

RECENT RESULTS ON MULTI-STAGE
SELECTION PROCEDURES

by

K. J. Miescke
Purdue University

Technical Report #82-25

Department of Statistics
Purdue University

July 1982

*Research supported by ONR Contract N00014-75-C-0455 at Purdue University.

ABSTRACT

RECENT RESULTS ON MULTI-STAGE SELECTION PROCEDURES

Klaus J. Miescke (West-Lafayette)

During the past few years, several new developments took place in the area of sequential selection procedures. The purpose of the present paper is to describe the major results and to point out some open problems and interesting questions for further research in the future.

The basic goal is to find that one of k populations which is associated with the largest parameter of a given underlying family of distributions. Additionally, in the control setting, one wishes to decide whether this parameter is large enough, i.e., larger than a control value. A major topic of interest is to find procedures which are reasonably economical, i.e., which perform well without requiring too many observations. The traditional criterion, due to R.E. Bechhofer, which is to guarantee the probability of a correct selection outside of a certain indifference zone, combined with the criterion of keeping the expected total sampling amount small, constitutes the main stream of current research. On the other hand, some work has also been done in the decision theoretic approach, but due to the inherent analytical difficulties, the results are rather incomplete up to now. One promising direction of further research appears to be the construction of procedures which are not too complicated and, at the same time, are at least approximately optimum in a decision theoretic sense.

RECENT RESULTS ON MULTI-STAGE
SELECTION PROCEDURES

K. J. Miescke*
Purdue University

1. Introduction. The problem of how to find the best (in terms of a distributional parameter) of $k \geq 2$ populations, by means of observations drawn from them, has been studied by many research workers in the past. A thorough introduction into the field of Ranking and Selection as well as a complete overview over all relevant developments up to 1979 is provided by Gupta and Panchapakesan [16]. The bibliography contained therein can be complemented with the help of Dudewicz and Koo [7].

The purpose of the present paper is to survey recent results in the sub-field of sequential or, respectively, multi-stage selection procedures which are not already discussed in [16]. Moreover, several open problems will be pointed out to encourage further work in this direction. The most remarkable publication in this respect, without doubt, has been Bechhofer, Kiefer and Sobel [1], which still serves as an inspiring source of results and ideas.

Let the populations, as usual, be denoted by π_1, \dots, π_k . From every π_i , a sequence of observations X_{i1}, X_{i2}, \dots is available to the experimenter. The observations altogether are assumed to be independent. For every i , let the X_{ij} 's have a density f_{θ_i} w.r.t. a sigma-finite measure μ on \mathbb{R} , which is the Lebesgue- or a counting measure, respectively. The family of densities $\mathfrak{F} = \{f_{\theta}\}_{\theta \in \Omega}$, $\Omega \subseteq \mathbb{R}$, is assumed to be known. In many papers, \mathfrak{F} is a one-parameter exponential family. In the continuous case, the most prominent example consists of k normal populations with unknown means $\theta_1, \dots, \theta_k \in \Omega = \mathbb{R}$ and a common known variance $\sigma^2 > 0$ (Normal Case). In the discrete case, it consists of k Bernoulli populations with unknown success probabilities $\theta_1, \dots, \theta_k \in \Omega = [0, 1]$. (Bernoulli Case).

*Reserach supported by ONR Contract N00014-75-C-0455 at Purdue University.

A multi-stage selection procedure for finding that π_j with $\theta_j = \max\{\theta_1, \dots, \theta_k\}$ consists of four different types of decisions which have to be made anew at each subsequent Stage $m = 1, 2, \dots$. More precisely, at Stage m , based on the observations drawn up to that point, the experimenter has to decide

- (a) whether or not he would like to stop (stopping rule);
- (b) In case of not stopping: Which populations to eliminate from further considerations (elimination rule), and what kind of observations to take at the next following Stage $m+1$ (sampling rule);
- (c) In case of stopping: Which population(s) finally to select (terminal decision rule).

Procedures can be categorized in different ways according to their characteristics in (a), (b) or (c), respectively. Some of the terms which are used frequently are the following.

Closed (open) sequential procedure: The number of observations which can be drawn from π_1, \dots, π_k is a bounded (unbounded) random variable.

Truncated procedure: The number of stages is a bounded random variable.

q-stage procedure: The number of stages is a fixed constant $q \geq 1$.

Procedure with elimination: At every stage, populations (which appear to be inferior) can be eliminated from terminal decisions. The remaining populations constitute a subset selection. Typically, eliminated populations are excluded also from further sampling. A justification for this will be given later (cf. Fact 5).

Vector at a time sampling: At every Stage m , exactly n_m observations are taken from every non-eliminated population. The sample sizes n_1, n_2, \dots are determined before the experiment starts.

Adaptive sampling: At every stage, the decision which observations to be taken next depends on the data collected up to that stage. A well known example for the Bernoulli Case is the "play the winner sampling rule" which is due to H. Robbins (cf. [16], p. 64).

Subset selection procedure: The final decisions are subsets of $\{\pi_1, \dots, \pi_k\}$ of random size. A correct selection (CS) occurs if the best population is included.

Fixed size t subset selection procedure: The terminal decisions are subsets of $\{\pi_1, \dots, \pi_k\}$ of fixed size t. In the case of $t = 1$, one calls such a procedure simply a selection procedure. A correct selection has the same meaning as before.

All of the terms above are used quite consistently in the literature, except for the term adaptive. Some authors call their procedures, though employing vector at a time sampling, adaptive because of certain other reasons. See for example Büringer, Martin and Schriever [5] (the nonparametric part), Hsu and Edwards [19] and Tong [48].

The classical approach to find reasonable procedures is the following. Let $\Omega^k(\delta^*) = \{\underline{\theta} \in \Omega^k \mid D(\theta_{[k-1]}, \theta_{[k]}) \geq \delta^*\}$ be the so-called preference zone and $\Omega^k \setminus \Omega^k(\delta^*)$ be the indifference zone, where D is a distance measure (usually to be $\theta_{[k]} - \theta_{[k-1]}$ or $\theta_{[k]}/\theta_{[k-1]}$, resp.), $\delta^* > 0$ is fixed and $\theta_{[1]} \leq \theta_{[2]} \leq \dots \leq \theta_{[k]}$ denote the ordered values of $\theta_1, \dots, \theta_k$. In the indifference zone approach, due to

R. E. Bechhofer (cf. [16], p. 8), only those procedures are considered which satisfy

$$(1) \quad \inf\{P_{\underline{\theta}}(\text{CS}) \mid \underline{\theta} \in \Omega^k(\delta^*)\} \geq P^*,$$

where $P^* > k^{-1}$ is prespecified. A $\underline{\theta} \in \Omega^k(\delta^*)$ is called a least favorable configuration (LFC) if the infimum in (1) occurs at this $\underline{\theta}$. A first step towards establishment of (1) for a suitable type of procedure thus is usually to find its LFC. Among those procedures satisfying (1) one then naturally tries to find a candidate with a small expected terminal subset size and, moreover, with a small average sample number (ASN).

From a decision point of view, this (minimax type) approach means that the risk w.r.t. a 0-1 terminal decision loss should be less than $1-P^*$ on $\Omega^k(\delta^*)$ and that, subject to this condition, objective functions (expected terminal subset size and ASN) are tried to get small. One major objection to the indifference

zone approach is, however, that nothing is actually controlled if $\underline{\theta} \notin \Omega^k(\delta^*)$. Procedures based on this approach will be the main topic of Sections 3-5.

An alternative way of treating the problem is the one prescribed by the decision theory. Hereby one has to incorporate all losses due to inappropriate decisions and costs of sampling into one (stage-dependent) loss function, and then to consider the risk function, i.e. the expected loss, as the objective function which has to be minimized. Within the class of permutation invariant procedures, i.e. among those which give no apriori preference to any of the k populations, several optimality results can be derived, especially in the Bayesian approach. This will be described in Section 2 below.

2. The Decision Theoretic Approach. In this section we assume that \mathcal{F} is a one-parameter exponential family. More precisely, let

$$(2) \quad \mathcal{F} = \{c(\theta)\exp(\theta x)d(x), x \in \mathbb{R}\}_{\theta \in \Omega}, \text{ where } \Omega \subseteq \mathbb{R} \text{ is an interval.}$$

We consider the class \mathcal{X}_I , say, of permutation invariant sequential procedures with (or without) elimination, which are based on vector at a time sampling.

Let $W_{im} = X_{i1} + X_{i2} + \dots + X_{i(n_1 + \dots + n_m)}$ denote the sufficient statistic for θ_i , based on all observation which are available from π_i up to Stage m , $i = 1, \dots, k$, and let $\underline{W}_m = (W_{1m}, \dots, W_{km})$, $m = 1, 2, \dots$.

Let $L_m(\underline{\theta}, (t_1, \dots, t_{m+1}))$ be the loss which occurs at $\underline{\theta} \in \Omega^k$ if a procedure stops at Stage m and finally selects populations $t_{m+1} \subseteq \{\pi_1, \dots, \pi_k\}$, after it has eliminated at Stage j populations $t_j \subseteq \{\pi_1, \dots, \pi_k\}$, $j = 1, \dots, m$, where t_1, \dots, t_{m+1} is a disjoint decomposition of $\{\pi_1, \dots, \pi_k\}$. Assume that for every m , L_m has the following properties:

$$(3a) \quad L_m(\underline{\theta}, (\sigma(t_1), \dots, \sigma(t_{m+1}))) = L_m((\theta_{\sigma(1)}, \dots, \theta_{\sigma(k)}), (t_1, \dots, t_{m+1})),$$

where $\sigma(t_j) = \{\sigma(i) | i \in t_j\}$, $j = 1, \dots, m+1$, for every permutation σ of $(1, \dots, k)$, and

(3b) $L_m(\underline{\theta}, (\tilde{t}_1, \dots, \tilde{t}_{m+1})) \leq L_m(\underline{\theta}, (t_1, \dots, t_{m+1}))$, if for a certain pair (i, j) with $\theta_i \leq \theta_j$ there exist integers $\alpha < \beta \leq m+1$ such that $i \in t_\beta$, $j \in t_\alpha$, $\tilde{t}_\alpha = (t_\alpha \setminus \{j\}) \cup \{i\}$, $\tilde{t}_\beta = (t_\beta \setminus \{i\}) \cup \{j\}$ and $\tilde{t}_\gamma = t_\gamma$ for $\gamma \neq \alpha, \beta$.

Condition (3a) states that L_m is permutation invariant, and (3b) states that a better population should be eliminated later than an inferior one. It is not difficult to see that $-L_m$, if L_m has these two properties, can be represented by a function of $\underline{\theta}$ and a permutation $(\sigma(1), \dots, \sigma(k))$ of $(1, \dots, k)$ which is decreasing in transposition (DT). Functions with this property have been studied by Hollander, Proschan and Sethuraman [18], and their results can be used to derive several optimality results in the present context.

Since the risk function $R(\underline{\theta}, \rho)$, $\underline{\theta} \in \Omega^k$, of a procedure ρ from \mathcal{X}_I is permutation symmetric in $\underline{\theta}$, uniformly (in $\underline{\theta}$) best results in terms of the risk function can be derived more easily in a Bayesian approach under a permutation symmetric prior τ . Thus, let τ be such a prior for the now random parameter vector $\underline{\theta}$. Then one can prove, one after another, the following facts (cf. Gupta and Miescke [12, 14, 15]). Let $m \geq 1$ be fixed in the sequel.

Fact 1. The density of W_m , defined on $\mathbb{R}^k \times \Omega^k$, is (DT).

Fact 2. The posterior density of $\underline{\theta}$ is (DT).

Fact 3. $-E\{L_m(\underline{\theta}, (t_1, \dots, t_{m+1})) | W_m = \underline{w}\}$, as a function of $(\sigma(1), \dots, \sigma(k))$ and \underline{w} , is (DT). Hereby, $\{\sigma(1), \dots, \sigma(q_1)\} = t_1$, $\{\sigma(q_1+1), \dots, \sigma(q_2)\} = t_2$, and so forth, for certain numbers q_1, \dots, q_{m+1} with $q_1 \leq \dots \leq q_{m+1} = k$.

A natural terminal decision, at Stage m , selects only those populations among the non-eliminated ones which yield the largest W_{im} -values. In the discrete case, ties have to be split at random to get a procedure within \mathcal{X}_I . For one-stage procedures it is well known that this type of decision is optimum in several senses (cf. [16], p. 42 and Miescke [30]). The next result can be considered as a generalization of the so-called "Bahadur-Goodman Theorem" (cf. [16], p. 46).

Fact 4. For any $\rho \in \mathcal{X}_I$, let ρ^* be the same procedure as ρ except for the terminal decisions where ρ^* uses the natural ones. Then

$$(4) \quad R(\underline{\theta}, \rho^*) \leq R(\underline{\theta}, \rho), \quad \text{uniformly in } \underline{\theta} \in \Omega^k.$$

Actually, Fact 4 remains unchanged if one assumes that, at every Stage m , the complete vector \underline{W}_m has been observed. With other words, one can state

Fact 5. Observations from eliminated populations are irrelevant for terminal decisions.

The next following, rather negative, statement may be considered, in certain situations, as an argument against the use of adaptive sampling techniques.

Fact 6. For all situations where terminal decisions of a procedure with adaptive sampling are based on unequal numbers of observations from the non-eliminated populations, there is no, uniformly in $\underline{\theta}$, optimal terminal decision.

One might now expect that, within stages where a procedure with elimination does not stop, natural subset selections (i.e. subset selections associated with largest W_{im} -values) have similar strong optimality properties as the natural terminal decisions. It turns out, however, that this is only the case under certain circumstances. First of all, results analogous to Fact 4 can be proved only for strongly unimodal exponential families \mathfrak{F} , i.e. where $f_{\theta}(x)$ or $d(x)$, respectively, is log-concave. The following is a key-lemma.

Fact 7. If \mathfrak{F} is strongly unimodal, then the conditional density of \underline{W}_{m+1} , given $\underline{W}_m = \underline{w}$, which is derived from the joint distribution of $\underline{\theta}$, \underline{W}_m and \underline{W}_{m+1} , is (DT).

If one assumes that at every Stage m , all observations \underline{W}_m are known (just to simplify the proofs), then with backward induction it is not possible to overcome the points where decisions have to be made on how many populations to eliminate. Optimal decision would utilize here all observations, including those from already eliminated populations. The following three results, however, can at least be proved.

Fact 8. Let the number of stages q , say, as well as all subset sizes of selections $r_1 \geq r_2 \geq \dots \geq r_q$, say, at Stages $1, 2, \dots, q$ be fixed. Then the procedure which uses the natural subset selections and natural terminal decision is the unique, uniformly in θ , best procedure in the sub-class of procedures in \mathcal{X}_I with these properties.

Fact 9. Within the sub-class of fixed size t two-stage procedures in \mathcal{X}_I , the procedures which use natural subset selections at Stage 1 and the natural terminal decision at Stage 2 constitute an essentially complete class.

Fact 10. Assume that L_m depends on θ only through those θ_j 's which are associated with the, at Stage m , not eliminated populations, $m = 1, 2, \dots$. If a priori, $\theta_1, \dots, \theta_k$ are independently identically distributed, then every truncated Bayes procedure in \mathcal{X}_I uses natural subset selections at all stages, and the natural terminal decision.

After these considerations under rather mild assumptions upon the loss functions, it is now natural to look for specific procedures which are optimal in more concrete situations. In the control case, where one wishes to select the best population, provided that it is better than the control (i.e. $\theta_{[k]} > \theta_0$), two-stage procedures with elimination have been derived by Miescke [31] and Gupta and Miescke [14]. A Γ -minimax approach, like the one discussed by Miescke [32] for one-stage procedures, appears to be appropriate for such problems, but no work has been done here up to now.

The difficulties arising in concrete problems with the backward induction in sequential selection problems are considerable (cf. Edwards [8]). It seems to be more realistic for future work to look for simple structured procedures which are approximately optimal in a reasonable sense. Ad hoc procedures like the ones proposed by Washburn [52] may have good performance characteristics and deserve to be studied in more detail. Washburn's procedures are open and closed procedures

without elimination. They are based on prior knowledge and use adaptive sampling which, as well as the terminal decision, is based on the posterior expectations of $\theta_1, \dots, \theta_k$. These procedures are not Bayes solutions in a decision theoretic sense since no overall loss is actually considered here. However, as it is shown by Washburn [52], they perform well compared with other procedures given in the literature.

Two papers which give more detailed Bayes solutions in concrete problems, but which do not completely fit into the distributional framework considered so far, are due to Ramey and Alam [41] and Gulati [9]. Ramey and Alam [41] derive a Bayes truncated procedure for the most probable of k cells in a multinomial model under a Dirichlet prior. Gulati [9] considers the problem of finding that one of k uniform distributions which has slipped to the left (and has a shorter support). He finds the Bayes solution with respect to any prior specifying the slipped population, within the class of closed sequential procedures based on adaptive sampling.

In the sections to come, procedures will be discussed which are not derived from the decision theoretic approach. Most of them are fixed size t (especially $t = 1$) subset selection procedures. All of them use the natural subset selections within stages and the natural terminal decisions.

3. The Bernoulli Case and Related Models. The case of k Bernoulli populations with unknown success probabilities $\theta_1, \dots, \theta_k \in \Omega = [0, 1]$ will be discussed only briefly, since two detailed publications on recent developments in this area are readily available. The first one is Büringer, Martin and Schriever [5]. In their book, various sequential selection procedures with vector at a time as well as play the winner sampling and different stopping rules are studied under the indifference zone approach in great detail. The second one is the paper by Bechhofer and Kulkarni [2] which provides an excellent overview not only over the

k-population selection problem but also over related areas like clinical trials and multi-armed bandit problems, where the objective functions differ from those in the selection problem. These related areas are also covered by Dudewicz and Koo [7], where further references can be found.

The main topic of Bechhofer and Kulkarni [2], however, is to propose a closed non-eliminating adaptive fixed size t subset selection procedure, which has several optimal properties in terms of the $P(\text{CS})$ and the expected number of observations taken from certain populations. Some results, which are proved for $k = 2$ only, are conjectured to hold also for $k \geq 3$. In a subsequent paper, Bechhofer and Kulkarni [3] provide additional results on the performance of their procedure.

Levin and Robbins [27] consider an open non-eliminating procedure with vector at a time sampling which stops if one population has r more "successes" than all the remaining ones. Among others, a conjecture is stated for a procedure with elimination, which is proved to hold for the non-eliminating version.

The related multinomial case (note that here independence of the cell-frequencies is clearly not given, but that the joint distribution is (DT)), where the goal is to find the cell with the largest probability, is also treated by Levin and Robbins [27]. A closed sequential (inverse sampling) version of their procedure has been studied already by Ramey and Alam [40], where a conjecture concerning the LFC for $k \geq 3$ is stated. A Bayes procedure by Ramey and Alam [41] has been mentioned already in Section 2. Further work has been done by Hwang [24] and Hsuang, Hwang and Parnes [22]. The latter is actually a one-stage result, but the question of whether more sampling is more informative is certainly of relevance for sequential selection problems, too.

4. The Normal Case and Related Models. The problem of finding that one of k normal populations $N(\theta_i, \sigma_i^2)$, $i = 1, \dots, k$, which has the largest mean in more than one stage has been studied under several aspects (but mainly under the indifference

zone approach) by many research workers. In the following we shall distinguish between three different situations depending on the status of knowledge about $\sigma_1^2, \dots, \sigma_k^2$. The first one is the simplest one: Here, $\sigma_1^2 = \dots = \sigma_k^2 = \sigma^2$ and $\sigma^2 > 0$ is known. For this model, Bechhofer, Kiefer and Sobel [1] proposed and studied their, meanwhile classical, open sequential procedure without elimination which is based on vector at a time sampling with $1 = n_1 = n_2 = \dots$. The terminal decision is the natural one (which is optimum, cf. Fact 4), and the stopping rule is

$$(4) \quad N_{\text{BKS}} = \inf\{m \mid \sum_{i=1}^{k-1} \exp(-\delta^*(Y_{[k]m} - Y_{[i]m})) \leq (1-P^*)/P^*\}$$

where $Y_{im} = (X_{i1} + \dots + X_{im})/\sigma^2$, $i = 1, \dots, k$,

which establishes the procedure in the indifference zone approach (i.e. (1)). An upper bound for the first moment of N_{BKS} is given in Huang [23], and asymptotic properties of this stopping rule, i.e. the behavior of the ASN under $P^* \rightarrow 1$ and/or $\delta^* \rightarrow 0$ have been studied by Bechhofer, Kiefer and Sobel [1], Tong [49] and Jennison, Johnstone and Turnbull [25].

As with the BKS-procedure, many others can be viewed, in one way or another, as being generalizations of Wald's SPRT (cf. [16], p. 127). This is also the case with that one in Mukhopadhyay [37]. It differs from the BKS-procedure only through its stopping rule, N_M , say.

$$(5) \quad N_M = \inf\{m \mid (k-1) \max_{i \neq k} \exp(-\delta^*(Y_{[k]m} - Y_{[i]m})) \leq 1-P^*\}.$$

Apparently, $N_{\text{BKS}} \leq N_M$ and $E_{\underline{\theta}}(N_{\text{BKS}}) < E_{\underline{\theta}}(N_M)$ for all $\underline{\theta} \in \mathbb{R}^k$, which implies that the BKS-procedure is more efficient. Usually it is more difficult to compare procedures directly in this manner and asymptotic techniques are then the only way to do this.

Under the indifference zone approach, open sequential procedures with elimination and vector at a time sampling, which are based on paired comparisons of the sample means, are discussed by Swanepoel and Geertsema [44], Kao and Lai [26], Hsu and Edwards [20], Turnbull, Kaspi and Smith [50] and Jennison, Johnstone

and Turnbull [25]. In the latter two papers, however, procedures with adaptive sampling are the main topic.

An open sequential procedure without elimination based on vector at a time sampling is proposed by Tong [48] in a more general setting including the normal case. To avoid the indifference zone approach, the single stage $P(\text{CS})$ is hereby estimated stage by stage (by replacing the unknown parameters by estimators), and the procedure stops as soon as this estimate is greater or equal to P^* . It would be interesting to derive a probability guarantee for the $P(\text{CS})$ in $\Omega^k(\delta^*)$. In $\Omega^k \setminus \Omega^k(\delta^*)$, however, one has to encounter (similar as with the BKS-procedure) a large ASN (cf. McCulloch [28]).

Even if the ASN is finite, there remains the uncertainty of how long it actually takes until an open procedure eventually stops (see also Bechhofer and Kulkarni [2], 2.2). From a practical point of view, truncated and q -stage procedures with elimination seem to meet more likely the needs of practitioners. A two-stage procedure with elimination and vector at a time sampling is proposed by Tamhane and Bechhofer [47]. The elimination hereby is made by means of Gupta's maximum means rule (cf. [16], p. 232):

(6) Select π_i if $Y_{i n_1} \geq Y_{[k] n_1} - d$, $i = 1, \dots, k$,

where $d = d(\delta^*, P^*, n_1, n_2)$ in the present context. A multi-stage procedure which is a direct generalization of Tamhane and Bechhofer's procedure is proposed by Tamhane [46].

Several optimality properties of Gupta's subset selection rule have been pointed out recently by Gupta and Miescke [11], Miescke [29], Gupta and Kim [10] and Bickel and Yahav [4]. Thus the use of this rule at the first stage is intuitively justified. No theoretical results, however, which support this idea are known at present.

In a conservative approach, Tamhane and Bechhofer [47] use a lower bound for the $P(\text{CS})$ to find a most economical pair (n_1, n_2) in the indifference zone approach. Their conjecture that the slippage configuration $\theta_{[1]} = \dots = \theta_{[k-1]} = \theta_{[k]} - \delta^*$ is the LFC has been proved to be correct for $k = 3$ by Miescke and Sehr [33]. The case of $k > 3$ is still unproved. Techniques for finding LFC's as well as results for other procedures are given in Gupta and Miescke [13].

For the control problem, several two-stage procedures are considered and discussed in Gupta and Miescke [14] and Miescke [31].

The second situation, where still $\sigma_1^2 = \dots = \sigma_k^2 = \sigma^2$, but $\sigma^2 > 0$ is now unknown, is also considered by Kao and Lai [26] and Jennison, Johnstone and Turnbull [25].

Open sequential procedures without elimination based on vector at a time sampling are studied by Wackerley [51] (in a more general approach) and by Mukhopadhyay and Chou [38]. The latter perform a similar analysis as before Mukhopadhyay [37] has done before. Mukhopadhyay [36] deals with the case of $k = 2$.

The third situation, where $\sigma_1^2, \dots, \sigma_k^2$ are unknown and possibly unequal, is the most difficult one. It is questionable, however, whether it is still reasonable to look for a population with the largest mean which perhaps might also have the largest variance.

Two-stage procedures without elimination, based on the classical Stein-technique (cf. [16], p. 23) to determine the common sample size at Stage 2 in dependence of the estimated variances at Stage 1, are considered by Rinott [42] and Mukhopadhyay [34] under the indifference zone approach. A three-stage procedure employing the Stein approach, followed by elimination via Gupta's rule at Stage 2, is proposed by Hochberg and Marcus [17] in the indifference zone approach. For the case of $k = 2$ populations, other procedures have been considered by Mukhopadhyay [35]. Procedures are also given by Swanepoel and Geertsema [44].

For the control case, open and closed sequential procedures with elimination and vector at a time sampling are proposed by Hsu and Edwards [19].

5. Other results. For the problem of finding the normal population with the smallest variance, Mukhopadhyay and Chou [39] give an open sequential procedure without elimination based on vector at a time sampling. The basic construction is the same as before in Mukhopadhyay [37]. For the linear regression model, another type of sequential procedure is proposed by Hsu and Huang [21].

Finally, several papers dealing with nonparametric sequential procedures are presented by Swanepoel and Venter [45], Swanepoel [43], Carroll [6], Büringer, Martin and Schriever [5] and Hsu and Edwards [20].

Acknowledgement. The author wishes to express his sincere thanks to Professor S. S. Gupta for his encouragement and support. Thanks are also due to Professor S. Panchapakesan for several helpful comments.

References

1. Bechhofer, R. E., Kiefer, J., Sobel, M.: Sequential Identification and Ranking Procedures. Chicago-London: The University of Chicago Press 1968.
2. Bechhofer, R. E., Kulkarni, R. V.: Closed adaptive sequential procedures for selecting the best of $k \geq 2$ Bernoulli populations. In: Statistical Decision Theory and Related Topics III, Vol. 1, pp. 61-108, Gupta, S. S. and Berger, J. O. eds., New York: Academic Press 1982.
3. Bechhofer, R. E., Kulkarni, R. V.: On the performance characteristics of a closed adaptive sequential procedure for selecting the best Bernoulli population. Tech. Rep. #532, School of Operations Research and Industrial Engineering, College of Engineering, Cornell University, Ithaca 1982.
4. Bickel, P. J., Yahav, J. A.: Asymptotic theory of selection procedures and optimality of Gupta's rules. In: Statistics and Probability: Essays in Honor of C. R. Rao, pp. 109-124. Kallianpur, G., Krishnaiah, P. R., Ghosh, J. K. eds., Amsterdam-New York-Oxford: North-Holland 1982.
5. Büringer, H., Martin, H., Schriever, K.-H.: Nonparametric Sequential Selection Procedures. Boston-Basel-Stuttgart: Birkhäuser 1980.

6. Carroll, R. J.: Distributions of stopping times for sequential elimination ranking procedures. *South African Statist. J.* 12, 33-45 (1978).
7. Dudewicz, E. J., Koo, J. O.: *The Complete Categorized Guide to Statistical Selection and Ranking Procedures*. Columbus, Ohio: American Sciences Press 1982.
8. Edwards, H. P.: Bayes sequential design procedures: Theory and applications. (Abstract). *Bull. Austral. Math. Soc.* 21, 157-158 (1980).
9. Gulati, C. M.: Optimal sequential procedure for finding the slipped population. *Commun. Statist. C*, submitted for publication, 1982.
10. Gupta, S. S., Kim, W. C.: On the problem of selecting good populations. *Commun. Statist. A10*, 1043-1077 (1981).
11. Gupta, S. S., Miescke, K. J.: Optimality of subset selection procedures for ranking means of three normal populations. *Sankhyā B43*, 1-17 (1981).
12. Gupta, S. S., Miescke, K. J.: An essentially complete class of two-stage procedures with screening at the first stage. Mimeo. Ser. #81-53, Dept. of Statist., Purdue University, W. Lafayette, 1981.
13. Gupta, S. S., Miescke, K. J.: On the least favorable configurations in certain two-stage selection procedures. In: *Statistics and Probability: Essays in Honor of C. R. Rao*, pp. 295-305. Kallianpur, G., Krishnaiah, P. R., Ghosh, J. K. eds., Amsterdam-New York-Oxford: North-Holland 1982.
14. Gupta, S. S., Miescke, K. J.: On the problem of finding a best population with respect to a control in two stages. In: *Statistical Decision Theory and Related Topics III*, Vol. 1 pp. 473-496, Gupta, S. S., Berger, J. O. eds., New York: Academic Press 1982.
15. Gupta, S. S., Miescke, K. J.: Sequential selection procedures: A decision theoretic approach. Mimeo. Ser. #82-6, Dept. of Statist., Purdue University, W. Lafayette, 1982.
16. Gupta, S. S., Panchapakesan, S.: *Multiple Decision Procedures: Theory and Methodology of Selecting and Ranking Populations*. New York-Chichester-Brisbane-Toronto: Wiley 1979.
17. Hochberg, Y., Marcus, R.: Three stage elimination type procedures for selecting the best normal population when variances are unknown. *Commun. Statist. A10*, 597-612 (1981).
18. Hollander, M., Proschan, F., Sethuraman, J.: Functions decreasing in transposition and their applications in ranking problems. *Ann. Statist.* 5, 722-733 (1977).
19. Hsu, J., Edwards, D. G.: Adaptive sequential procedures for comparing new treatments with a standard. Tech. Rep. #235, Dept. of Statist., Ohio State University, Columbus 1981.
20. Hsu, J., Edwards, D. G.: Sequential multiple comparison with the best. Tech. Rep. #260, Dept. of Statist., Ohio State University, Columbus, 1982.

21. Hsu, T. A., Huang, D. Y.: Some sequential selection procedures for good regression models. Mimeo. Ser. #81-21, Dept. of Statist., Purdue University, W. Lafayette, 1981.
22. Hsuan, F., Hwang, F. K., Parnes, M.: A class of selection problems for which more sampling is more informative. *Biometrika* 69, 280-283 (1982).
23. Huang, D. Y.: A note on an efficiency of the Bechhofer-Kiefer-Sobel sequential selection procedure. *Bull. Inst. Math. Acad. Sinica* 4, 249-252 (1976).
24. Hwang, F. K.: A multi-stage selection scheme for the most probable event. Unpublished manuscript, 1980.
25. Jennison, C., Johnstone, I. M., Turnbull, B. W.: Asymptotically optimal procedures for sequential adaptive selection of the best of several normal means. In: *Statistical Decision Theory and Related Topics III*, Vol. 2, pp. 55-86, Gupta, S. S., Berger, J. O. eds., New York: Academic Press 1982.
26. Kao, S. C., Lai, T. L.: Sequential selection procedures based on confidence sequences for normal populations. *Commun. Statist.* A9, 1657-1676 (1980).
27. Levin, B., Robbins, H.: Selecting the highest probability in binomial or multinomial trials. *Proc. Nat. Acad. Sci. U.S.A.*, submitted for publication, 1981.
28. McCulloch, C. E.: Conditions under which $E\{N_1\} = \infty$ for Tong's adaptive solution to ranking and selection problems. *Commun. Statist.* A11, 815-819 (1982).
29. Miescke, K. J.: Bayesian subset selection procedures for additive and linear loss functions. *Commun. Statist.* A8, 1205-1226 (1979).
30. Miescke, K. J.: Identification and selection procedures based on tests. *Ann. Statist.* 7, 207-219 (1979).
31. Miescke, K. J.: On two-stage procedures for finding a population better than a control. Mimeo. Ser. #80-28, Dept. of Statist., Purdue University, W. Lafayette, 1980.
32. Miescke, K. J.: Γ -Minimax selection procedures in simultaneous testing problems. *Ann. Statist.* 9, 215-220 (1981).
33. Miescke, K. J., Sehr, J.: On a conjecture concerning least favorable configurations in certain two-stage selection procedures. *Commun. Statist.* A9, 1609-1617 (1980).
34. Mukhopadhyay, N.: Some comments on two-stage selection procedures. *Commun. Statist.* A8, 671-683 (1979).
35. Mukhopadhyay, N.: A note on selecting the better treatment. Tech. Rep. #21, Dept. of Statist., Oklahoma State University, Stillwater, 1981.
36. Mukhopadhyay, N.: Theoretical investigations on some sequential and two-stage procedures to select the larger mean. Tech. Rep. #22, Dept. of Statist., Oklahoma State University, Stillwater, 1981.

37. Mukhopadhyay, N.: Selecting the largest normal mean through likelihoods. *Statistics and Decisions*, to appear, 1982.
38. Mukhopadhyay, N., Chou, W. S.: Sequential procedures for selecting the largest normal mean when the common variance is unknown. Tech. Rep. #30, Dept. of Statist., Oklahoma State University, Stillwater, 1982.
39. Mukhopadhyay, N., Chou, W. S.: Selecting the smallest normal variance through the comparisons of several likelihoods. *Commun. Statist. C*, to appear, 1982.
40. Ramey, J. T., Alam, K.: A sequential procedure for selecting the most probable multinomial event. *Biometrika* 66, 171-173 (1979).
41. Ramey, J. T., Alam, K.: A Bayes sequential procedure for selecting the most probable multinomial event. *Commun. Statist. A9*, 265-276 (1980).
42. Rinott, Y.: On two-stage selection procedures and related probability-inequalities. *Commun. Statist. A7*, 799-811 (1978).
43. Swanepoel, J.W.H.: Nonparametric elimination selection procedures based on robust estimators. *South African Statist. J.* 11, 27-41 (1977).
44. Swanepoel, J.W.H., Geertsema, J. C.: Sequential procedures with elimination for selecting the best of k normal populations. *South African Statist. J.* 10, 9-36 (1976).
45. Swanepoel, J.W.H., Venter, J. H.: A class of elimination selection procedures based on ranks. *South African Statist. J.* 9, 119-128 (1975).
46. Tamhane, A. C.: On a class of multi-stage selection procedures with screening for the normal means problem. *Sankhyā B42*, 197-216 (1980).
47. Tamhane, A. C., Bechhofer, R. E.: A two-stage minimax procedure with screening for selecting the largest normal mean (II): An improved PCS lower bound and associated tables. *Commun. Statist. A8*, 337-358 (1979).
48. Tong, Y. L.: An adaptive solution to ranking and selection problems. *Ann. Statist.* 6, 658-672 (1978).
49. Tong, Y. L.: Some new properties of the Bechhofer-Kiefer-Sobel stopping rule. *Ann. Inst. Statist. Math.* 31, 423-433 (1979).
50. Turnbull, B. W., Kaspi, H., Smith, R. L.: Adaptive sequential procedures for selecting the best of several normal populations. *J. Statist. Comput. Simul.* 7, 133-150 (1978).
51. Wackerley, D. D.: An alternative sequential approach to the problem of selecting the best of k populations. *J. American Statist. Assoc.* 72, 464-468 (1977).
52. Washburn, A. R.: NOMBAS-A Bayesian procedure for selecting the greatest mean. Tech. Rep. #NPS55-82-017, Naval Postgraduate School, Monterey, California, 1982.

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER Technical Report #82-25	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Recent Results on Multi-Stage Selection Procedures		5. TYPE OF REPORT & PERIOD COVERED
		6. PERFORMING ORG. REPORT NUMBER Technical Report #82-25
7. AUTHOR(s) Klaus-J. Miescke		8. CONTRACT OR GRANT NUMBER(s) N00014-75-C-0455
9. PERFORMING ORGANIZATION NAME AND ADDRESS Purdue University Department of Statistics West Lafayette, IN 47907		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Washington, DC		12. REPORT DATE July 1982
		13. NUMBER OF PAGES 16
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report) Approved for public release, distribution unlimited.		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) Multiple decision procedures, sequential selection procedures, Bayesian analysis, indifference zone approach.		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) During the past few years, several new developments took place in the area of sequential selection procedures. The purpose of the present paper is to describe the major results and to point out some open problems and interesting questions for further research in the future. The basic goal is to find that one of k populations which is associated with the largest parameter of a given underlying family of distributions. Additionally, in the control setting, one wishes to decide whether this parameter is large enough, i.e., larger than a control value. A major topic of interest is to find procedures		

DD FORM 1 JAN 73 1473

UNCLASSIFIED

which are reasonably economical, i.e., which perform well without requiring too many observations. The traditional criterion, due to R. E. Bechhofer, which is to guarantee the probability of a correct selection outside of a certain indifference zone, combined with the criterion of keeping the expected total sampling amount small, constitutes the main stream of current research. On the other hand, some work has also been done in the decision theoretic approach, but due to the inherent analytical difficulties, the results are rather incomplete up to now. One promising direction of further research appears to be the construction of procedures which are not too complicated and, at the same time, are at least approximately optimum in a decision theoretic sense.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE(When Data Entered)