ELSEVIER

European Journal of Operational Research 96 (1996) 407–416

EUROPEAN JOURNAL OF OPERATIONAL RESEARCH

Theory and Methodology

A method for selecting a subset of alternatives for future decision making

Chaim Meyer Ehrman [a], Morris Hamburg [b], Abba M. Krieger [b,*]

[a] Loyola University of Chicago, Chicago, USA

[b] Wharton School of the University of Pennsylvania, Department of Statistics and Operations Research, Pennsylvania, PA 19104-6302, USA

Received 30 August 1995; accepted 23 February 1996

Abstract

In this paper we present an algorithm for finding a subset from a large number of alternatives. The criterion for selecting this subset is based on the assumption that ultimately one alternative will be chosen and implemented from this subset. Some areas of application for the subset selection techniques are presented. Extensions of this research are suggested.

Keywords: Subset selection; Decision making

1. Introduction

There is a class of problems in which several items from many are selected for information search, and the ultimate objective is to choose, after the procurement of information, the best item for some management objective. Current information concerning the potential performances of each item is imprecise, and searching for information with higher precision is costly. Therefore, it is infeasible to collect information for each item. The decision maker may want to select a small group of items or a "subset" for which to obtain information with greater precision. The final selection is assumed to be restricted to those items with more precise information, i.e., to items in the subset. The problem we are addressing is how to select this subset. Initially we present some applications of this type of problem.

In *foreign direct investment*, assume that a manufacturer with large export sales is interested in investing in a manufacturing facility abroad. The motivation for establishing such a facility can be lower labor costs, lower costs of raw materials, greater proximity to foreign markets, lower transportation costs, increased identification of marketing opportunities with the foreign market, etc. Suppose there are over 150 countries with some potential for investment. The investor wants to conduct an information search on a subset of countries before committing his firm to a foreign direct investment decision. It is important that the countries that provide good opportunities for the manufacturing facility be included in the subset. A country that has not been included in

* Corresponding author.

PII S0377-2217(96)00086-0

the subset and has not been investigated will not be considered a viable location for foreign direct investment. The investor has the decision problem of determining which countries to select for additional information search. Subset selection is widely referenced in the literature of foreign direct investment (cf. Aharoni (1966), Stobaugh (1969a), Stobaugh (1969b) and Root (1982)).

In *new product development*, product concepts are usually subjected to a series of screening tests with the goal of getting more information on chances for success before the product concept actually enters the market as a finished good. Assume that a brand manager selects a subset of product concepts with the purpose of producing them in limited quantity for test markets. The brand manager wants the concept that receives the highest score within the subset to be a successful one. Subset selection is referenced in the marketing literature of product planning and development (cf. Kotler (1980), Silk and Urban (1978)).

In application areas that use subset selection, the technique used has primarily been to score each alternative and to choose the subset of alternatives corresponding to the highest scores. Somewhat less frequently a variance has been assigned to each alternative as a measure of uncertainty attached to the original scores. The subset that is selected is determined by the alternatives with the smallest variance. Although more sophisticated techniques have been proposed in the statistical literature, (see Bechhofer (1954) for problem definition and Gupta and Panchapakesan (1979) for a review of the literature on the techniques), the focus has been on sample size.

The emphasis in this paper is very different. Our aim is to find a subset of alternatives so that at least one member of the subset will do well, once the uncertainties are resolved. Intuitively, the subset may very well include some items with relatively high scores and low uncertainty and other items with relatively low scores and high uncertainty particularly when the means of the scores of the alternatives are not too disparate. An alternative, A, with relatively low mean score and high variance will often do poorly in which case the best item will come from an alternative with high mean score. However, because of the high uncertainty, the realized value of alternative A could be quite high. When this occurs, there is an opportunity to choose this alternative. Subsets of this form arise from the analysis in this paper.

In Section 2, we formulate the problem mathematically. The algorithm for selecting the subset is presented in Section 3. Properties of this algorithm and a form of sensitivity analysis are also given. The results are illustrated with examples in Section 4. Finally, in Section 5, conclusions are stated and suggestions are given for future research.

2. Mathematical preliminaries

Let the values of the n alternatives be denoted by the random variables X_i; $i = 1, \ldots, n$, where $(X_1, \ldots, X_n)$ are mutually independent and X_i has cumulative distribution function (cdf) $F_i(x)$. For part of the discussion we assume that $X_i \sim N(\mu_i, \sigma_i^2)$. Since we want to choose a subset of size k from among the n alternatives we let S denote a subset of the integers from 1 to n with k elements. Given the subset S of alternatives, the random variable $Y(S) = \max_{i \in S} X_i$. The cdf of $Y(S)$ is denoted by $F_{Y(S)}(y)$.

Example: There are ten alternatives from which three are to be selected. Hence, $n = 10$, $k = 3$ and there are $\binom{10}{3} = 120$ possible solutions to the problem. If $S = \{2, 5, 9\}$, for example, then the objective function $Y(S) = \max(X_2, X_5, X_9)$.

We want to decide which subset, from among the $\binom{n}{k}$ subsets of size k, is "best". In order to decide which subset is "best", we need to be able to compare $Y(S_1)$ and $Y(S_2)$, for two subsets S_1 and S_2. The problem is trivial if $Y(S_1)$ is stochastically larger than $Y(S_2)$ (i.e., $F_{Y(S_1)}(y) \le F_{Y(S_2)}(y)$ for all y). In practice, however, there will rarely be a situation in which a solution S^* gives rise to a payoff $Y(S^*)$ which is stochastically larger than the payoffs for all other subsets.

The general approach is to choose a suitable functional, ψ: $F \to R$ (e.g., mean, median and variance). If we focus on measures of central tendency, then the class of functionals that is often considered is of the form $\int_{t=0}^{1} F_Y^{-1}(t)\, d\mu(t)$ where μ is a measure on [0, 1] and where this implicitly assumes that Y is absolutely continuous (there is an analogous formulation if Y is discrete). This formulation includes the median if μ places probability one at $t = 1/2$, the mean if μ is uniformly distributed on [0, 1] and trimmed mean if μ is uniformly distributed on $[\alpha, 1-\alpha]$ for a given $\alpha \in (0, 1/2)$. Although we focus on the mean, median and other quantiles, it is important to place these parameters in the wider setting as discussed above.

The random variable that is the maximum of other random variables is often skewed to the right; this would be the case if the random variables are normal. The mean of a skewed random variable is highly sensitive to the tail behavior. Hence maximizing the mean might not be appropriate. For example, in many income distributions the mean is a figure such that as many as 75% of the items in the distribution are lower than that figure.

There are of course many criteria that could be used. The criterion that should be used is subjective and should depend on the context. We only note that the more commonly used criteria, means and variances, might not be appropriate when one is interested in the maximum score chosen from a subset. [1]

It is interesting to note that the mean of a random variable can be written as $\int_0^1 F^{-1}(p)\,\mathrm{d}p$. In essence, this says that the mean is the average performance over all quantiles. In this paper, we consider a specific quantile as our criterion. This suggests that an interesting extension would be to use a weighted combination of user supplied quantiles as a criterion (e.g. maximizing the average of the median and the 75th percentile). The results in this paper can be used in a 0-1 integer program framework to deal with this extension.

How we compare $Y(S_1)$ to $Y(S_2)$, whether through the means of the random variables or the medians of the random variables, is somewhat arbitrary and should depend on qualitative considerations. There are clearly cases in which the median of $Y(S_1)$ exceeds the median of $Y(S_2)$, but the mean of $Y(S_1)$ is less than the mean of $Y(S_2)$. In order to illustrate this point, we assume that $X_1 \sim N(\mu_1, \sigma_1^2)$ and $X_2 \sim N(\mu_2, \sigma_2^2)$ independent of X_1. If $Y = \max(X_1, X_2)$, then

$$E(Y) = \sigma_0 \phi\left(\frac{\mu_1 - \mu_2}{\sigma_0}\right) + \mu_2 \Phi\left(\frac{\mu_2 - \mu_1}{\sigma_0}\right) + \mu_1 \Phi\left(\frac{\mu_1 - \mu_2}{\sigma_0}\right) \tag{2.1}$$

where $\sigma_0 = (\sigma_1^2 + \sigma_2^2)^{1/2}$ and $\phi(x)$ and $\Phi(x)$ denote the density and distribution functions of the standard normal distribution. The median of Y is y_0 defined by the equation

$$\Phi\left(\frac{y_0 - \mu_1}{\sigma_1}\right)\Phi\left(\frac{y_0 - \mu_2}{\sigma_2}\right) = \frac{1}{2}. \tag{2.2}$$

Eq. (2.1) is proved in Appendix A and equation Eq. (2.2) follows from the fact that the distribution of the maximum of two independent random variables is the product of the distributions.

Example 1. Let $X_1 \sim N(\mu, 1)$, $X_2 \sim N(\mu, 1)$ and $X_3 \sim N(-\mu^2, \mu^4)$ with $\mu > 0$. Let $Y_1 = \max(X_1, X_2)$ and $Y_2 = \max(X_1, X_3)$. It is easy to verify that the median of Y_1 is $\mu + \Phi^{-1}(\sqrt{2}/2)$ and the median of Y_2 is less than $\mu + \Phi^{-1}(\sqrt{2}/2)$ because

$$P\left(X_3 \le \mu + \Phi^{-1}\left(\frac{\sqrt{2}}{2}\right)\right) > \frac{\sqrt{2}}{2}.$$

But,

$$E(Y_1) = \sqrt{2}\,\phi(0) + \mu = \frac{1}{\sqrt{\pi}} + \mu$$

and

$$\frac{E(Y_2)}{\mu^2} = \frac{\mu_0 \phi((\mu + \mu^2)/\mu_0)}{\mu^2} - \Phi\left(\frac{-\mu^2 - \mu}{\mu_0}\right) + \frac{\Phi((\mu + \mu^2)/\mu_0)}{\mu} \underset{\mu \to \infty}{\to} [\phi(1) - \Phi(-1)] \approx 0.08$$

[1] The normal distribution (and hence means and variances) is used to describe the behavior of the scores of each of the alternatives. When we consider an objective function which is the maximum of normal random variables, then we use a quantile such as the median as the parameter of interest.

where $\mu_0 = (1 + \mu^4)^{1/2}$. The point is that Y_1 is preferred by the median criterion, but for sufficiently large μ, Y_2 is far superior in terms of expected value. In the next example, we find an instance where Y_1 is preferred by the expected value criterion, but for sufficiently large μ, Y_2 is far superior in terms of its median.

Example 2. Let $X_1 = \mu$ with probability one, $X_2 \sim N(-\mu^2, 4\mu^4)$ and $X_3 \sim N(-\mu^2, 4\mu^4)$ with $\mu > 0$. Let $Y_1 = \max(X_1, X_2)$ and $Y_2 = \max(X_2, X_3)$. Then,

$$\frac{E(Y_1)}{\mu^2} = 2\phi\left(\frac{\mu + \mu^2}{2\mu^2}\right) - \Phi\left(\frac{-\mu^2 - \mu}{2\mu^2}\right) + \frac{\Phi((\mu^2 + \mu)/2\mu^2)}{\mu}$$

$$\underset{\mu \to \infty}{\rightarrow} [2\phi(1/2) - \Phi(-1/2)] \approx 0.4$$

and

$$E(Y_2) = \sqrt{8}\,\mu^2\phi(0) - \mu^2 = \mu^2\left[\sqrt{8}\,\phi(0) - 1\right] \approx 0.13\mu^2.$$

But the median of Y_1 is μ and the median of Y_2 is

$$-\mu^2 + 2\mu^2\Phi^{-1}\left(\frac{\sqrt{2}}{2}\right) = \mu^2\left[2\Phi^{-1}\left(\frac{\sqrt{2}}{2}\right) - 1\right] \approx 0.09\mu^2,$$

so Y_1 is preferred to Y_2 in terms of expected value but the median of Y_2 is far greater than the median of Y_1 for large μ.

We showed, by way of examples, that the choice of criterion (i.e., median versus mean) might affect the optimal subset. These examples used random variables with high means and low variances and random variables with low means and high variances. This makes intuitive sense since it is often beneficial to include an alternative with a low mean and high variance as one of the alternatives because the realization of this random variable might be very high; we are guarded against a low realization of this random variable by the other alternatives in the subset. These examples are, of course, carefully constructed. In practice, we might expect that the two criteria lead to the same subset. The following proposition provides a necessary condition for this to be the case.

Proposition. *Assume that X_i are normally distributed. Index the alternatives so that $\mu_1 \geq \mu_2 \geq \cdots \geq \mu_n$. Let $S = \{1,\ldots,k\}$. If $\sigma_1^2 \geq \cdots \geq \sigma_n^2$ then*

(i) *the median of $Y(S) \geq$ the median of $Y(S')$ for all subsets S' of size k and*

(ii) *$E[Y(S)] \geq E[Y(S')]$ for all subsets S' of size k.*

Proof. (i) The proof of (i) is given as Proposition (A.2) in Appendix A.

(ii) The proof of (ii) follows by applying Corollary A.5 as stated in Appendix A. □

3. Mathematical results

We are interested in comparing $Y(S_1) = \max_{i \in S_1} X_i$ to $Y(S_2) = \max_{i \in S_2} X_i$. This will enable us to find ultimately the subset S^* of size k, such that $Y(S^*)$ is preferred to $Y(S)$ for all subsets S of size k. We need to specify what we mean by preferred. Since $Y(S)$ involves a maximum over k random variables, $E(Y(S))$ is not easily obtainable (it involves numerical integration) for most distributional assumptions on X_i including normality. However, the median or any quantile of $Y(S)$ is easily obtainable. We compare two subsets S_1 and S_2 by the pth quantiles of $Y(S_1)$ and $Y(S_2)$.

This section is divided into three parts. We first provide an algorithm for finding S^* such that $F^{-1}_{Y(S^*)}(p) \geq F^{-1}_{Y(S)}(p)$ for all subsets S of size k. Note that when $p = 1/2$ this corresponds to maximizing the median of $Y(S)$. We then present a sensitivity analysis to determine the values of p for which S^* remains optimal. Finally, we illustrate the results from the first two subsections in the third subsection.

3.1. Optimal subset algorithm

In this subsection, we present an algorithm that finds the subset S^* that maximizes $F^{-1}_{Y(S^*)}(p)$ for any p. We then discuss the logic behind the algorithm.

We are given n alternatives X_i, $i = 1, ..., n$ with respective cdf $F_i(x)$; $i = 1, ..., n$. We are also given a percentile of interest, p, and are required to find the best subset of size k. The following algorithm accomplishes this:

3.1.1. Algorithm

Step 0 (*Initialization*): *Let* $C_\ell = \max_i F_i^{-1}(p)$ *and* $C_u = \max_i F_i^{-1}(p^{1/k})$.

Step 1 (*Computation*): *Let* $c = (C_\ell + C_u)/2$, *and* $\gamma_i = F_i(c)$.
Let $i_1, ..., i_n$ *be defined so that* $\gamma_{i_1} \le \cdots \le \gamma_{i_n}$.
Let $S = \{i_1, ..., i_k\}$ *and* $\beta = \prod_{j=1}^{k} \gamma_{i_j}$.

Step 2 a) *If* $|\beta - p| < \epsilon$ *then STOP*; $S^* = S$.
b) *If* $|\beta - p| > \epsilon$ *and*
(*i*) $\beta > p$; $C_u = c$
GO TO STEP 1
(*ii*) $\beta < p$; $C_\ell = c$
GO TO STEP 1

Note that ϵ is chosen to be a very small positive number.

The basis of the algorithm is that the subset S at any iteration is optimal for the corresponding given percentile β (see Step 1). This follows because if we choose any other subset S', then $F_{Y(S')}(c) = \beta' = \prod_{j \in S'} \gamma_j \ge \beta = F_{Y(S)}(c)$. Hence, $F_{Y(S')}^{-1}(\beta) \le c$ because $F_{Y(S')}$ is a nondecreasing function.

The remainder of the algorithm adjusts β at each iteration so that ultimately β is arbitrarily close to p. This is accomplished by finding C_ℓ and C_u at each iteration so that $F_{Y(S^*)}^{-1}(p)$ is necessarily between C_ℓ and C_u and halving $(C_u - C_\ell)$ from one iteration to the next so that $C_u - C_\ell$ is eventually arbitrarily small. We initialize C_ℓ and C_u in Step 0. Clearly, $F_{Y(S^*)}^{-1}(p) \ge F_{X_i}^{-1}(p)$ for all i so that C_ℓ is a lower bound; the best subset of size k that maximizes a percentile must necessarily do at least as well as the best subset of size 1. By the definition of C_u, $F_{X_i}(C_u) \ge p^{1/k}$; $i = 1, ..., n$. Hence, $F_{Y(S)}(C_u) \ge p$ for any subset S of size k. This implies $F_{Y(S)}^{-1}(p) \le C_u$. In Step 2 we either replace C_u or C_ℓ for c. If the optimal subset at c yields a percentile $\beta < p$ ($\beta > p$) then clearly c is too small (large).

3.2. Sensitivity analysis

In this subsection, we first find conditions on F_i such that if a subset is optimal for percentiles p_1 and p_2 it is necessarily optimal for p between p_1 and p_2. This provides the basis for a sensitivity analysis if the condition is satisfied.

The issue is simply if $X \sim F$ and $Y \sim G$ are two random variables such that F and G cross once (say, $F(x) \le G(x)$ for $x \le x_o$ and $G(x) \le F(x)$ for $x_o \le x$), then for $p \le p_o \equiv F^{-1}(x_o) = G^{-1}(x_o)$ random variable X would be preferred to random variable Y in any subset of any size. Similarly, if $p > p_o$, then random variable Y would be preferred. This precludes the possibility of having a subset S that is optimal for p_1 and p_2 which is not optimal for some p between p_1 and p_2. We discuss this issue of 1-crossing in more detail in Appendix B. For our discussion, it is interesting to note that the condition is satisfied if X_i, $i = 1, ..., n$ belong to the same location-scale family such as the normal.

We have a subset S that is found to be optimal for percentile p. We want to find p_ℓ and p_u such that this subset remains optimal for all $p \in [p_\ell, p_u]$. We assume that $X_i \sim F_i$; $i = 1, ..., n$ where F_i and F_j cross once for all $i \ne j$.

We need to compare each alternative not in S with each alternative in S. Since in comparing alternative i to j, F_i can cross F_j from above or below, $I_{ij} = 1$ if the ith alternative crosses the jth alternative from below and zero otherwise; for $i \in S$ and $j \notin S$. We only define I_{ij} for those cases for which $i \in S$ and $j \notin S$. Note that $I_{ij} = 1$ under normality if $\sigma_i < \sigma_j$. We also let C_{ij} be the point at which the distribution functions for the ith alternative and the jth alternative are equal. Let $\underline{C} = \max_{(i,j), I_{ij}=0} C_{ij}$ and $\overline{C} = \min_{(i,j), I_{ij}=1} C_{ij}$. Then $p_\ell = F_{Y(S)}(\underline{C})$ and $p_u = F_{Y(S)}(\overline{C})$ and S remains optimal for all $p_\ell \le p \le p_u$.

3.3. Illustration

Consider three alternatives ($n = 3$) with $X_1 \sim N(0, 25)$, $X_2 \sim N(2.5, 1)$ and $X_3 \sim N(5, 0.04)$. We want to choose a subset of size 2 ($k = 2$). The optimal subset for $p = 1/2$ is found using the algorithm from Section 3.1. Assume $\epsilon = 0.01$.

Step 0 $C_\ell = \max\ (F_1^{-1}(0.5) = 0,\ F_2^{-1}(0.5) = 2.5,\ F_3^{-1}(0.5) = 5) = 5$
$C_u = \max (F_1^{-1}(\sqrt{0.5}) = 2.73,\ F_2^{-1}(\sqrt{0.5}) = 3.04,\ F_3^{-1}(\sqrt{0.5}) = 5.109) = 5.109$

Step 1 $c = 5.0545$
$\gamma_1 = F_1(5.0545) = 0.84$; $\gamma_2 = F_2(5.0545) = 0.99$; $\gamma_3 = F_3(5.0545) = 0.61$
This implies $S = \{1, 3\}$ and $\beta = (0.84)(0.61) = 0.5124$.

Step 2 b(i): $|\beta - p| = 0.0124 > 0.01$ and $\beta > p$; $C_u = 5.0545$

Step 1 $c = (5 + 5.0545)/2 = 5.0273$
$\gamma_1 = F_1(5.0273) = 0.84$; $\gamma_2 = F_2(5.0273) = 0.99$; $\gamma_3 = F_3(5.0273) = 0.55$
This implies $S = \{1, 3\}$ and $\beta = (0.84)(0.55) = 0.4620$

Step 2 b(ii): $|\beta - p| = 0.038 > 0.01$ and $\beta < p$; $C_\ell = 5.0273$

Step 1 $c = (5.0273 + 5.0545)/2 = 5.0409$
$\gamma_1 = F_1(5.0409) = 0.84$; $\gamma_2 = F_2(5.0409) = 0.99$; $\gamma_3 = F_3(5.0409) = 0.58$
This implies $S = \{1, 3\}$ and $\beta = (0.84)(0.58) = 0.4872$

Step 2 b(ii): $|\beta - p| = 0.0128 > 0.01$ and $\beta < p$; $C_\ell = 5.0409$

Step 1 $c = (5.0409 + 5.0545)/2 = 5.0477$
$\gamma_1 = F_1(5.0477) = 0.84$; $\gamma_2 = F_2(5.0477) = 0.99$; $\gamma_3 = F_3(5.0477) = 0.59$
This implies $S = \{1, 3\}$ and $\beta = (0.84)(0.59) = 0.4956$

Step 2 Since $|p - \beta| = 0.0044 \le 0.01$ STOP.

To illustrate the sensitivity analysis we have $I_{12} = 0$ since $\sigma_1 > \sigma_2$ and $I_{23} = 1$ since $\sigma_3 < \sigma_2$. $C_{12} = 3.125$ which is the point at which the cdfs of $N(0, 25)$ and $N(2.5, 1)$ cross. Similarly $C_{23} = 5.625$ is the point at which the distribution functions of $N(5, 0.04)$ and $N(2.5, 1)$ cross. Hence $p_\ell = \Phi((3.125 - 0)/(5))\Phi((3.125 - 5)/(0.2)) \approx 0$ and $p_u = \Phi((5.625 - 0)/(5))\Phi((5.625 - 5)/(0.2)) \cong 0.87$. This implies that $S = \{1, 3\}$ is optimal for all quantiles up to 0.87.

4. Examples

Many applications have a context within which there are a large number of alternatives, but only a few alternatives can be explored because of time and cost constraints. Ultimately, one alternative will be chosen. We discuss two such applications in an attempt to show the richness of the problem setting and to demonstrate the mathematical results from the previous sections.

Example 1: An academic department wants to hire an Assistant Professor. There are ten applicants; however, only three can be interviewed because of time considerations. Note that in order to restrict the output, we assume there are only ten applicants. (The same approach can be used on much larger problems without any difficulty.) Collectively, the department is able to measure the qualifications of the ten candidates by recording the means (to reflect the department's best estimates) and standard deviations [2] (to reflect their uncertainties) as outlined below and a normal distribution is assumed:

Applicant	1	2	3	4	5	6	7	8	9	10
Mean	10.0	10.0	10.0	8.0	8.0	7.0	7.0	6.0	6.0	6.0
Standard deviation	1.0	1.5	3.0	3.5	4.0	5.0	6.5	5.5	6.0	10.0

Assume we establish an objective function of maximizing the median of the best candidate interviewed. If we interview based on means, then applicant 1, 2 and 3 would be interviewed giving rise to a median of the best candidate interviewed of 11.35. However, the optimal choice of candidates, using the algorithm described in the previous section, is to interview candidates 3, 7 and 10. This leads to a median of the best candidate interviewed of 12.75 (i.e., approximately 12% better than the naive ap-

[2] In this case where the evaluation of a candidate is collective, then the mean and standard deviation can be estimated by summarizing the scores across the members of the department. There is an approach that is commonly used in applications for estimating the mean and standard deviation from an individual. The respondent is first asked to give the number s so that the chance is 50% that the candidate will be better than s. The respondent is then told to assume that the candidate's performance is better than s. The number t is then elicited, where the chance that the candidate is better than t given that the candidate is better than s is 50%. The mean and standard deviation can then be derived from s and t.

proach of choosing the highest means). Furthermore, using the methodology described in Section 3, the subset of (3, 7, 10) not only maximizes the median, but also all percentiles between 0.3085 and 0.6154.

Example 2: A multinational corporation is interested in building a plant in a foreign country. Twelve countries are under consideration. Management wants to reduce this list to four countries that can be studied more carefully by teams of employees and consultants. The success in building a plant in a given country depends on many factors including: political climate, cost of labor and materials, import and export tariffs and availability of raw materials. After a careful preliminary analysis the twelve countries are scored as follows:

Country	1	2	3	4	5	6	7	8	9	10	11	12
Mean	1.0	2.0	3.0	4.0	6.0	6.5	7.0	7.5	8.0	8.5	8.75	9.0
Standard deviation	12.0	10.0	8.0	7.0	6.5	6.0	5.5	5.0	4.75	4.5	4.25	4.0

The naive approach chooses countries (9, 10, 11, 12) with a median of the best choice from countries in the subset of 12.38. In contrast, assuming normality, the optimal subset of countries is (1, 5, 9, 10) with a median of best choice from among countries in this subset of 12.80. The optimal subset not only maximizes the median but also all percentiles between 0.47 and 0.56.

Although these data are fabricated, it is not unlikely for the means and standard deviations to be negatively correlated. In example 2, it is plausible that countries 9, 10, 11 and 12 are all well known and developed countries (e.g. Canada, France, Germany and Great Britian) while countries 1, 2 and 3 are countries in which such ventures are rare (e.g. Tanzania, Kenya and New Zealand). The optimal subset, that considers countries from both ends of the spectrum, is intuitively appealing. Learning more about lesser known alternatives might prove to be beneficial.

Example 3: A manufacturer of frozen dinners is interested in testing new food product concepts. The name of the company is not given and the concepts are not described in order to retain confidentiality. Twenty seven concepts are evaluated by 305 consumers. Based on the results of the survey the consulting company recommended the subset of concepts $A = \{12, 13, 21, 25\}$ for further study.

Means and standard deviations across respondents were calculated for each of the twenty seven concepts. We assume a normal distribution for likelihood of purchase, rated on a 1-10 scale, for each of the concepts. The optimal subset algorithm in Section 3.1 was used to produce the subset of size four that maximizes the median. The subset that the algorithm produced was $B = \{11, 18, 21, 22\}$. Note that only concept 21 is in both subsets.

The median of the best concept from among concepts in subset A is 6.42. In contrast, the optimal subset B has a median of 6.57. Furthermore, the sensitivity analysis described in Section 3.2 shows that subset B is optimal for any quantile between 0.48 and 0.73.

5. Conclusions

In this paper, we presented an algorithm for finding a subset of alternatives from a larger number of alternatives. The criterion for selecting this subset is based on the assumption that ultimately one alternative will be chosen from this subset. The intuition behind the approach is that the subset will probably contain alternatives with high expectation and little uncertainty as well as alternatives with relatively low expectation and high uncertainty. Since in the end only one alternative will be chosen and implemented, it might be beneficial to explore some unusual alternatives that are likely to have low expectations a priori.

The criterion that we used was to maximize a quantile, for example, the median of the alternative that finally is chosen and implemented. One area for future research is to consider different criteria, perhaps incorporating the cost of gathering the information about the alternatives. In this context, optimal subset size can be investigated in conjunction with the determination of the optimal subset. In this paper we assumed that alternatives were independent. However, it may be the case that knowing the characteristics of one alternative provides information about the characteristics of other alternatives. Therefore, it is worth exploring the problem considered in this paper when the assumption of independence is relaxed.

Appendix A

Note that each random variable has a normal distribution throughout the discussion in Appendix A.

Proposition A.1. *Let $Y=\max(X_1,X_2)$ where X_1 and X_2 are independent with $X_i \sim N(\mu_i,\sigma_i^2)$, $i=1,2$. Then*

$$E(Y)=\sigma_0\phi\left(\frac{\mu_1-\mu_2}{\sigma_0}\right)+\mu_2\Phi\left(\frac{\mu_2-\mu_1}{\sigma_0}\right)+\mu_1\Phi\left(\frac{\mu_1-\mu_2}{\sigma_0}\right), \tag{A.1}$$

where $\sigma_0=(\sigma_1^2+\sigma_2^2)^{1/2}$.

Proof. $Y=X_1+(X_2-X_1)^+$ where $(X)^+\equiv\max(0, X)$.

$$E(Y)=E(X_1)+E\left[(X_2-X_1)^+\right]. \tag{A.2}$$

But $(X_2-X_1)\sim N(\mu_2-\mu_1,\ \sigma_1^2+\sigma_2^2)$. Hence $X_0\equiv(X_2-X_1)^+=0$ with probability $\Phi\ ((\mu_1-\mu_2)/(\sigma_0))$ and density $f_{X_0}(x_0)=\left(1/\sqrt{2\pi}\sigma_0\right)e^{-(x-\mu_0)^2/2\sigma_0^2}$ for $x_0>0$ where $\mu_0=\mu_2-\mu_1$ and $\sigma_0^2=\sigma_1^2+\sigma_2^2$. This implies

$$\begin{aligned}E(X_0)&=\int_0^\infty x_0\frac{1}{\sqrt{2\pi}\sigma_0}e^{-(x_0-\mu_0)^2/2\sigma^2_0}\,dx_0\\&=\int_{y=-\mu_0/\sigma_0}^\infty(\sigma_0y+\mu_0)\frac{1}{\sqrt{2\pi}}e^{-y^2/2}\,dy\\&=\mu_0\left[1-\Phi\left(\frac{-\mu_0}{\sigma_0}\right)\right]+\sigma_0\phi\left(\frac{-\mu_0}{\sigma_0}\right).\end{aligned} \tag{A.3}$$

Substituting Eq. (A.3) into Eq. (A.2) yields Eq. (A.1). □

Proposition A.2. *Let the alternatives be indexed so that $\mu_1\ge\mu_2\ge\cdots\ge\mu_n$. If $\sigma_1^2\ge\sigma_2^2\ge\cdots\ge\sigma_n^2$ and $S=\{1,\cdots,k\}$, then the median of $Y(S)$ is at least as great as the median of $Y(S_1)$ for any subset S_1 of size k.*

Proof. Let m be the median of $Y(s)$. Clearly, $m\ge\mu_1$. Then $P(X_i\le m)=\Phi((m-\mu_i)/(\sigma_i))\le P(X_j\le m)=\Phi((m-\mu_j)/(\sigma_j))$ for $i>j$ because $0\le m-\mu_i\le m-\mu_j$ and $\sigma_i\ge\sigma_j$. So $\prod_{j=1}^k P(X_{i_j}\le m)\ge 0.5$. Hence the median of $\max(X_{i_1},\cdots,X_{i_k})$ is no greater than m. □

Proposition A.3. *Let $X\sim N(\mu,\sigma^2)$ and $g(\mu,\sigma^2)\equiv E[(X-x_0)^+]$ where x_0 is an arbitrary constant. Then $g(\mu,\sigma^2)$ is nondecreasing in each of its arguments.*

Proof. Using the same approach as in Proposition (A.1), it is easy to verify that

$$g(\mu,\sigma^2)=(\mu-x_0)\left[1-\Phi\left(\frac{x_0-\mu}{\sigma}\right)\right]+\sigma\phi\left(\frac{x_0-\mu}{\sigma}\right).$$

It follows from $\phi'(x)=-x\phi(x)$ that the partial derivatives of $g(\mu,\ \sigma^2)$ in each of its arguments is nonnegative. □

Corollary A.4. *Let X_0 be an arbitary random variable, then $E[\max(X_0,X)]$ is nondecreasing in μ and σ where $X\sim N(\mu,\sigma^2)$.*

Proof. $E[\max(X_0,X)]=E(X_0)+E[(X-X_0)^+]$. But $E[(X-X_0)^+|X_0=x_0]$ is nondecreasing in μ and σ for any x_0 from Proposition (A.3). Hence $E[(X-X_0)^+]=E_{X_0}E[(X-X_0)^+|X_0]$ is nondecreasing in μ and σ. □

Corollary A.5. *Let $Z_1=\max(Y_1,\cdots,Y_{n-1},X_1)$ and $Z_2=\max(Y_1,\cdots,Y_{n-1},X_2)$ where $X_1\sim N(\mu_1,\sigma_1^2)$, $X_2\sim N(\mu_2,\sigma_2^2)$ with $\mu_1\ge\mu_2$ and $\sigma_1^2\ge\sigma_2^2$. Then $E(Z_1)\ge E(Z_2)$.*

Proof. The result follows directly from Corollary A.4 by letting $X_0=\max(Y_1,\cdots,Y_{n-1})$. □

Appendix B

Consider a pair of random variables $X \sim F$ and $Y \sim G$. We are interested in the number of times F and G cross. This problem has been studied extensively in the applied probability literature. Note that if F and G do not cross at all then one random variable is stochastically larger than the other. We are interested in situations in which F and G cross once. We assume throughout that X and Y are absolutely continuous. To this end,

Definition B.1. A pair of random variables (X, Y) is 1-crossing if there exists only one x such that $0 < F_X(x) = F_Y(x) < 1$.

Definition B.2. A family of random variables F_θ; $\theta \in \Theta$ is 1-crossing if F_{θ_1} and F_{θ_2} is 1-crossing for all $\theta_1, \theta_2 \in \Theta$ where $\theta_1 \neq \theta_2$.

We are interested in studying when families of random variables are 1-crossing. We first prove the following lemma.

Lemma B.1. *If $Y = g(X)$ where g is strictly increasing and $g(x) = x$ once then (X,Y) is 1-crossing.*

Proof. $F_Y(y) = P(Y \leq y) = P(X \leq g^{-1}(y)) = F_X(g^{-1}(y))$. But $y = g^{-1}(y) \Leftrightarrow g(y) = y$ which can only occur once. □

Corollary B.2. *If F_θ is a location scale family then F_θ is 1-crossing.*

Proof. We have $\theta_1 = (\mu_1, \sigma_1)$ and $\theta_2 = (\mu_2, \sigma_2)$ with σ_1 and $\sigma_2 \geq 0$. Hence $X = \mu_1 + \sigma_1 Z$ and $Y = \mu_2 + \sigma_2 Z$ where Z has location zero and scale one. This implies $Y = (\mu_2 - \sigma_2 \mu_1/\sigma_1) + \sigma_2 X/\sigma_1$. Since Y is a linear function of X the condition in Lemma B.1 is satisfied.

Example. The normal family is 1-crossing as it satisfies Corollary B.2.

Corollary B.3. *If F_θ is a 1-crossing family and G_θ is created by $Y \in G_\theta \Leftrightarrow Y = h(X)$ where $X \in F_\theta$ and h is a monotonically increasing function, then G_θ is 1-crossing.*

Proof. Consider θ_1 and θ_2. $G_{\theta_1}(y) = G_{\theta_2}(y)$ iff $F_{\theta_1}(h^{-1}(y)) = F_{\theta_2}(h^{-1}(y))$. The results follows. □

Example. The lognormal family is 1-crossing as it satisfies Corollary B.3 with $h(x) = e^x$ and F_θ denoting the normal family.

Lemma B.4. *If X and Y have the same support and their respective density functions $f(x)$ and $g(x)$ cross twice, then (X,Y) is 1-crossing.*

Proof. Assume that $f(x) = g(x)$ only at $x = a$ and $x = b$, $a < b$. Also, $f(x) < g(x)$ only if $x < a$ or $x > b$. Clearly, $F(a) < G(a)$ and $F(b) > G(b)$. There is at least one point x_o with $a < x_o < b$ such that $F(x_o) = G(x_o)$. But $F(x) < G(x)$ for $x < x_o$ and $F(x) > G(x)$ for $x > x_o$ because $f(x) > g(x)\ \forall x \in (a, b)$. □

Example. The gamma family is 1-crossing since

$$\frac{\beta_1^{\alpha_1} x^{\alpha_1 - 1} e^{-\beta_1 x}}{G(\alpha_1)} = \frac{\beta_2^{\alpha_2} x^{\alpha_2 - 1} e^{-\beta_2 x}}{G(\alpha_2)}$$

$$\Leftrightarrow \quad x^{(\alpha_1 - \alpha_2)} e^{-(\beta_1 - \beta_2) x} = k = \frac{G(\alpha_1) \beta_2^{\alpha_2}}{G(\alpha_2) \beta_1^{\alpha_1}}.$$

Since in general $x^\alpha\ e^{-\beta x}$ is a unimodal function it can only cross the line $y = k$ at most twice.

References

Aharoni, Yair (1966), *The Foreign Investment Decision Process*, Harvard University Graduate School of Business Administration.

Bechhofer, R. E. (1954), "A Single-Sample Multiple-Decision Procedure for Ranking Means of Normal Populations with Known Variances", *Annals of Mathematical Statistics*, Vol. 25, pp. 16–39.

Gupta, Shanti, S. and Panchapakesan, S. (1979), *Multiple Decision Procedures*, New York: John Wiley and Sons, Inc.

Kotler, Philip (1980), *Marketing Management*, Englewood Cliffs, New Jersey: Prentice Hall.

Root, Franklin (1982), *Foreign Market Entry Strategies*, New York: American Management Association.

Silk, Alvin J. and Glen L. Urban (1978), ''Pre-Test-Market Evaluation of New Package Goods: A Model and Measurement Methodology'', *Journal of Marketing Research* (May), pp. 171–191.

Stobaugh, Robert B. Jr. (1969a), ''Where in the World Should We Put That Plant?'', *Harvard Business Review* (January-February), pp. 129–136.

Stobaugh, Robert B. Jr. (1969b), ''How to Analyze Foreign Investment Climates'', *Harvard Business Review* (September-October), pp. 100–108.