ..., W^{(i)}_{0M})^t$ and $W^{(i)}_\zeta = (W_{i1\zeta}, ..., W_{iM\zeta})^t$, $i = 1, ..., k$, $\zeta$ being *st* or *co*, with elements

$$W^{(i)}_{0j} = \frac{\partial}{\partial\theta_{0j}} D^{\phi}_{i,co} = \int f'_{\theta_0(j)} \phi\left(\frac{f_{\theta_i}}{f_{\theta_0}}\right) d\mu - \int \frac{f_{\theta_i} f'_{\theta_0(j)}}{f_{\theta_0}} \phi'\left(\frac{f_{\theta_i}}{f_{\theta_0}}\right) d\mu, \tag{11}$$

for $i = 1, ..., k$, $j = 1, ..., M$, and

$$W_{ij\zeta} = \frac{\partial}{\partial\theta_{ij}} D^{\phi}_{i,\zeta} = \int f'_{\theta_i(j)} \phi'\left(\frac{f_{\theta_i}}{f_{\theta_0}}\right) d\mu,\ i = 1, ..., k,\ j = 1, ..., M,\ \zeta \text{ being } st \text{ or } co, \tag{12}$$

with $f'_{\theta_\nu(j)}$ the partial derivative $f'_{\theta_\nu(j)} = (\partial/\partial\theta_{\nu j}) f_{\theta_\nu}$, $\nu = 0, i$ and $i = 1, ..., k$.

**Theorem 1** *Let the densities $f_{\theta_i}$ and $f_{\theta_0}$, $i = 1, ..., k$, and the convex function $\phi$ satisfying regularity conditions (R1)-(R4) and ($\phi$1)-($\phi$2), respectively. Then,*

*a) As $n_i \to \infty$, $i = 1, ..., k$,*

$$\sqrt{n_i}\left(\widehat{H}^{\varphi}_i - H^{\varphi}_i\right) \xrightarrow{L} N(0, \sigma^2_i), \qquad i = 1, ..., k,$$

*with* $\sigma_i^2 = T_i^t I^{-1}(\theta_i) T_i$, *where* $T_i^t = (t_{i1}, ..., t_{iM})$, $i = 1, ..., k$, *and* $t_{ij} = \int \varphi'(f_{\theta_i}) \frac{\partial f_{\theta_i}}{\partial \theta_{ij}} d\mu$, *for* $i = 1, ..., k$, $j = 1, ..., M$.

*b) As* $n_i \to \infty$, $i = 1, ..., k$,

$$\sqrt{n_i}\left(\widehat{D}_{i,st}^{\phi} - D_{i,st}^{\phi}\right) \xrightarrow{L} N(0, \tau_{i,st}^2), \quad i = 1, ..., k,$$

*where* $\tau_{i,st}^2 = (W_{st}^{(i)})^t I^{-1}(\theta_i) W_{st}^{(i)}$, *and* $W_{st}^{(i)}$, $i = 1, ..., k$, *is given by (12).*

*c) If* $\frac{n_0}{n_0+n_i} \to \lambda_0^{(i)} > 0$ *and* $\frac{n_i}{n_0+n_i} \to \lambda^{(i)} > 0$, *as* $n_0 \to \infty$, $n_i \to \infty$, $i = 1, ..., k$, *then we have*

$$\sqrt{n_0 + n_i}\left(\widehat{D}_{i,co}^{\phi} - D_{i,co}^{\phi}\right) \xrightarrow{L} N(0, \tau_{i,co}^2), \quad i = 1, ..., k,$$

*where*

$$\tau_{i,co}^2 = \frac{1}{\lambda_0^{(i)}} (W_0^{(i)})^t I^{-1}(\theta_0) W_0^{(i)} + \frac{1}{\lambda^{(i)}} (W_{co}^{(i)})^t I^{-1}(\theta_i) W_{co}^{(i)}, \quad i = 1, ..., k,$$

*with* $W_0^{(i)}$ *and* $W_{co}^{(i)}$, $i = 1, ..., k$, *are, respectively, the vectors defined by (11) and (12). By* $I(\theta_i)$, $i = 1, ..., k$, *and* $I(\theta_0)$ *we denote the corresponding Fisher information matrices.*

**Proof:** a) A Taylor series expansion of $H^{\varphi}(\widehat{\theta}_i)$ around $\theta_i$ gives

$$\sqrt{n_i}\left(\widehat{H}_i^{\varphi} - H_i^{\varphi}\right) = \sqrt{n_i} T_i^t (\widehat{\theta}_i - \theta_i) + o_p(1), \; i = 1, ..., k,$$

and the results follows in view of the asymptotic normality and efficiency of the maximum likelihood estimator $\widehat{\theta}_i$ of $\theta_i$, $i = 1, ..., k$.

b-c) A proof for b) can be obtained by an argument similar to that of a). To prove case c), a Taylor series expansion of $\widehat{D}_{i,co}^{\phi} = D^{\phi}(f_{\widehat{\theta}_i}, f_{\widehat{\theta}_0})$ around $\theta_i$, $\theta_0$ gives

$$\sqrt{n_0+n_i}\left(\widehat{D}_{i,co}^{\phi} - D_{i,co}^{\phi}\right) = \sqrt{n_0+n_i}(W_0^{(i)})^t(\widehat{\theta}_0 - \theta_0) + \sqrt{n_0+n_i}(W_{co}^{(i)})^t(\widehat{\theta}_i - \theta_i) + o_p(1), \quad (13)$$

for $i = 1, ..., k$. Taking into account that $\widehat{\theta}_0$ and $\widehat{\theta}_i$ are the maximum likelihood estimators of $\theta_0$ and $\theta_i$, respectively, we have that

$$\sqrt{n_0 + n_i}(\widehat{\theta}_0 - \theta_0, \widehat{\theta}_i - \theta_i) \xrightarrow{L} N(0, I_i^*), \; i = 1, ..., k, \quad (14)$$

where $I_i^*$ is the block diagonal matrix

$$I_i^* = diag\left(\frac{1}{\lambda_0^{(i)}} I^{-1}(\theta_0) \; , \; \frac{1}{\lambda^{(i)}} I^{-1}(\theta_i)\right), \; i = 1, ..., k.$$

The result is now follows from relation (13). ■

**Theorem 2** *Under the assumptions and the notation introduced in Theorem 1,*

*a) The random vector* $\left(\sqrt{n_1}\left(\widehat{H}_1^{\varphi} - H_1^{\varphi}\right), ..., \sqrt{n_k}\left(\widehat{H}_k^{\varphi} - H_k^{\varphi}\right)\right)$ *is asymptotically distributed as multivariate normal with mean* 0 *and covariance matrix* $diag(\sigma_1^2, ..., \sigma_k^2)$.

*b) The random vector* $\left(\sqrt{n_1}\left(\widehat{D}^{\phi}_{1,st} - D^{\phi}_{1,st}\right), ..., \sqrt{n_k}\left(\widehat{D}^{\phi}_{k,st} - D^{\phi}_{k,st}\right)\right)$ *is asymptotically distributed as multivariate normal with mean* 0 *and covariance matrix* $diag(\tau^2_{1,st}, ..., \tau^2_{k,st})$.

*c) The joint distribution of the random vector*

$$\left(\sqrt{n_0+n_1}\left(\widehat{D}^{\phi}_{1,co} - D^{\phi}_{1,co}\right), ..., \sqrt{n_0+n_k}\left(\widehat{D}^{\phi}_{k,co} - D^{\phi}_{k,co}\right)\right),$$

*is asymptotically multivariate normal with mean* 0 *and covariance matrix* $\Sigma = [\sigma_{ij}]_{k\times k}$, *with elements*

$$\sigma_{ij} = \begin{cases} \tau^2_{i,co}, & i=j \\ \frac{1}{\sqrt{\lambda_0^{(i)}\lambda_0^{(j)}}}(W_0^{(i)})^t I^{-1}(\theta_0)W_0^{(j)}, & i \neq j \end{cases}, \; i,j = 1, ..., k.$$

**Proof:** We will outline the proof of c). The proof of a) and b) is immediate, in view of Theorem 1 (a), (b) and independence of the estimated $\varphi-$entropies $\widehat{H}^{\varphi}_1, ..., \widehat{H}^{\varphi}_k$ and $\phi-$divergences $\widehat{D}^{\phi}_{1,st}, ..., \widehat{D}^{\phi}_{k,st}$. To prove c), we have to show that every linear combination of $\sqrt{n_0+n_i}\left(\widehat{D}^{\phi}_{i,co} - D^{\phi}_{i,co}\right)$, $i = 1, ..., k$, follows a univariate normal distribution. Because of independence of the random samples considered, the right hand side of (13) corresponds to a weighted sum of independent asymptotically normal distributions and therefore every linear combination of $\sqrt{n_0+n_i}\left(\widehat{D}^{\phi}_{i,co} - D^{\phi}_{i,co}\right)$, $i = 1, ..., k$, is asymptotically normaly distributed. Then the joint distribution of $\sqrt{n_0+n_i}\left(\widehat{D}^{\phi}_{i,co} - D^{\phi}_{i,co}\right)$, $i = 1, ..., k$, is asymptotically $k-$variate normal with mean vector 0 in view of Theorem 1. The elements $\sigma_{ij}$ of the covariance matrix are

$$\sigma_{ij} = Cov\left(\sqrt{n_0+n_i}\left(\widehat{D}^{\phi}_{i,co} - D^{\phi}_{i,co}\right), \sqrt{n_0+n_j}\left(\widehat{D}^{\phi}_{j,co} - D^{\phi}_{j,co}\right)\right), i,j = 1, ..., k.$$

After considerable algebra we obtain that

$$\sigma_{ij} = \begin{cases} \sqrt{n_0+n_i}\left[(W_0^{(i)})^t Var(\widehat{\theta}_0 - \theta_0)W_0^{(i)} + (W_{co}^{(i)})^t Var(\widehat{\theta}_i - \theta_i)W_{co}^{(i)}\right], & i=j \\ \sqrt{n_0+n_i}\sqrt{n_0+n_j}(W_0^{(i)})^t Var(\widehat{\theta}_0 - \theta_0)W_0^{(j)}, & i \neq j \end{cases},$$

for $i, j = 1, ..., k$. The proof is now completed in view of (14). ■

# 4 Probabilities of correct selection

In order to obtain the probabilities of correct selection and their infimum we need the following lemma which is proved in Menéndez et. al (2000) and it is based on the results of Curnow and Dunnett (1962). We will describe the proof for the sake of completeness.

**Lemma 1** *Let* $Z_1, ..., Z_k$ *be* $k$ *standardized normal random variables with mean* 0, *variance* 1 *and correlation coefficients* $\rho_{ij}$ *given by* $\rho_{ij} = \sum_{s,\ell=1}^{M} \beta^{(i)}_{s\ell}\beta^{(j)}_{s\ell}$, *for* $i, j = 1, ..., k$

*and $i \neq j$. Then*

$$\Pr\left(Z_i < h_i,\ i=1,...,k\right) = \int_{R^{M^2}} \prod_{i=1}^{k} \Phi\left(\frac{h_i - \sum_{s,\ell=1}^{M} \beta_{s\ell}^{(i)} y_{s\ell}}{\left(1 - \sum_{s,\ell=1}^{M} \left(\beta_{s\ell}^{(i)}\right)^2\right)^{1/2}}\right) \prod_{s,\ell=1}^{M} f(y_{s\ell}) dy_{s\ell},$$

*where $\Phi$ and $f$ are respectively the c.d.f. and the p.d.f. of the standard normal distribution.*

**Proof** Consider $k+M^2$ independent standard normal variables $X_1,...,X_k,Y_{11},...,Y_{1M},...,Y_{M1},...,Y_{MM}$, and define the random variables $Z_i$, $i=1,...,k$, as follows:

$$Z_i = \left[1 - \sum_{s,\ell=1}^{M} \left(\beta_{s\ell}^{(i)}\right)^2\right]^{1/2} X_i + \sum_{s,\ell=1}^{M} \beta_{s\ell}^{(i)} Y_{s\ell}, \quad i=1,...,k.$$

Variables $Z_i$, $i=1,...,k$, satisfy the assumptions of the lemma and the proof is completed in view of relation (1.4) of Curnow and Dunnett (1962). ■

Using the above lemma, we will obtain in the sequel the probabilities of correct selection $\Pr(CS_1|R_1)$ and $\Pr(CS_{2,\zeta}|R_{2,\zeta})$, with $\zeta$ being *st* or *co* depending on whether $\Pi_0$ is a standard or a control population.

**Theorem 3** *Based on the selection rules $R_1$ and $R_{2,\zeta}$, given by (9) and (10), respectively, we have*

*a)*

$$\Pr(CS_1|R_1) = \int_{-\infty}^{+\infty} \prod_{i=1}^{k-1} \Phi\left(\frac{\sqrt{n_{(i)}}\left(bx - H_{[i]}^{\varphi}\right)}{\sigma_{(i)}}\right) f(x)dx,$$

*where $\sigma_{(i)}^2 = Var(\sqrt{n_{(i)}}\widehat{H}_{(i)}^{\varphi})$, $i=1,...,k-1$, is obtained in Theorem 1 (a), the random variable $\widehat{H}_{(i)}^{\varphi}$ is associated with $H_{[i]}^{\varphi}$, and $n_{(i)}$, for $i=1,...,k$, denote the sample size from the population associated with $H_{[i]}^{\varphi}$. $\Phi$, $f$ are respectively the c.d.f. and the p.d.f. of the standard normal distribution and $b$, $b>1$, is the constant of the selection rule $R_1$ given by (9).*

*b)*

$$\Pr(CS_{2,st}|R_{2,st}) = \int_{-\infty}^{+\infty} \prod_{i=2}^{k} \Phi\left(\frac{\sqrt{n_{(i)}}\left(D_{[i],st}^{\phi} - c_{st}x\right)}{\tau_{(i),st}}\right) f(x)dx,$$

*with $\tau_{(i),st}^2 = Var(\sqrt{n_{(i)}}\widehat{D}_{(i),st}^{\phi})$, $i=2,...,k$, is obtained in Theorem 1 (b) and the random variable $\widehat{D}_{(i),st}^{\phi}$ is associated with $D_{[i],st}^{\phi}$, for $i=1,...,k$, while $c_{st}$, $c_{st}<1$, is the constant related to the selection rule $R_{2,st}$ given by (10).*

*c)*

$$\Pr(CS_{2,co}|R_{2,co}) = \int_{-\infty}^{+\infty}\int_{R^{M^2}}\prod_{i=2}^{k}\Phi\left(\frac{D^{\phi}_{[i],co} - c_{co}x - u_{(i)}\sum_{s,\ell=1}^{M}\beta^{(i)}_{s\ell}y_{s\ell}}{u_{(i)}\left(1-\sum_{s,\ell=1}^{M}\left(\beta^{(i)}_{s\ell}\right)^2\right)^{1/2}}\right)f(x)\prod_{s,\ell=1}^{M}f(y_{s\ell})dxdy_{s\ell},$$

*where* $u^2_{(i)} = Var\left(\widehat{D}^{\phi}_{(i),co}\right) = \tau^2_{(i),co}/\left(n_0 + n_{(i)}\right)$, $i = 2, ..., k$, *with* $\tau^2_{i,co}$ *is obtained in Theorem 1 (c), and* $c_{co}$, $c_{co} < 1$, *is the constant of the selection rule* $R_{2,co}$ *given by (10). The random variable* $\widehat{D}^{\phi}_{(i),co}$ *is associated with* $D^{\phi}_{[i],co}$, *and* $W_0^{((i))}$ *is defined by (11), in terms of* $D^{\phi}_{[i],co}$, *for* $i = 1, ..., k$, *while* $\beta^{(i)}_{s\ell}$, *are defined by* $\beta^{(i)}_{s\ell} = W^{((i))}_{0s}\left[I^{-1}_{s\ell}(\theta_0)\right]^{1/2}/\sqrt{n_0}u_{(i)}$, *with* $I^{-1}_{s\ell}(\theta_0)$ *the* $(s,\ell)-$*element of the* $M \times M$ *matrix* $I^{-1}(\theta_0)$ *and* $s, \ell = 1, ..., M$, $i, j = 1, ...k$ *with* $i \neq j$.

**Proof:** a) Taking into account relation (9),

$$\begin{aligned}\Pr(CS_1|R_1) &= \Pr(\text{Selecting } \Pi_{(k)}|R_1)\\ &= \Pr(b\widehat{H}^{\varphi}_{(k)} \geq \widehat{H}_{\max})\\ &= \Pr(b\widehat{H}^{\varphi}_{(k)} \geq \widehat{H}^{\varphi}_{(i)}, \quad \forall \ i = 1, ..., k-1)\\ &= \Pr(\bigcap_{i=1}^{k-1}\widehat{H}^{\varphi}_{(i)} \leq b\widehat{H}^{\varphi}_{(k)})\\ &= \int_{-\infty}^{+\infty}\Pr(\bigcap_{i=1}^{k-1}\widehat{H}^{\varphi}_{(i)} \leq bx)f(x)dx.\end{aligned}$$

The last equality follows from page 44 of Johnson and Kotz (1972). But from Theorem 2 (a), $\widehat{H}^{\varphi}_{(1)}, ..., \widehat{H}^{\varphi}_{(k-1)}$ are independent asymptotically normal random variables with $E(\widehat{H}^{\varphi}_{(i)}) = H^{\varphi}_{[i]}$ and $Var(\widehat{H}^{\varphi}_{(i)}) = \sigma^2_{(i)}/n_{(i)}$, $i = 1, ..., k-1$. Thus,

$$\Pr(CS_1|R_1) = \int_{-\infty}^{+\infty}\prod_{i=1}^{k-1}\Pr(\widehat{H}^{\varphi}_{(i)} \leq bx)f(x)dx,$$

which completes the proof of part (a).

The proof of part (b) can be obtained in a similar manner and it is omitted.
c) We have

$$\begin{aligned}\Pr(CS_{2,co}|R_{2,co}) &= \Pr(\text{Selecting } \Pi_{(1)}|R_{2,co})\\ &= \Pr(c_{co}\widehat{D}^{\phi}_{(1),co} \leq \widehat{D}_{\min,co})\\ &= \Pr(\widehat{D}^{\phi}_{(i),co} \geq c_{co}\widehat{D}^{\phi}_{(1),co}, \quad \forall \ i = 2, ..., k)\\ &= \Pr(\bigcap_{i=2}^{k}\widehat{D}^{\phi}_{(i),co} \geq c_{co}\widehat{D}^{\phi}_{(1),co})\\ &= \int_{-\infty}^{+\infty}\Pr(\bigcap_{i=2}^{k}\widehat{D}^{\phi}_{(i),co} \geq c_{co}x)f(x)dx,\end{aligned} \tag{15}$$

where $f$ is the p.d.f. of the standard normal distribution. The last equality follows again from page 44 of Johnson and Kotz (1972). But taking into account Theorem 2 (c), $\widehat{D}^{\phi}_{(2),co}, ..., \widehat{D}^{\phi}_{(k),co}$, have jointly an asymptotic $(k-1)-$dimensional normal distribution with $E(\widehat{D}^{\phi}_{(i),co}) = D^{\phi}_{[i],co}$, $Var(\widehat{D}^{\phi}_{(i),co}) = \frac{\tau^2_{(i),co}}{n_0+n_{(i)}}$, and $Cov(\widehat{D}^{\phi}_{(i),co}, \widehat{D}^{\phi}_{(j),co}) = \frac{1}{n_0}\left(W_0^{((i))}\right)^t I^{-1}(\theta_0) W_0^{((j))}$, for $i, j = 2, ..., k$ and $i \neq j$. In order to apply Lemma 1, let $Z_i = \frac{\widehat{D}^{\phi}_{(i),co} - D^{\phi}_{[i],co}}{u_{(i)}}$ and $h_i(x) = \frac{c_{co}x - D^{\phi}_{[i],co}}{u_{(i)}}$, with $u^2_{(i)} = Var\left(\widehat{D}^{\phi}_{(i),co}\right) = \tau^2_{(i),co}/\left(n_0 + n_{(i)}\right)$ for $i = 2, ..., k$. In this case $Z_2, ..., Z_k$ are $k-1$ standardized normal random variables with mean 0, variance 1 and the correlation coefficient $\rho_{ij}$ of $Z_i$ and $Z_j$ is of the form $\rho_{ij} = \sum_{s,\ell=1}^{M} \beta^{(i)}_{s\ell}\beta^{(j)}_{s\ell} = \left(W_0^{((i))}\right)^t I^{-1}(\theta_0) W_0^{((j))}/n_0 u_{(i)} u_{(j)}$, with $\beta^{(i)}_{s\ell} = W^{((i))}_{0s}\left[I^{-1}_{s\ell}(\theta_0)\right]^{1/2}/\sqrt{n_0}u_{(i)}$, $s, \ell = 1, ..., M$, $i, j = 1, ...k$ with $i \neq j$. In this context relation (15) can be stated as

$$\begin{aligned}\Pr(CS_{2,co}|R_{2,co}) &= \int_{-\infty}^{+\infty} \Pr(Z_i \geq h_i(x), i = 2, ..., k) f(x)dx \\ &= \int_{-\infty}^{+\infty} \Pr(Z_i \leq -h_i(x), i = 2, ..., k) f(x)dx,\end{aligned}$$

because probabilities and correlations are unchanged by replacing each $Z_i$ by $-Z_i$ and taking into account the fact that the event $\bigcap_{i=2}^{k} (Z_i \leq -h_i(x))$ is equivalent to the event $\max_{2\leq i\leq k} (Z_i + h_i(x)) \leq 0$ (cf. Johnson and Kotz (1972), p. 45). The last equation leads, by using Lemma 1, to the desired result. ■

**Remark 1** The above theorem helps to determine the constants $b$, $c_{co}$ and $c_{st}$ of the selection rules stated in relations (9) and (10). Indeed, based on the case (a) of the theorem we have that

$$\Pr(CS_1|R_1) = \int_{-\infty}^{+\infty} \prod_{i=1}^{k-1} \Phi\left(\frac{\sqrt{n_{(i)}}\left(bx - H^{\varphi}_{[i]}\right)}{\sigma_{(i)}}\right) f(x)dx,$$

and taking into account that $H^{\varphi}_{[1]} \leq H^{\varphi}_{[2]} \leq ... \leq H^{\varphi}_{[k]}$ and the assumption $H^{\varphi}_{[i]} \geq 0$, $i = 1, ..., k$, we get that

$$\Pr(CS_1|R_1) \geq \int_{-\infty}^{+\infty} \prod_{i=1}^{k-1} \Phi\left(\frac{\sqrt{n_{(i)}}\left(bx - H^{\varphi}_{[k]}\right)}{\sigma_{(i)}}\right) f(x)dx.$$

Hence, the constant $b$ required for the definition of the selection rule (9) can be determined, in view of (7), so as to satisfy

$$\int_{-\infty}^{+\infty} \prod_{i=1}^{k-1} \Phi\left(\frac{\sqrt{n_{(i)}}\left(bx - H^{\varphi}_{[k]}\right)}{\sigma_{(i)}}\right) f(x)dx = p^*_1, \tag{16}$$

where $p_1^*$ is defined by (7). In a similar manner, the constants $c_{st}$ and $c_{co}$ of the selection rule (10), can be obtained by the solution of the following equations,

$$\int_{-\infty}^{+\infty} \prod_{i=2}^{k} \Phi\left(\frac{\sqrt{n_{(i)}}(D_{[1],st}^{\phi} - c_{st}x)}{\tau_{(i),st}}\right) f(x)dx = p_{2,st}^*, \tag{17}$$

and

$$\int_{-\infty}^{+\infty}\int_{R^{M^2}} \prod_{i=2}^{k} \Phi\left(\frac{D_{[1],co}^{\phi} - c_{co}x - u_{(i)} \sum_{s,\ell=1}^{M} \beta_{s\ell}^{(i)} y_{s\ell}}{u_{(i)}\left(1 - \sum_{s,\ell=1}^{M} \left(\beta_{s\ell}^{(i)}\right)^2\right)^{1/2}}\right) f(x) \prod_{s,\ell=1}^{M} f(y_{s\ell})dxdy_{s\ell} = p_{2,co}^*. \tag{18}$$

**Remark 2** The last equation can be considerably simplified if the parameters $\theta_0$ and $\theta_i$, $i = 1, ..., k$, of the populations are univariate. In this case an application of $\beta_{s\ell}^{(i)} = \frac{W_{0s}^{((i))}\left[I_{s\ell}^{-1}(\theta_0)\right]^{1/2}}{\sqrt{n_0}u_{(i)}}$ for $M = 1$ gives that

$$\beta_i \equiv \beta_{11}^{(i)} = W_{01}^{((i))}\left[I_{11}^{-1}(\theta_0)\right]^{1/2} / \sqrt{n_0}u_{(i)},$$

and the correlation coefficient of $\frac{\widehat{D}_{(i),co}^{\phi} - D_{[i],co}^{\phi}}{u_{(i)}}$ and $\frac{\widehat{D}_{(j),co}^{\phi} - D_{[j],co}^{\phi}}{u_{(j)}}$ is of the form $\rho_{ij} = \beta_i\beta_j$, $i, j = 2, ..., k$ and $i \neq j$. In this context equation (18) is simplified to

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \prod_{i=2}^{k} \Phi\left(\frac{D_{[1],co}^{\phi} - c_{co}x - u_{(i)}\beta_i y}{u_{(i)}\left(1 - \beta_i^2\right)^{1/2}}\right) f(x)f(y)dxdy = p_{2,co}^*. \tag{19}$$

# 5 Illustrative Examples

**Example 1.** (*Homoscedastic Normal Populations*)

Consider $k$ multivariate normal $N_M(\mu_i, \Sigma)$ populations $\Pi_i$, $i = 1, ..., k$, and denote by $\Pi_0$ the standard or control population described by $N_M(\mu_0, \Sigma)$ distribution with known variance-covariance matrix $\Sigma > 0$. Consider the Kullback-Leibler divergence between $\Pi_i$ and the standard or control population $\Pi_0$. This divergence is denoted by $D_{i,st}$ or $D_{i,co}$ and can be obtained from (3), for $\phi(u) = u \log u$, $u > 0$. It can be easily seen that

$$D_{i,st} = D_{i,co} = \frac{1}{2}(\mu_i - \mu_0)^t \Sigma^{-1}(\mu_i - \mu_0), \quad i = 1, ..., k, \tag{20}$$

i.e, a half of the Mahalanobis distance between populations $\Pi_i$ and $\Pi_0$. Consider firstly the case of standard population $\Pi_0$, that is, the mean vector $\mu_0$ is known. In this case

consider a random sample of size $n_i$ from the population $N_M(\mu_i, \Sigma)$ and denote by $\overline{\mathbf{X}}_i$, $i = 1, ..., k$, the sample mean vector. Then, in view of (20), the estimator of $D_{i,st}$ is given by

$$\widehat{D}_{i,st} = \frac{1}{2}\left(\overline{\mathbf{X}}_i - \mu_0\right)^t \Sigma^{-1}\left(\overline{\mathbf{X}}_i - \mu_0\right), \quad i = 1, ..., k. \tag{21}$$

Taking into account Theorem 1 (b), the asymptotic distribution of $\sqrt{n_i}\,\widehat{D}_{i,st}$ is univariate normal with variance $\tau^2_{i,st} = \left(W^{(i)}_{st}\right)^t I^{-1}(\mu_i) W^{(i)}_{st}$, where $I(\mu_i)$ is the Fisher information matrix associated to $N_M(\mu_i, \Sigma)$ and $W^{(i)}_{st} = \frac{\partial D_{i,st}}{\partial \mu_i}$, for $i = 1, ..., k$. We can obtain that $I(\mu_i) = \Sigma^{-1}$ and $W^{(i)}_{st} = \frac{\partial D_{i,st}}{\partial \mu_i} = \Sigma^{-1}(\mu_i - \mu_0)$, for $i = 1, ..., k$. Hence,

$$\tau^2_{i,st} = 2D_{i,st}, \quad i = 1, ..., k. \tag{22}$$

Using Theorem 3(b), relation (22) and the rank $D_{[1],st} \leq D_{[2],st} \leq ... \leq D_{[k],st}$, we have that

$$\int_{-\infty}^{+\infty} \prod_{i=2}^{k} \Phi\left(\frac{\sqrt{n_{(i)}}(D_{[1],st} - c_{st}x)}{\sqrt{2D_{[k],st}}}\right) f(x)dx = p^*_{2,st}. \tag{23}$$

This relation can be used in order to determine the constant $c_{st}$ of the selection rule (10). If the sample sizes are equal, that is, $n_1 = n_2 = ... = n_k = n$, then (23) is simplified as follows,

$$\int_{-\infty}^{+\infty} \Phi^{k-1}\left(\frac{\sqrt{n}(D_{[1],st} - c_{st}x)}{\sqrt{2D_{[k],st}}}\right) f(x)dx = p^*_{2,st}. \tag{24}$$

Consider now the case of control population $\Pi_0$, that is, the mean vector $\mu_0$ is unknown. In this case consider an additional random sample of size $n_0$ from the population $N_M(\mu_0, \Sigma)$ and denote by $\overline{\mathbf{X}}_0$, the sample mean vector. Then, in view of (20), the estimator of $D_{i,co}$ is given by

$$\widehat{D}_{i,co} = \frac{1}{2}\left(\overline{\mathbf{X}}_i - \overline{\mathbf{X}}_0\right)^t \Sigma^{-1}\left(\overline{\mathbf{X}}_i - \overline{\mathbf{X}}_0\right), \quad i = 1, ..., k. \tag{25}$$

The asymptotic distribution of $\sqrt{n_0 + n_i}\,\widehat{D}_{i,co}$ is univariate normal according to Theorem 1(c), with variance

$$\tau^2_{i,co} = 2\left(\frac{1}{\lambda^{(i)}_0} + \frac{1}{\lambda^{(i)}}\right) D_{i,co}, \quad i = 1, ..., k,$$

when $\frac{n_0}{n_0+n_i} \to \lambda^{(i)}_0 > 0$ and $\frac{n_i}{n_0+n_i} \to \lambda^{(i)} > 0$, as $n_0 \to \infty$, $n_i \to \infty$, $i = 1, ..., k$. After a little algebra, we can see that the quantities $u^2_{(i)}$ and $\beta^{(i)}_{s\ell}$ of Theorem 3(c) are given by

$$u^2_{(i)} = \frac{2(n_0 + n_{(i)})}{n_0 n_{(i)}} D_{[i],co}, \quad \text{and} \quad \beta^{(i)}_{s\ell} = \frac{W^{((i))}_{0s}\sigma^{1/2}_{s\ell}}{\sqrt{n_0}u_{(i)}}, \tag{26}$$

with $W_{0s}^{((i))}$ the $s$th element of the vector $W_0^{((i))} = \frac{\partial D_{[i],co}}{\partial \mu_0} = \Sigma^{-1}\left(\mu_0 - \mu_{(i)}\right)$ and $\sigma_{s\ell}$ the $(s,\ell)-$element of the variance-covariance matrix $\Sigma$ for $s,\ell = 1,...,M$, $i = 1,...,k$ and $M$ the dimension of multivariate normal distributions. Using Theorem 3(c), relations (26) and the order $D_{[1],co} \leq D_{[2],co} \leq ... \leq D_{[k],co}$, we have that

$$\int\limits_{R^{M^2+1}} \prod_{i=2}^{k} \Phi\left(\frac{\sqrt{n_0 n_{(i)}}\left(D_{[1],co}^{\phi} - c_{co}x - \sum\limits_{s,\ell=1}^{M} W_{0s}^{((i))}\sigma_{s\ell}^{1/2} y_{s\ell}\right)}{\left(2n_{(i)}^{*} D_{[k],co}^{\phi} - n_{(i)} \sum\limits_{s,\ell=1}^{M}\left(W_{0s}^{((i))}\sigma_{s\ell}^{1/2}\right)^2\right)^{1/2}}\right) f(x) \prod_{s,\ell=1}^{M} f(y_{s\ell})dy_{s\ell}dx = p_{2,co}^{*},$$

with $n_{(i)}^{*} = n_0 + n_{(i)}$, $i = 2,...,k$. This relation can be, at least theoretically, used in order to determine the constant $c_{co}$ of the selection rule (10). The above equation is considerably simplified in the case of univariate normal populations.

Consider the simplest case where the populations $\Pi_i$ are characterized by the univariate normal distributions $N(\mu_i, \sigma^2)$, with $\mu_i$ unknown, $i = 1,...,k$ and the control population $\Pi_0$ is $N(\mu_0, \sigma^2)$, with $\mu_0$ unknown and $\sigma^2$ known. If $f_{\mu_i}$ and $f_{\mu_0}$ are the densities of $N(\mu_i, \sigma^2)$ and $N(\mu_0, \sigma^2)$, respectively, then

$$D_{i,co} = \int\limits_{-\infty}^{+\infty} f_{\mu_i} \log \frac{f_{\mu_i}}{f_{\mu_0}} dx = \frac{\left(\mu_i - \mu_0\right)^2}{2\sigma^2}, i = 1,...,k.$$

Thus $W_0^{(i)} = \frac{\partial D_{i,co}}{\partial \mu_0} = -\frac{\mu_i - \mu_0}{\sigma^2}$ and $I\left(\mu_0\right) = I\left(\mu_i\right) = \frac{1}{\sigma^2}$, $i = 1,...,k$. Hence, taking into account (26), we have

$$u_{(i)}^2 = \frac{\left(n_0 + n_{(i)}\right)}{n_0 n_{(i)}} \frac{\left(\mu_{(i)} - \mu_0\right)^2}{\sigma^2}, \quad \beta_i \equiv \beta_{11}^{(i)} = \frac{W_{01}^{((i))}\left[I_{11}^{-1}(\theta_0)\right]^{1/2}}{\sqrt{n_0} u_{(i)}} = -\frac{\sqrt{n_{(i)}}\left(\mu_{(i)} - \mu_0\right)}{\sqrt{n_0 + n_{(i)}}\left|\mu_{(i)} - \mu_0\right|},$$

for $i = 1,...,k$. Then, taking into account relation (19), we have that

$$\int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{+\infty} \prod_{i=2}^{k} \Phi\left(\frac{\sigma\sqrt{n_0 n_{(i)}}\left(D_{[1],co} - c_{co}x\right) + \sqrt{n_{(i)}}\left(\mu_{(i)} - \mu_0\right) y}{\sqrt{n_0}\left|\mu_{(i)} - \mu_0\right|}\right) f(x)f(y)dxdy = p_{2,co}^{*}.$$

Various approximations of the above integral had been proposed but many mathematical packages allow also this computation.

**Example 2.** Remaining in the framework of the previous example, consider now the real data set used in Section 4 of Thabane and Safiul Haq (1999). These data are provided by Hoque et al. (1997) in their cardio-vascular research, devoted to a preliminary experiment on hearts of rats to determine wheather $Na^{+}-H^{+}$ exchange (NHE) inhibitor 4-isopropyl-3-methylsulfonylbezoyl-guanidine (HOE 642) protects against Lysophosphatidylcholine (LPC)-induced cardiac injuries. For details of the experiment we refer the reader

to Thabane and Safiul Haq (1999) and Hoque et al. (1997). Based on Thabane and Safiul Haq (1999), in this experiment a large group of hearts of rats was divided into small groups. The first group, denoted by $\Pi_0$, represents the group of healthy hearts and is considered to be the standard population. The rest of the hearts were divided into 3 more groups, which we denote by $\Pi_1$, $\Pi_2$ and $\Pi_3$. Three different treatments, (i.e. dose level of a remedy) have applied to the hearts of the three groups $\Pi_1$, $\Pi_2$ and $\Pi_3$, respectively. At the end of the experiment, measurements of ATP, creatine phosphate (CrP) and lactate (LAC) were taken from frozen cardiac tissues. The main object is to select the best treatment relative to the standard population $\Pi_0$. We assume that random samples of size $n_i = n = 6$ are coming from each of the normal $N_3(\mu_i, \Sigma)$ populations $\Pi_i$, with known common variance-covariance matrix $\Sigma$ and unknown mean vectors $\mu_i$, $i = 1, 2, 3$. Based on the data of Table 1 of Thabane and Safiul Haq (1999), the following matrix

$$\Sigma = \begin{pmatrix} 7.9154 & 7.4655 & 0.1271 \\ 7.4655 & 24.2169 & 0.0354 \\ 0.1271 & 0.0354 & 3.5078 \end{pmatrix},$$

is the pooled sample covariance matrix obtained from the sample covariance matrices given by Thabane and Safiul Haq (1999). Denote by $\overline{\mathbf{X}}_i$ the sample mean vector obtained from a sample of size $n = 6$ from the population $N_3(\mu_i, \Sigma)$, $i = 1, 2, 3$. Let $\overline{\mathbf{X}}_1 = (12.9, 12.2, 9.1)^t$, $\overline{\mathbf{X}}_2 = (16.9, 18.6, 5.6)^t$ and $\overline{\mathbf{X}}_3 = (18.6, 15.7, 3.1)^t$. Consider also the standard population $\Pi_0$, described by $N_3(\mu_0, \Sigma)$ distribution with $\mu_0 = (26.2, 18.8, 3.1)^t$. If we have decided to use Kullback-Leibler divergence, then the selection rule leading to the selection of the best population, which is closest to the standard population $\Pi_0$, is in view of (10),

$$\text{Select } \Pi_i \text{ if and only if } c_{st}\widehat{D}_{i,st} \leq \widehat{D}_{\min,st},\ c_{st} < 1,$$

where $\widehat{D}_{\min,st} = \min\{\widehat{D}_{1,st}, \widehat{D}_{2,st}, \widehat{D}_{3,st}\}$. $\widehat{D}_{i,st} = \frac{1}{2}\left(\overline{\mathbf{X}}_i - \mu_0\right)^t \Sigma^{-1}\left(\overline{\mathbf{X}}_i - \mu_0\right)$, $i = 1, 2, 3$, has been obtained by (21) and the constant $c_{st}$ can be determined by (24), if the unknown $D_{[i],st}$ are replaced by their estimators $\widehat{D}_{[i],st}$, $i = 1, 2, 3$. We can easily obtain that

$$\widehat{D}_{1,st} = 17.76, \quad \widehat{D}_{2,st} = 8.634, \quad \widehat{D}_{3,st} = 4.132.$$

It is obvious that

$$\widehat{D}_{[1],st} = 4.132 \leq \widehat{D}_{[2],st} = 8.634 \leq \widehat{D}_{[3],st} = 17.76.$$

Because the sample sizes are equal, $n = 6$, the constant $c_{st}$ can be determined, in view of (24), by the solution of the equality

$$\int_{-\infty}^{+\infty} \Phi^2\left(1.6982 - 0.411 c_{st} x\right) f(x) dx = p^*_{2,st}.$$

We observe that the above integral is a decreasing function of the constant $c_{st}$. In the following table, the value of the constant $c_{st}$ is determined for some values of the probability of correct selection $p^*_{2,st}$,

| $p^*_{2,st}$ | 0.91299 | 0.91156 | 0.90867 | 0.90652 | 0.90089 | 0.89363 | 0.89158 |
|---|---|---|---|---|---|---|---|
| $c_{st}$ | 0.1 | 0.2 | 0.4 | 0.5 | 0.7 | 0.9 | 0.95 |

If we decide to use probability of correct selection $p^*_{2,st} = 0.90089$, then the value $c_{st} = 0.7$. The selection rule (10) gives that

$$c_{st}\widehat{D}_{3,st} = 2.8924 \leq 4.132 = \widehat{D}_{\min,st},$$

and the third population $\Pi_3$ is the best population with respect to the standard population $\Pi_0$. If we consider $p^*_{2,st} = 0.90867$, then $c_{st} = 0.4$. We observe that

$$\begin{matrix} c_{st}\widehat{D}_{3,st} = 1.6528 \\ c_{st}\widehat{D}_{2,st} = 3.4536 \end{matrix} \leq 4.132 = \widehat{D}_{\min,st}.$$

Hence, populations $\Pi_3$ and $\Pi_2$ can be selected as the best populations in comparison with the standard population $\Pi_0$, with a high probability of correct selection equal to 0.90867. Hoque et al. (1997) analyzed the same real data set by comparing the means of each group with the mean of the population $\Pi_0$ for each characteristic. Their conclusion was that $\Pi_3$ was the best population for the cardio-vascular problem. Thabane and Safiul Haq (1999) considered $\Pi_2$ as the best population by using their Bayesian approach to study the problem of selecting the best population. These two methods are not accompanied their conclusions with a probabilistic statement related to the correct selection.

We have to notice at this point, that the asymptotic results of the paper are used, in this example, to analyze a small sample real data set. However, although we have presented this example as a simple illustration of the results of this paper, the conclusions are in harmony with the similar ones by Hoque et al. (1997) and Thabane and Haq (1999).

**Example 3.** (*Regular Exponential Populations*)

Consider $k$ independent populations $\Pi_i$, $i = 1, ..., k$, described by the regular exponential models

$$f_{\theta_i}(x) = \exp\left\{\sum_{j=1}^{M} T_j(x)\theta_{ij} - b(\theta_i)\right\}, \quad \theta_i \in \Theta \subseteq R^M, \quad i = 1, ..., k, \tag{27}$$

with

$$b(\theta_i) = \ln \int \exp\left\{\sum_{j=1}^{M} T_j(x)\theta_{ij}\right\} d\mu, \quad i = 1, ..., k,$$

and $\mu$ a probability measure on the Euclidean space. Then, the Shannon entropy, defined by (1), for the above exponential models is given by (cf. Menendez (2000)),

$$H_i = H(\theta_i) = -\sum_{j=1}^{M} \theta_{ij}\frac{\partial b(\theta)}{\partial \theta_{ij}} + b(\theta).$$

If $\widehat{\theta}_i$ is the maximum likelihood estimator of $\theta_i$, then the estimator $\sqrt{n_i}\widehat{H}_i$, $i = 1, ..., k$, follows, in view of Theorem 1(a), a normal distribution with variance $\sigma_i^2 = T_i^t I^{-1}(\theta_i) T_i$, where $T_i^t = (t_{i1}, ..., t_{iM})$ and

$$t_{ir} = -\sum_{j=1}^{M} \theta_{ij} \frac{\partial^2 b(\theta)}{\partial\theta_{ir}\partial\theta_{ij}}, \quad I(\theta_i) = \left[\frac{\partial^2 b(\theta)}{\partial\theta_{ir}\partial\theta_{i\ell}}\right]_{M\times M}, \quad i = 1, ..., k, \quad r, \ell = 1, ..., M.$$

These asymptotic results and relation (16) can be used in order to determine the constant $b$ of the selection rule (9).

Consider now that $k$ independent populations $\Pi_i$, $i = 1, ..., k$, are described by the well known gamma distribution with parameters $\alpha_i$ and $\beta_i$, $i = 1, ..., k$, and densities

$$f_{(\alpha_i,\beta_i)}(x) = \exp\{-\alpha_i x + \beta_i \ln x + (\beta_i + 1)\ln\alpha_i - \ln\Gamma(\beta_i + 1)\}, i = 1, ..., k.$$

These densities are immediately obtained from (27) for $b(\alpha_i, \beta_i) = -(\beta_i + 1)\ln\alpha_i - \ln\Gamma(\beta_i + 1)$, $i = 1, ..., k$, $T_1(x) = \ln x$ and $T_2(x) = -x$. The Shannon entropy of $f_{(\alpha_i,\beta_i)}(x)$ is given by

$$H(\alpha_i, \beta_i) = -\beta_i\Psi(\beta_i + 1) + (\beta_i + 1) + \ln\frac{\Gamma(\beta_i + 1)}{\alpha_i}, i = 1, ..., k,$$

where $\Gamma$ and $\Psi = \Gamma'/\Gamma$ are the well known gamma and digamma functions, respectively. The maximum likelihood estimators $\widehat{\alpha}_i, \widehat{\beta}_i$ of the parameters $\alpha_i, \beta_i$ can be obtained by the solution of the equations

$$\Psi(\widehat{\alpha}_i) - \ln\widehat{\alpha}_i - \ln(\widetilde{X}_i/\overline{X}_i) = 0 \quad \text{and} \quad \widehat{\beta}_i + 1 = \overline{X}_i/\widehat{\alpha}_i, \quad i = 1, ..., k,$$

where $\overline{X}_i$ is the sample mean and $\widetilde{X}_i = \prod_{j=1}^{n} X_{ij}^{1/n}$ is the sample geometric mean based on a random sample $X_{i1}, ..., X_{in_i}$ of size $n_i$ from the gamma distribution $f_{(\alpha_i,\beta_i)}(x)$, $i = 1, ..., k$. Applying Theorem 1(a) we have that $\sqrt{n_i}H(\widehat{\alpha}_i, \widehat{\beta}_i)$ has asymptotically normal distribution with variance (cf. Menendez (2000)),

$$\sigma_i^2 = \frac{\delta(\beta_i + 1) + (\beta_i + 1)[1 - \beta_i\delta(\beta_i + 1)]^2 + 2[1 - \beta_i\delta(\beta_i + 1)]}{(\beta_i + 1)\delta(\beta_i + 1) - 1},$$

where

$$\delta(\beta_i + 1) = \frac{\Psi'(\beta_i + 1)\Gamma(\beta_i + 1) - [\Psi(\beta_i + 1)]^2}{\Gamma(\beta_i + 1)}, \quad i = 1, ..., k.$$

The formulas obtained above and relation (16) can be used in order to use the rule (9) for the selection of the best from $k$ independent gamma populations.

## Acknowledgements

Research supported in part by Greek Ministry for Development, General Secretariat for Research and Technology, and Spanish Foreign Office, through a bilateral program of scientific and technologic cooperation (1999-2000) and DGI BFM2000-0800.

# 6 References


Ali, S. M. and Silvey, S. D. (1966). A general class of coefficients of divergence of one distribution from another. *J. Roy. Statist. Soc.* B, **28**, 131-142.

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

Burbea, J. and Rao, C.R. (1982). On the convexity of some divergence measures based on entropy functions. *IEEE Trans. Inform. Theory*, **IT-28**, 489-495.

Chattopadhyay, A. K. (1981). Selecting the normal population with largest (smallest) value of Mahalanobis distance from the origin. *Commun. Statist.-Theor. Meth.*, **10**, 31-37.

Chen, P. and Sobel, M. (1987). An integrated formulation for selecting the $t$ best of $k$ normal populations. *Commun. Statist.-Theor. Meth.*, **16**, 121-146.

Csiszár, I. (1963) . Eine informationstheoretische ungleichung und ihre anwendung auf den beweis der ergodizitat von Markoffschen Ketten. *Mayar Tud. Acad. Mat. Kutato Int. Kozl.*, **8**, 85-108.

Curnow, R. N. and Dunnett, C. W. (1962). The numerical evaluation of certain multivariate normal integrals. *Ann. Math. Statist.*, **33**, 571-579.

Dudewicz, E. J. and van der Meulen, E. C. (1983). Entropy-based statistical inference, II: Selection-of-the-best/complete ranking for continuous distributions on (0,1), with applications to random number generators. *Statistics and Decisions*, **1**, 131-145.

Gibbons, J. D., Olkin, I. and Sobel, M. (1977). *Selecting and ordering populations: a new statistical methodology.* Wiley, New York,

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

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

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

Gupta, S. S. and Studden, W. J. (1970). On some selection and ranking procedures with applications to multivariate populations. In: *Essays in Probability and Statistics* (Eds. R. C. Bose et al.), pp. 327-338, The University of North Carolina Press, Chapel Hill, NC.

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

Gupta, S. S. and Panchapakesan, S. (1979). *Multiple decision procedures: theory and methodology of selecting and ranking populations.* Wiley, New York.

Gupta, S. S. and Panchapakesan, S. (1987). Statistical selection procedures in multivariate models, In *Advances in Multivariate Statistical Analysis* (Ed. A. K. Gupta), pp.141-160, D. Reidel Publishing Company.

Hoque, A. N. E., Haist, J. V. and Karmazyn, M. (1997). $Na^+-H^+$ exchange inhibition protects against mechanical, ultrastructural and biochemical impairment induced by low concentration of Lysophosphatidylcholine in isolated rat hearts. *Circulation Research*, **80**, 95-102.

Jeyaratnam, S. and Panchapakesan, S. (1997). An integrated formulation for selecting the best from several normal populations in terms of the absolute values of their means: common known variance case. In: *Advances in Statistical Decision Theory and Applications* (Eds. S. Panchapakesan and N. Balakrishnan), pp. 277-287, Birkhäuser, Boston.

Johnson, N.L. and Kotz, S. (1972). *Distributions in Statistics: Continuous Multivariate Distributions.* Wiley, New York.

Liese, F. and Miescke, K. J. (1999). Selection of the best population in models with nuisance parameters. *Mathematical Methods of Statistics*, **8**, 371-396.

Liese, F. and Vajda, I. (1987). *Convex Statistical Distances.* Teubner, Leipzig.

Menéndez, M. L. (2000). Shannon's entropy in exponential families: Statistical applications. *Applied Mathematics Letters*, **13**, 37-42.

Menéndez, M. L., Pardo, L. and Zografos, K. (2000). Minimum $\varphi-$dissimilarity approach in classification. Submitted.

Morales, D., Pardo, L. and Vajda, I. (1997). Some new statistics for testing hypotheses in parametric models. *J. Multivariate Analysis*, **62**, 137-168.

Panchapakesan, S. and Santner, T. J. (1977). Subset selection procedures for $\Delta_p$-superior populations. *Commun. Statist.-Theor. Meth.*, **6**, 1081-1090.

Serfling, R. J. (1980). *Approximations theorems of mathematical statistics.* Wiley, New York.

Thabane, L. and Haq, Safiul M. (1999). On Bayesian selection of the best normal population using the Kullback-Leibler divergence measure. *Statistica Neerlandica*, **53**, 342-360.