Distributional and efficiency results for subset selection

EINDHOVEN UNIVERSITY OF TECHNOLOGY
Department of Mathematics and Computing Science

Memorandum COSOR 94-03

Distributional and efficiency results for subset selection

P. van der Laan

Eindhoven, January 1994
The Netherlands

Distributional and efficiency results for subset selection

Paul van der Laan
Eindhoven University of Technology
Department of Mathematics and Computing Science
The Netherlands

Summary

Assume $k(k \geq 2)$ populations are given. The associated independent random variables have continuous distribution functions with an unknown location parameter. The statistical selection goal is to select a non-empty subset which contains the best population, that is the population with largest value of the location parameter with confidence level $P^*(k^{-1} < P^* < 1)$. Some distributional results for subset selection are given and proved. Explicit expressions for expectation and variance of the subset size are presented. Also some distributional and efficiency results are given concerning a generalized selection goal using subset selection. This generalized subset selection goal is to select a non-empty subset of populations that contains at least one ε-best population. An ε-best population is any population with a parameter value within $\varepsilon(\varepsilon \geq 0)$ of the largest parameter value. The subset selection goal of Gupta is a special case of the generalized selection goal.

AMS Classification: 62F07.

Key Words: Selection; subset selection; location parameter; subset size; expectation; variance; generalized selection goal.

1 Introduction

A general goal in many statistical experiments is to indicate the best population from a set of $k(k \geq 2)$ populations $\pi_1, ..., \pi_k$. The associated independent random variables, which may be sample means, have continuous distribution functions with an unknown location parameter. In the context of location parameter the best population is defined as the population with largest value of the location parameter. The main approaches in handling with selection of the best population are the subset selection approach of Gupta (1965) and the indifference zone approach by Bechhofer (1954). The subset selection procedure, which will be considered in this paper, has as its goal the selection of a non-empty subset, as small as possible, containing the best population with a given probability. This probability requirement has to be met for all possible parameter values. A possible practical "objection" to subset selection procedures is of the type "large subsets are sometimes the result". For a strong probability requirement one pays, for fixed sample sizes, with a large size of the selected subset, where the size of the subset is defined as the number of populations in the subset. These large subsets are mainly due to the fact that the probabiltiy requirement has to be met for the least favourable configuration (LFC) of the parameters. For the location model the LFC is, in many cases, the configuration consisting of equal parameter values for the k populations. The performance of selection procedures can be improved by either increasing the sample sizes or by weakening the probability requirement. From an application point of view the (expected) subset size is a characteristic and crucial quantity. After a short description of Gupta's subset selection procedure in Section 2 some general distributional results for the subset size will be given in Section 3. Chen an Sobel (1987) give some distributional results for normal populations. Section 4 contains some results for the expectation and variance of the subset size. An alternative to weakening the probability requirement and increasing the sample sizes, in order to get smaller subset sizes, is to be content when the subset contains an ε-best population. Here, an ε-best population is any population with a parameter value within $\varepsilon(\varepsilon > 0)$ of the value of the largest parameter. This extension was considered by van der Laan (1992a and b). Section 5 contains some efficiency results based on the results of Section 4. Finally, some remarks concerning recent research results are made in Section 6.

2 Subset selection

Let, for $i = 1, ..., k, X_{i1}, ..., X_{in}$ be a sample from π_i and suppose these k samples are independent. It is assumed that X_{i1} has a continuous distribution function $F(x; \theta_i)$, where $\theta_i \in \Theta \subset I\!R$ is an unknown location parameter, $i = 1, ..., k$. It is further assumed that $F(x; \theta)$ is, for each $\theta \in \Theta$, a known function of x. $F(x; 0)$ is indicated by $F(x)$ with density $f(x)$. The ordered parameters are denoted by $\theta_{[1]} \leq ... \leq \theta_{[k]}$ and the population associated with $\theta_{[i]}$ is denoted by $\pi_{(i)}, i = 1, ..., k$. The best population is $\pi_{(k)}$. It is assumed that $\pi_{(k)}$ is unique, otherwise appropriate flagging is used. A sufficient statistic X_i for θ_i, e.g. the sample mean, is used for the selection. The random variable associated with $\pi_{(j)}, j = 1, ..., k$, is denoted by $X_{(j)}$. The number of populations in the selected subset, the subset size, is indicated by S. The size S is a random variable and the outcome is denoted by s, where $1 \leq s \leq k$. The selected subset of size s is denoted by $C_s, s = 1, ..., k$.
A correct selection (CS) is the selection of a subset from $\pi_1, ..., \pi_k$, which contains the best population $\pi_{(k)}$. The goal of Gupta's selection procedure is to select a non-empty subset, as

small as possible, such that

$$P_\theta(CS) \geq P^* \quad (k^{-1} < P^* < 1)$$

for all $\theta = (\theta_1, ..., \theta_k)$. P^* is called the confidence level of the selection. The selection rule R is given by

$$R: \text{put } \pi_i \text{ into the subset if and only if} \quad X_i \geq \max_{1\leq j\leq k} X_j - d\ ,$$

where the selection constant $d \geq 0$.
Distributional properties of the subset size S are derived in the next section.

3 Distributional results for S

The subset size S can be considered as a crucial and characteristic quantity for subset selection. A relatively large subset size means, apart from random fluctuations, that the location parameters are close together in comparison with the variation of the populations. The distribution of the subset size S is given in the next theorem.

Theorem 3.1. For $s = 1, ..., k$ and integers $1 \leq i_j \leq k, i_j \neq i_l$ for $l \neq j, j, l = 1, ..., k$, we have

$$P[S = s] = \sum_{i_1 < \ldots < i_s} \sum_{r=1}^{s} \int_{-\infty}^{\infty} \prod_{j=s+1}^{k} F(x - \theta_{[i_j]} + \theta_{[i_r]} - d)$$

$$\prod_{\substack{j=1 \\ j\neq r}}^{s} \{F(x - \theta_{[i_j]} + \theta_{[i_r]}) - F(x - \theta_{[i_j]} + \theta_{[i_r]} - d)\} dF(x)\ . \tag{1}$$

Proof.

$$\begin{aligned} P[S = s] &= \sum_{i_1 < \ldots < i_s} P[\pi_{(i_j)} \in C_s, j = 1, ..., s\ ; \\ &\qquad \pi_{(i_l)} \notin C_s, l = s+1, ..., k; i_{s+1} < \ldots < i_k] \\ &= \sum_{i_1 < \ldots < i_s} P[X_{(i_1)}, ..., X_{(i_s)} \geq \max_{1\leq i\leq k} X_{(i)} - d\ ; \\ &\qquad X_{(i_j)} < \max_{1\leq i\leq k} X_{(i)} - d, j = s+1, ..., k; i_{s+1} < \ldots < i_k] \\ &= \sum_{i_1 < \ldots < i_s} \sum_{r=1}^{k} P[X_{(i_1)}, ..., X_{(i_s)} \geq X_{(i_r)} - d; X_{(i_r)} = \max_{1\leq i\leq k} X_{(i)}\ ; \\ &\qquad X_{(i_j)} < X_{(i_r)} - d, j = s+1, ..., k; i_{s+1} < \ldots < i_k] \end{aligned}$$

$$= \sum_{i_1<\ldots<i_s} \sum_{r=1}^{s} P[X_{(i_1)},\ldots,X_{(i_s)} \geq X_{(i_r)} - d; X_{(i_1)},\ldots,X_{(i_s)} \leq X_{(i_r)} ;$$
$$X_{(i_j)} < X_{(i_r)} - d, j = s+1,\ldots,k; i_{s+1} < \ldots < i_k]$$

$$= \sum_{i_1<\ldots<i_s} \sum_{r=1}^{s} P[X_{(i_r)} - d \leq X_{(i_1)},\ldots,X_{(i_s)} \leq X_{(i_r)} ;$$
$$X_{(i_j)} < X_{(i_r)} - d, j = s+1,\ldots,k; i_{s+1} < \ldots < i_k]$$

and the result follows. □

In the next theorem the distribution of S in a specific subspace of the parameter space $\Omega = \{\theta = (\theta_1,\ldots,\theta_k), \theta_i \in \Theta, i = 1,\ldots,k\}$ will be given. This subspace, indicated by $\Omega(\delta)$, consists of the following configurations of θ's:

$$\theta_{[1]} = \theta_{[k-1]} = \theta_{[k]} - \delta \quad (\delta > 0) .$$

Theorem 3.2. For $s = 1,\ldots,k$ the following holds:

$$P[\ S = s|\theta \in \Omega(\delta)] =$$

$$s\binom{k-1}{s}\int_{-\infty}^{\infty} F(x-\delta-d)F^{k-s-1}(x-d)\{F(x)-F(x-d)\}^{s-1}dF(x)+$$

$$+\ (s-1)\binom{k-1}{s-1}\int_{-\infty}^{\infty} F^{k-s}(x-d)\{F(x)-F(x-d)\}^{s-2}\{F(x-\delta)-F(x-\delta-d)\}dF(x)+$$

$$+\binom{k-1}{s-1}\int_{-\infty}^{\infty} F^{k-s}(x+\delta-d)\{F(x+\delta)-F(x+\delta-d)\}^{s-1}dF(x) . \qquad (2)$$

Proof.

$$P[S = s|\theta_{[1]} = \theta_{[k-1]} = \theta_{[k]} - \delta] =$$

$$\sum_{i_1<\ldots<i_s\leq k-1} s\int_{-\infty}^{\infty} F(x-\delta-d)F^{k-s-1}(x-d)\{F(x)-F(x-d)\}^{s-1}dF(x)+$$

$$+\sum_{i_1<\ldots<i_s=k} (s-1)\int_{-\infty}^{\infty} F^{k-s}(x-d)\{F(x)-F(x-d)\}^{s-2}$$

$$\{F(x-\delta)-F(x-\delta-d)\}dF(x)+$$

$$+\sum_{i_1<\ldots<i_s=k}\int_{-\infty}^{\infty} F^{k-s}(x+\delta-d)\{F(x+\delta)-F(x+\delta-d)\}^{s-1}dF(x) ,$$

and the result follows. □

A special case of the subspace $\Omega(\delta)$ is $\Omega(0)$, the so called Least Favourable Configuration (LFC) for subset selection in terms of location parameters. This case is considered in Corollary 3.1.

Corollary 3.1. For $s = 1, ..., k$ one has

$$P[S = s|\theta \in \Omega(0)] = P[S = s|\theta \in LFC]$$

$$= s \binom{k}{s} \int_{-\infty}^{\infty} \{F(x) - F(x-d)\}^{s-1} F^{k-s}(x-d) dF(x) . \tag{3}$$

Proof. The result follows directly from Theorem 3.2 using the identity $\binom{k-1}{s} + \binom{k-1}{s-1} = \binom{k}{s}$. □

It can be verified that $\sum_{s=1}^{k} P[S = s|\theta \in LFC] = 1$ by expanding $\{F(x) - F(x-d)\}^{s-1}$ in (3), changing the summation order, and noticing that the coefficient of $F^i(x)F^{k-1-i}(x-d)$ for $i = 0, 1, ..., k-2$ is equal to

$$\sum_{s=i+1}^{k} s \binom{k}{s} \binom{s-1}{i} (-1)^{s-1-i} = k \binom{k-1}{i} \sum_{s=i+1}^{k} \binom{k-i-1}{s-i-1} (-1)^{s-i-1} = 0 ,$$

while for $i = k-1$, and thus $s = k$, this coefficient is equal to k, and the result follows immediately.

Some special distributions are considered in the next corollaries:

Corollary 3.2. For uniform populations on $(0, \frac{1}{\lambda})$ with scale parameter $\lambda > 0$, and $F(x) = \lambda x, 0 \leq x \leq \frac{1}{\lambda}$, one gets:

$$P[S = s|\theta \in LFC] = \binom{k}{s-1} (\lambda d)^{s-1}(1-\lambda d)^{k-s+1} + (\lambda d)^k I(s = k) , \tag{4}$$

where $s = 1, ..., k$ and $I(A)$ is the indicator function of the set A.

Proof. For $s < k$ one gets from (3)

$$P[S = s|\theta \in LFC] = s \binom{k}{s} \lambda^k \int_{d}^{\frac{1}{\lambda}} d^{s-1}(x-d)^{k-s} dx$$

$$= \binom{k}{s-1} (\lambda d)^{s-1}(1-\lambda d)^{k-s+1}$$

and for $s = k$ the probability in (3) equals

$$k\lambda^k \left\{ \int_d^{\frac{1}{\lambda}} d^{k-1} dx + \int_0^d x^{k-1} dx \right\} ,$$

and the result follows. □

Corollary 3.3. For exponential populations with $F(x) = 1 - e^{-\lambda x}, x \geq 0$, with scale parameter $\lambda > 0$, and 0 elsewhere, one gets

$$P[S = s|\theta \in LFC] = (1 - e^{-\lambda d})^{s-1} e^{-\lambda d I(s<k)} \text{ for } s = 1, ..., k \ . \tag{5}$$

Proof. For $s < k$ one gets

$$P[S = s|\theta \in LFC] = s \binom{k}{s} \lambda \int_d^\infty \{e^{-\lambda(x-d)} - e^{-\lambda x}\}^{s-1} (1 - e^{-\lambda(x-d)})^{k-s} e^{-\lambda x} dx$$

$$= s \binom{k}{s} (1 - e^{-\lambda d})^{s-1} e^{-\lambda d} \int_0^1 y^{s-1} (1-y)^{k-s} dy$$

$$= e^{-\lambda d} (1 - e^{-\lambda d})^{s-1}$$

and for $s = k$ this probability equals

$$e^{-\lambda d} (1 - e^{-\lambda d})^{k-1} + k\lambda \int_0^d (1 - e^{-\lambda x})^{k-1} e^{-\lambda x} dx$$

$$= (1 - e^{-\lambda d})^{k-1} ,$$

and the result follows. □

Corollary 3.4. For logistic populations with $F(x) = (1 + e^{-\lambda x})^{-1}, -\infty < x < \infty$, with scale parameter $\lambda > 0$, one gets, with $a = e^{\lambda d}$ and $s = 1, ..., k$:

$$P[S = s|\theta \in LFC] = s \binom{k}{s} (a-1)^{s-1} \sum_{r=0}^{s-1} \binom{s-1}{r}$$

$$\sum_{i=0}^{r} (-1)^{r+i} \frac{\binom{k+r-1}{k-2}}{\binom{k+r-2}{k+i-2}} a^i \left\{ \frac{1}{k+r-1} - aI(i=r) C_{k+r-1}\left(\frac{1}{a}\right) \right\} , \tag{6}$$

where $C_r(c) = \left(\frac{c}{c-1}\right)^{r+1} \left[\ln c - \sum_{i=1}^{r} \frac{1}{i} \left(1 - \frac{1}{c}\right)^i \right], c > 0$ and integer $r \geq 1$.

Proof. See appendix 1. □

4 The expectation and variance of S under the LFC

It is possible to determine the expected subset size S under the LFC, denoted by $\mathcal{E}\{S|LFC\}$, using (3) (see Appendix 2). The ultimate result

$$\mathcal{E}\{S|LFC\} = k \int_{-\infty}^{\infty} F^{k-1}(x+d)dF(x) \ , \tag{7}$$

derived in a different way by Gupta (1965), is well known. This result holds for a general value of the selection constant $d \geq 0$. To fulfill the confidence requirement $P(CS) \geq P^*$, then d has to be solved from

$$\int_{-\infty}^{\infty} F^{k-1}(x+d)dF(x) = P^* \ .$$

For this value of d one has $\mathcal{E}\{S|LFC\} = kP^*$ (cf. Gupta (1965)). Expressions for the expected value of S will be derived for some well known distributions.

Corollary 4.1. For *uniform populations* with $F(x) = \lambda x, 0 \leq x \leq \frac{1}{\lambda}$, with scale parameter $\lambda > 0$, one gets for $0 \leq d < \frac{1}{\lambda}$:

$$\mathcal{E}\{S|LFC\} = 1 - \lambda^k d^k + k\lambda d \ . \tag{8}$$

Proof.

$$\mathcal{E}\{S|LFC\} = k\lambda \int_{0}^{\frac{1}{\lambda}-d} \{\lambda(x+d)\}^{k-1}dx + k\lambda \int_{\frac{1}{\lambda}-d}^{\frac{1}{\lambda}} dx = 1 - \lambda^k d^k + k\lambda d \ .$$

□

Corollary 4.2. For *exponential populations* with $F(x) = 1 - e^{-\lambda x}, 0 \leq x < \infty$ and with scale parameter $\lambda > 0$, one gets for $0 \leq d$:

$$\mathcal{E}\{S|LFC\} = e^{\lambda d}\{1 - (1 - e^{-\lambda d})^k\} \ . \tag{9}$$

Proof.

$$\mathcal{E}\{S|LFC\} = k\lambda \int_{0}^{\infty} \{1 - e^{-\lambda(x+d)}\}^{k-1} e^{-\lambda x} dx$$

$$= e^{\lambda d}\{1 - (1 - e^{-\lambda d})^k\} \ .$$

□

Corollary 4.3. For *logistic populations* with $F(x) = (1 + e^{-\lambda x})^{-1}, -\infty < x < \infty$ with scale parameter $\lambda > 0$, one gets for $0 \leq d$ and $a = e^{\lambda d}$:

$$\mathcal{E}\{S|LFC\} = k\left\{1 - \frac{k-1}{a}C_{k-1}(a)\right\} . \tag{10}$$

Proof.

$$\mathcal{E}\{S|LFC\} = k \int_{-\infty}^{\infty} (1 + e^{-\lambda(x+d)})^{-k+1}\lambda e^{-\lambda x}(1 + e^{-\lambda x})^{-2}dx$$

$$= ka^{k-1}\int_{0}^{\infty}(z + a)^{-k+1}(z + 1)^{-2}dz$$

$$= k\left\{1 - \frac{k-1}{a}C_{k-1}(a)\right\} ,$$

where the last equality follows from Theorem 1 in van der Laan (1992c). □

For a general value of d the variance of S under the LFC, denoted by $\text{var}\{S|LFC\}$, is given in the next theorem.

Theorem 4.1.

$$\text{var}\{S|LFC\} = k \int_{-\infty}^{\infty} F^{k-1}(x + d)dF(x)\{2k - 1 - \int_{-\infty}^{\infty} F^{k-1}(x + d)dF(x)\}$$

$$-2k(k-1)\int_{-\infty}^{\infty} F^{k-2}(x + d)F(x)dF(x) . \tag{11}$$

Proof. See Appendix 3.

For a general value of d this variance has been determined for uniform, exponential and logistic distributions. The results are given in the next corollaries.

Corollary 4.4. For *uniform populations* with distribution function $F(x) = \lambda x, 0 \leq x \leq \frac{1}{\lambda}$, with scale parameter $\lambda > 0$ one, gets for $0 \leq d < \frac{1}{\lambda}$:

$$\text{var}\{S|LFC\} = k\lambda d(1 - \lambda d)(1 - 2\lambda^{k-1}d^{k-1}) + \lambda^k d^k(1 - \lambda^k d^k) . \tag{12}$$

Proof.

$$\text{var}\{S|LFC\} = (2k-1)(1 - \lambda^k d^k + k\lambda d) -$$

$$-2k(k-1)\left\{\lambda^k \int_{0}^{\frac{1}{\lambda}-d} (x + d)^{k-2}x dx + \lambda^2 \int_{\frac{1}{\lambda}-d}^{\frac{1}{\lambda}} x dx\right\} - (1 - \lambda^k d^k + k\lambda d)^2$$

$$= k\lambda d(1 - \lambda d)(1 - 2\lambda^{k-1}d^{k-1}) + \lambda^k d^k(1 - \lambda^k d^k) .$$ □

Corollary 4.5. For *exponential populations* with distribution function $F(x) = 1 - e^{-\lambda x}, 0 \leq x < \infty$, and with scale parameter $\lambda > 0$, one has

$$\mathrm{var}\{S|LFC\} = e^{\lambda d}(e^{\lambda d} - 1)\{1 - (2k-1)e^{-\lambda d}(1 - e^{-\lambda d})^{k-1} - (1 - e^{-\lambda d})^{2k-1}\} . \quad (13)$$

Proof.

$$\mathrm{var}\{S|LFC\} = (2k-1)e^{\lambda d}\{1 - (1 - e^{-\lambda d})^k\} - e^{2\lambda d}\{1 - (1 - e^{-\lambda d})^k\}^2 -$$

$$-2k(k-1)\int_0^\infty (1 - e^{-\lambda(x+d)})^{k-2}(1 - e^{-\lambda x})\lambda e^{-\lambda x} dx$$

$$= (2k-1)e^{\lambda d}\{1 - (1 - e^{-\lambda d})^k\} - e^{2\lambda d}\{1 - (1 - e^{-\lambda d})^k\}^2 -$$

$$-2k(k-1)\int_0^1 (1 - e^{-\lambda d}t)^{k-2}(1 - t)dt ,$$

and the result follows. □

Corollary 4.6. For *logistic population* with distribution function $F(x) = (1+e^{-\lambda x})^{-1}, -\infty < x < \infty$, with scale parameter $\lambda > 0$, and density $f(x)$ one has

$$\mathrm{var}\{S|LFC\} = \frac{k(k-1)}{a}\left\{(k-2)\left(1 - \frac{k-1}{a}C_{k-1}(a)\right) + C_{k-1}(a)\left(1 - \frac{k(k-1)}{a}C_{k-1}(a)\right)\right\} . \quad (14)$$

Proof.

$$\mathrm{var}\{S|LFC\} = k(2k-1)\left\{1 - \frac{k-1}{a}C_{k-1}(a)\right\} - k^2\left\{1 - \frac{k-1}{a}C_{k-1}(a)\right\}^2 -$$

$$-2k(k-1)\int_{-\infty}^\infty F^{k-2}(x)F(x-d)f(x-d)dx ,$$

where the last integral equals (with $a = e^{\lambda d}$):

$$\int_{-\infty}^\infty \frac{\lambda e^{-\lambda x}}{\{1 + e^{-\lambda(x+d)}\}^{k-2}(1 + e^{-\lambda x})^3} dx$$

$$= a^{k-2}\int_0^\infty (a+y)^{-k+2}(1+y)^{-3} dy$$

$$= \tfrac{1}{2} - \frac{k-2}{2}\frac{1}{a}\left\{1 - \frac{k-1}{a}C_{k-1}(a)\right\} ,$$

using partial integration. After some calculations the result follows. □

In the tables 4.1 and 4.2 some numerical results concerning expectation and standard deviation of the subset size S are given for normal populations.

Table 4.1. $\mathcal{E}\{S|LFC\}$ for standard normal populations for some values of k and d

k \ d	1	1.5	2	2.5	3	3.5	4	4.5	5
2	1.52	1.71	1.84	1.92	1.97	1.99	2.00	2.00	2.00
5	2.47	3.26	3.94	4.43	4.73	4.89	4.96	4.99	5.00
10	3.41	5.09	6.74	8.09	9.02	9.57	9.83	9.94	9.98
25	4.99	8.68	13.05	17.28	20.66	22.89	24.12	24.68	24.90

Table 4.2. Standard deviation of S under the LFC for standard normal distributions for some values of k and d

k \ d	1	1.5	2	2.5	3	3.5	4	4.5	5
2	.500	.453	.364	.267	.181	.115	.068	.038	.020
5	1.14	1.20	1.08	.848	.595	.381	.225	.125	.066
10	1.78	2.14	2.12	1.80	1.32	.863	.511	.281	.145
25	2.91	4.13	4.73	4.50	3.62	2.51	1.52	.083	.042

5 Efficiency of subset selection of an ε-best population relative to selecting the best one

In this section, for fixed $\varepsilon(\geq 0)$, the efficiency of subset selection for an ε-best population relative to subset selection for the best population is considered. A correct selection (CS) in the context of selection of an ε-best population is the selection of a subset which contains at least one ε-best population, where an ε-best population is defined as a population π_i for which $\theta_i \geq \theta_{[k]} - \varepsilon$. Note that there exists at least one ε-best population for each $\varepsilon \geq 0$. As for Gupta's procedure, the goal is to select a subset, as small as possible, such that a $P(CS) \geq P^*$ for all θ, where P^* is given and $k^{-1} < P^* < 1$. The selection rule is of the same form as Gupta's; the difference is in the choice of the selection constant d. Further, for normal populations with known standard deviation $\sigma(> 0)$, then it can easily be seen that the selection constant d for selecting an ε-best population is equal to $\sigma d_G - \varepsilon$, where d_G is Gupta's selection constant for standard normal populations and the same P^* is used for the two procedures. It can eaily be proved that, when using Gupta's procedure, the probability of selecting an ε-best population into the subset is at least equal to the probability of selecting the best population and this for every θ.

Definition 5.1. For a fixed selection constant $d(\geq 0)$ and a fixed $\varepsilon(\geq 0)$ the efficiency, G, of the ε-best selection procedure is defined as the relative difference in the minimal probabilities of reaching the selection goals – the new one, resp. Gupta's – relative to Gupta's.

Theorem 5.1. For fixed $\varepsilon \geq 0$ the relative efficiency $G = G(d)$ can be written as

$$G(d) = \frac{\int_{-\infty}^{\infty} F^{k-1}(x+d+\varepsilon)dF(x)}{\int_{-\infty}^{\infty} F^{k-1}(x+d)dF(x)} - 1 \;, \tag{15}$$

where d is the selection constant for subset selection of an ε-best population meeting the P^*-requirement. Further:

$$\lim_{d\uparrow\infty} G(d) = 0 \;, \tag{16}$$

with $d = d(P^*)$, an increasing function of P^*.

Proof. Straightforward. □

If the density $f(x)$ of $F(x)$ is strongly unimodel ($f(x)$ is log-concave; Dharmadhikari and Joag-dev (1988)) then the following theorem holds for the efficiency G.

Theorem 5.2. If $f(x)$ is strongly unimodal, then $G(d)$ is a decreasing function of P^*.

Proof. $H(d) := \int_{-\infty}^{\infty} F^{k-1}(x+d)f(x)dx$ can be written as

$$H(d) = P(\max(X_2, ..., X_k) - X_1 \leq d)$$

$$= F^{k-1} * F(d) \;.$$

F and hence (van der Laan (1970)) F^{k-1} are strongly unimodal. Since strong unimodality is preserved under convolution (Dharmadhikari and Joag-dev (1988)), $H(d)$ has a log-concave density. It follows (Karlin (1968)) that $H(d)$ itself is log-concave. From this it follows that log $H(d+\varepsilon) - H(d)$ is a decreasing function of d, so $G(d)$ is a decreasing function of d. □

Some results of the efficiency G is given in the tables 5.1 and 5.2 for normal populations with common known scale paramter σ, which can be assumed to be 1 without loss of generality. In table 5.2 it can be seen, as an illustration of Theorem 5.2, that, for fixed k and ε, G is a decreasing function of P^*.

Table 5.1. The relative efficiency G for different values of k and ε, and with $P^* = .90$

$\varepsilon = 0.2$

k	min. Pr. for best	G in %
2	0.868	3.7
3	0.865	4.0
4	0.864	4.2
5	0.863	4.3
6	0.863	4.3
7	0.862	4.4
8	0.862	4.4
9	0.862	4.4
10	0.862	4.4
25	0.860	4.7
50	0.859	4.8
100	0.859	4.8
500	0.858	4.9
1000	0.858	4.9
2000	0.857	5.0

$\varepsilon = 0.5$

k	min. Pr. for best	G in %
2	0.820	9.8
3	0.813	10.7
4	0.810	11.1
5	0.809	11.2
6	0.807	11.5
7	0.806	11.7
8	0.805	11.8
9	0.805	11.8
10	0.804	11.9
25	0.801	12.4
50	0.799	12.6
100	0.797	12.9
500	0.795	13.2
1000	0.794	13.4
2000	0.793	13.5

Table 5.2. The relative efficiency G for different values of P^* and ε, and with $k = 10$

$k = 10, \varepsilon = 0.2$

P^*	0.80	0.90	0.95	0.975	0.99	0.995	0.999
min P	0.739	0.862	0.927	0.962	0.983	0.9917	0.9977
G in %	8.3	4.4	2.5	1.4	0.7	0.33	0.13

$k = 10, \varepsilon = 0.5$

P^*	0.80	0.90	0.95	0.975	0.99	0.995	0.999
min P	0.648	0.804	0.893	0.942	0.973	0.9868	0.9959
G in %	23.5	11.9	6.4	3.5	1.7	0.83	0.31

Another way to compare the ε-best procedure with Gupta's is to investigate the ratio R defined by

$$R = \frac{\sup \mathcal{E}_\theta S_b}{\sup \mathcal{E}_\theta S_a}, \tag{17}$$

where S_b and S_a are the subset sizes for selecting the best and an ε-best population resp. when using Gupta's procedure and the ε-best selection procedure, respectively, with the same minimal probability of correct selection. For normal populations with $\sigma = 1$ some results are presented in table 5.3. In this case it can easily be proved that R can be written as

$$R = kP^* \left\{ \int_{-\infty}^{\infty} \Phi^{k-1}(x+d+\varepsilon)d\Phi(x) + (k-1)\int_{-\infty}^{\infty} \Phi^{k-2}(x+d)\Phi(x+d-\varepsilon)d\Phi(x) \right\}^{-1},$$

with $\Phi(\cdot)$ the standard normal cumulative distribution function (Gupta (1965)), and can be computed using numerical integration.

Table 5.3. Values of R for different values of k, ε and P^*

k	ε	P^* 0.50	0.80	0.90	0.95	0.99	k	ε	P^* 0.50	0.80	0.90	0.95	0.99
2	0.5	-	1.018	1.016	1.011	1.004	10	0.5	1.016	1.011	1.007	1.004	1.001
	1.0	-	1.072	1.063	1.046	1.017		1.0	1.077	1.051	1.033	1.021	1.007
	2.0	-	1.258	1.247	1.195	1.088		2.0	1.442	1.292	1.196	1.129	1.047
3	0.5	1.015	1.019	1.014	1.010	1.003	20	0.5	1.011	1.007	1.004	1.003	1.001
	1.0	1.063	1.082	1.061	1.042	1.015		1.0	1.056	1.033	1.021	1.013	1.004
	2.0	1.235	1.357	1.282	1.202	1.083		2.0	1.358	1.211	1.138	1.090	1.032
4	0.5	1.019	1.018	1.013	1.008	1.003	50	0.5	1.006	1.003	1.002	1.001	1.000
	1.0	1.082	1.078	1.055	1.037	1.012		1.0	1.032	1.018	1.011	1.007	1.002
	2.0	1.355	1.372	1.274	1.190	1.074		2.0	1.234	1.130	1.084	1.054	1.019
5	0.5	1.019	1.016	1.011	1.007	1.002	100	0.5	1.003	1.002	1.001	1.001	1.000
	1.0	1.087	1.072	1.050	1.032	1.011		1.0	1.020	1.011	1.007	1.004	1.001
	2.0	1.414	1.364	1.259	1.176	1.067		2.0	1.163	1.089	1.057	1.036	1.012
6	0.5	1.019	1.015	1.010	1.006	1.002	500	0.5	1.001	1.001	1.000	1.000	1.000
	1.0	1.088	1.067	1.045	1.029	1.009		1.0	1.006	1.003	1.002	1.001	1.000
	2.0	1.441	1.350	1.243	1.164	1.062		2.0	1.067	1.036	1.022	1.014	1.005
7	0.5	1.018	1.013	1.009	1.006	1.002	1000	0.5	1.001	1.001	1.000	1.000	1.000
	1.0	1.086	1.062	1.041	1.026	1.008		1.0	1.004	1.002	1.001	1.001	1.000
	2.0	1.452	1.334	1.229	1.153	1.057		2.0	1.045	1.024	1.015	1.009	1.003
8	0.5	1.017	1.012	1.008	1.005	1.002	1500	0.5	1.001	1.001	1.001	1.000	1.000
	1.0	1.083	1.058	1.038	1.024	1.008		1.0	1.003	1.002	1.001	1.001	1.001
	2.0	1.453	1.319	1.217	1.144	1.053		2.0	1.036	1.019	1.012	1.008	1.003
9	0.5	1.017	1.012	1.008	1.005	1.001	2000	0.5	1.001	1.001	1.001	1.001	1.001
	1.0	1.080	1.054	1.036	1.022	1.007		1.0	1.003	1.002	1.001	1.001	1.001
	2.0	1.449	1.305	1.205	1.136	1.050		2.0	1.031	1.016	1.010	1.007	1.002

6 A generalized subset selection procedure: Some recent results

Instead of asking for a probability of at least P^* of a correct selection, van der Laan and van Eeden (1993) use a loss function. They take the loss equal to zero when the subset contains an ε-best population and a nondecreasing function, h, of the difference $\theta_{[k]} - \varepsilon - \theta^{[s]}$ if the selected subset does not contain an ε-best population, where $\theta^{[s]} = \max(\theta_i | i$ such that π_i is in the subset). The subset selection goals are expressed in terms of an upper bound on the risk function and/or on the expected subset size. These upper bounds are required to hold either for all or for some θ. The selection rule is of the same form as the one used by Gupta (1965) and by van der Laan (1992a). The difference is in the value of the selection constant d. Gupta's and van der Laan's subset selection approaches are obtained as special cases by taking $h(x) \equiv 1$ for all $x \in \mathbb{R}^+$ with $\varepsilon = 0$ for Gupta and $\varepsilon > 0$ for van der Laan. The case of two normal populations with equal known variances and

$h(x) = x^p$ for some $p > 1$ has been studied in full detail. Properties of the expected subset size and the risk function are given. In order to apply in practice this new subset selection methodology, tables of the risk function and the expected subset size are needed. Such tables can be found in van der Laan and van Eeden (1993). If the introduction of a loss function of the form given in this paper is realistic for a practical problem, then the methodology based on a loss function can be considered as a flexible approach. The method is a generalization of Gupta's subset selection procedure and is designed to give a smaller expected subset size. A comparison with Gupta's subset selection procedure can be found in van der Laan and van Eeden (1993).

Appendix 1

Proof of Corollary 3.4.

$$P[S = s|\theta \in LFC]$$

$$= s\binom{k}{s}\lambda\int_{-\infty}^{\infty}\left\{\frac{1}{1+e^{-\lambda x}} - \frac{1}{1+e^{-\lambda(x-d)}}\right\}^{s-1}\left(\frac{1}{1+e^{-\lambda(x-d)}}\right)^{k-s}\frac{e^{-\lambda x}}{(1+e^{-\lambda x})^2}dx$$

$$= s\binom{k}{s}(a-1)^{s-1}a^{-k+1}\int_0^{\infty}\frac{z^{s-1}}{(z+1)^{s+1}(z+\frac{1}{a})^{k-1}}dz ,$$

with $a = e^{\lambda d}$ and using the transformation $z = e^{-\lambda x}$, gives

$$= s\binom{k}{s}(a-1)^{s-1}a^{-k+1}\sum_{i=0}^{s-1}\binom{s-1}{i}(-1)^i J(i+2, k-1) ,$$

using the Binomial expansion for $\{(z+1)-1\}^{s-1}$, where $J(m,n) = \int_0^{\infty}(z+1)^{-m}(z+\frac{1}{a})^{-n}dz$ with $2 \leq m \leq n+2$ and $1 \leq n < \infty$. Then

$$J(2,n) = a^n\left\{1 - anC_n(\tfrac{1}{a})\right\}$$

(van der Laan (1992c); Theorem 1), and after partial integration we get the following recurrence relation

$$J(m,n) = \frac{a^n}{m-1} - \frac{n}{m-1}J(m-1,n+1) .$$

Thus

$$J(m,n) = \sum_{i=1}^{m-1}(-1)^{i-1}\frac{\binom{n+m-2}{n-1}}{\binom{n+m-3}{n+i-2}}a^{n+i-1}\left\{\frac{1}{n+m-2} - aI(i = m-1)C_{n+m-2}(\tfrac{1}{a})\right\}$$

and

$$P[S=s|\theta \in LFC] = s \binom{k}{s} (a-1)^{s-1} a^{-k+1} \sum_{r=0}^{s-1} \binom{s-1}{r} (-1)^r$$

$$\sum_{i=1}^{r+1} (-1)^{i-1} \frac{\binom{k+r-1}{k-2}}{\binom{k+r-2}{k+i-3}} a^{k+i-2} \left\{ \frac{1}{k+r-1} - aI(i=r+1) C_{k+r-1}(\tfrac{1}{a}) \right\}$$

and the result follows.

Appendix 2

$$\mathcal{E}\{S|LFC\} = \sum_{s=1}^{k} s^2 \binom{k}{s} \int_{-\infty}^{\infty} \{F(x) - F(x-d)\}^{s-1} F^{k-s}(x-d) dF(x) .$$

From

$$s^2 \binom{k}{s} = k(k-1) \binom{k-2}{s-2} + k \binom{k-1}{s-1} ,$$

with $1 \le s \le k$, $\binom{0}{0} = 1$, $\binom{a}{b} = 0$ for all real a and integer $b < 0$, and, for $j = 1, 2$,

$$\sum_{s=j}^{k} \binom{k-j}{s-j} \{F(x) - F(x-d)\}^{s-j} F^{k-s}(x-d) = F^{k-j}(x) \tag{18}$$

it follows that

$$\mathcal{E}\{S|LFC\} = k(k-1) \int_{-\infty}^{\infty} \{F(x) - F(x-d)\} F^{k-2}(x) dF(x) +$$

$$+k \int_{-\infty}^{\infty} F^{k-1}(x) dF(x)$$

$$= k \int_{-\infty}^{\infty} F^{k-1}(y+d) dF(y) ,$$

using partial integration and the transformation $y = x - d$.

Appendix 3

Proof of Theorem 4.1.
We have

$$\mathcal{E}\{S^2|LFC\} = \sum_{s=1}^{k} s^3 \binom{k}{s} \int_{-\infty}^{\infty} F^{k-s}(x-d)\{F(x)-F(x-d)\}^{s-1} dF(x) \,.$$

The next identity can easily be verified

$$s^3 \binom{k}{s} = k(k-1)(k-2)\binom{k-3}{s-3} + 3k(k-1)\binom{k-2}{s-2} + k\binom{k-1}{s-1} \,,$$

with $1 \leq s \leq k$. Equality (18) is als valid for $j = 3 (k \geq 3)$. Then it follows that

$$\mathcal{E}\{S^2|LFC\} = k(k-1)(k-2) \int_{-\infty}^{\infty} F^{k-3}(x)\{F(x)-F(x-d)\}^2 dF(x) +$$

$$+3k(k-1) \int_{-\infty}^{\infty} F^{k-2}(x)\{F(x)-F(x-d)\} dF(x) + k \int_{-\infty}^{\infty} F^{k-1}(x) dF(x)$$

$$= k(2k-1) \int_{-\infty}^{\infty} F^{k-1}(x+d) dF(x) - 2k(k-1) \int_{-\infty}^{\infty} F^{k-2}(x+d) F(x) dF(x) \,,$$

using partial integration, and the result follows immediately.

References

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.

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

Dharmadhikari, S. and K. Joag-dev (1988). Unimodality, Convexity, and Applications. Academic Press, Inc., New York-London.

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

Karlin, S. (1968). Total Positivity. Stanford University Press, Vol. 1.

van der Laan, P. (1970). Simple distribution-free confidence intervals for a difference in location. Eindhoven University of Technology.

van der Laan, P. (1992a). Subset selection of an almost best treatment. Biometrical Journal 34, 647-656.

van der Laan, P. (1992b). Subset selection: Robustness and Imprecise Selection. Memorandum COSOR 92-10. Dept. of Mathematics and Computing Science, Eindhoven University of Technology.

van der Laan, P. (1992c). On subset selection from logistic populations. Statistica Neerlandica 46, 153-163.

van der Laan, P. and C. van Eeden (1993). Subset selection with a generalized selection goal based on a loss function. Technical Report #127, Dept. of Statistics, University of British Columbia, Vancouver and Memorandum COSOR 93-15, Dept. of Mathematics and Computing Science, Eindhoven University of Technology. Submitted.