Statistica Sinica **18**(2008), 203-218

A GENERALIZATION OF THE LEVIN-ROBBINS PROCEDURE FOR BINOMIAL SUBSET SELECTION AND RECRUITMENT PROBLEMS

Cheng-Shiun Leu[1,2] and Bruce Levin[2]

[1]*New York State Psychiatric Institute and* [2]*Columbia University*

Abstract: We introduce a family of sequential selection and recruitment procedures for the subset identification problem in binomial populations. We demonstrate the general validity of a simple formula providing a lower bound for the probability of correct identification in a version of the family without sequential elimination or recruitment. A new application of the non-central hypergeometric distribution is revealed. A similar theorem is conjectured to hold for the more efficient version which employs sequential elimination or recruitment.

Key words and phrases: Lower bound formula, probability of correct selection, recruitment, selection, sequential identification.

1. Introduction

Procedures to identify the best of several populations or subsets of best populations have a large literature. The textbooks by Gibbons, Olkin and Sobel (1977), Büringer, Martin and Schriever (1980) and Bechhofer, Santner and Goldsman (1995), and the references contained therein, give a good summary of the known theoretical and numerical properties of these procedures. The identification and ranking paradigm itself has wide-ranging applications in early phase clinical trials and industrial product testing, to mention only two areas of interest. The monograph of Bechhofer, Kiefer and Sobel (1968) on sequential procedures for identification and ranking is widely cited, although these authors do not discuss how to adapt their procedures for the general subset selection problem to allow for sequential elimination of inferior populations, nor are we aware of other candidate procedures.

In this paper we propose a procedure to solve the general subset selection problem for binomial populations that allows for sequential elimination of inferior populations. Easy to use, our procedure generalizes that of Hoel and Mazumdar (1968) and Levin and Robbins (1981) for selecting the single best binomial population. An important result in the latter paper and subsequent work by Leu and Levin (1999a,b) is a simple but useful lower bound formula for the probability of correct selection that is used to design the elimination criterion (see below), and which materially strengthens the Hoel and Mazumdar (1968) result. We refer to the combined technique of using the Hoel and Mazumdar elimination rule together with the lower bound formula to provide the strengthened elimination criterion as the Levin-Robbins procedure. Our object here is to establish some theoretical properties of the proposed method with or without elimination. Without elimination, we establish the validity of the lower bound formula for the probability of correctly selecting the best subset. With elimination, we have proven some partial results and conjecture the general validity of the formula. In addition, sequential elimination of inferior populations opens a new perspective on the class of such procedures, allowing us to distinguish between sequential *selection* versus *recruitment*.

We conclude this section by introducing notation and reviewing previous results for the Levin-Robbins procedure with and without elimination. In Section 2 we introduce the notion of sequential selection versus recruitment and discuss some theoretical relations between them. In particular we provide lower bound formulas conjectured to hold for both types of procedures. Section 3 provides a rigorous proof of the lower bound formula for the general subset selection procedure without elimination. We end in Section 4 with a novel application of the non-central hypergeometric distribution to provide a bound on the distribution of the number of "good" binomial populations selected from a collection of "good" and "poor" populations.

Suppose we have $c \geq 2$ coins, with labels in the set $C = \{1, \ldots, c\}$. For coin i $(i \in C)$, let p_i be the probability of heads on a single toss. For a given integer b $(1 \leq b < c)$, our goal is to identify a subset of b coins with the highest such probabilities, which we shall call a subset of b best coins. To accomplish this goal, consider the following method, which generalizes the sequential selection procedure of Levin and Robbins (1981). Toss the coins vector-at-a-time. For $n = 1, 2, \cdots$, let $\mathbf{X}^{(n)} = (X_1^{(n)}, X_2^{(n)}, \ldots, X_c^{(n)})$ be the vector that reports the cumulative number of heads observed for each coin after n tosses, and let $\mathbf{X}^{[n]} = (X_1^{[n]}, X_2^{[n]}, \ldots, X_c^{[n]})$ be the ordered $\mathbf{X}^{(n)}$ vector with $X_1^{[n]} \geq X_2^{[n]} \geq \cdots \geq X_c^{[n]}$. Let r be a positive integer chosen in advance of all tosses. Define the stopping time $M = M_r^{*(b,C)}$ to be the first time that the b^{th} largest tally exceeds the $(b+1)^{st}$ largest tally by r; that is to say,

$$M = M_r^{*(b,C)} = \inf\{n \geq 1 : X_b^{[n]} - X_{b+1}^{[n]} = r\}.$$

If $M = n$, we stop after n tosses of the set of coins and select those b coins with $X_b^{[n]}$ heads or more. There may be several coins tied for b^{th} best or for $(b+1)^{st}$ best, but the subset of b coins in the lead at $M = n$ is unique. We will also call these coins a subset of "b best coins" — the context will make clear whether these are truly best or only apparently best in the experiment.

$M_r^{*(b,C)}$ is a stopping time, i.e., $P[M_r^{*(b,C)} < \infty] = 1$; we omit the proof as it is essentially the same as for Wald's original SPRT. If S is any subset of C of size b ($S \subset C, |S| = b$), let $P_r^*[S]$ denote the probability of selecting S with this procedure. Thus $P_r^*[S] = P[X_i^{(M)} \geq X_j^{(M)} + r$ for all pairs (i,j) with $i \in S$ and $j \notin S]$. If $i_1, \ldots, i_b$ are integers satisfying $1 \leq i_1 < \cdots < i_b \leq c$, for notational convenience we write the subset $\{i_1, \ldots, i_b\}$ as $[i_1 \cdots i_b]$, without commas, and $P_r^*[\{i_1, \ldots, i_b\}]$ as $P_r^*[i_1 \cdots i_b]$.

Theorem 1. *For any set C of c coins with probabilities $\{p_1, \ldots, p_c\}$, let $w_i = p_i/(1-p_i)$, and suppose, without loss of generality, that $p_1 \geq p_2 \geq \cdots \geq p_c$. Then for any two subsets of integers $\{i_1, \ldots, i_b\} \subset C$ and $\{j_1, \ldots, j_b\} \subset C$ satisfying $i_k \leq j_k$ for all $k = 1, \ldots, b$, and for any positive integer r in procedure $M_r^{*(b,C)}$, we have*

$$\frac{P_r^*[i_1 \cdots i_b]}{P_r^*[j_1 \cdots j_b]} \geq \prod_{k=1}^{b} \left(\frac{w_{i_k}}{w_{j_k}}\right)^r.$$

The proof of Theorem 1 will be given in Section 3. Notice that a simple lower bound formula for the probability of correct selection $P[CS] = P_r^*[1 \cdots b]$ follows immediately from Theorem 1 because $P_r^*[1\cdots b] \geq \prod_{k=1}^{b} (w_k/w_{j_k})^r \cdot P_r^*[j_1\cdots j_b]$ for all $1 \leq j_1 < \cdots < j_b \leq c$, so that summing over all subsets of size b not equal to $[1 \cdots b]$, we have

$$P_r^*[1 \cdots b] = 1 - \sum_{[j_1,\ldots,j_b] \neq [1\cdots b]} P_r^*[j_1 \cdots j_b] \geq 1 - \sum_{[j_1\cdots j_b] \neq [1\cdots b]} \prod_{k=1}^{b} \left(\frac{w_k}{w_{j_k}}\right)^{-r} \cdot P_r^*[1 \cdots b].$$

This yields the following

Corollary 1. $P[CS] = P_r^*[1 \cdots b] \geq (w_1 \cdots w_b)^r / \sum_{[i_1\cdots i_b]} (w_{i_1} \cdots w_{i_b})^r$, *where the summation is over all subsets $[i_1 \cdots i_b]$ such that $1 \leq i_1 < i_2 < \cdots < i_b \leq c$.*

We refer to the expression on the right hand side of the inequality as a *lower bound formula* for the probability of correct selection of a subset of b best coins, and denote it by $L_+(b,c) = L_+(b, c|w_1, \ldots, w_c)$, although it clearly depends homogeneously on $w_1, \ldots, w_c$ only through their ratios. For example, for $b = 2$ and $c = 4$, $L_+(2,4) = (w_1w_2)^r/\{(w_1w_2)^r + (w_1w_3)^r + (w_1w_4)^r + (w_2w_3)^r + (w_2w_4)^r + (w_3w_4)^r\} = 1/(1 + w_{23}^{-r} + w_{24}^{-r} + w_{13}^{-r} + w_{14}^{-r} + w_{13}^{-r}w_{24}^{-r})$, where the $w_{ij} = w_i/w_j$ are odds ratios.

To see how the corollary is applied in practice, suppose we wish to guarantee $P[CS] \geq 0.95$ for selecting the best $b = 2$ out of $c = 4$ coins with odds w_1, w_2, w_3, w_4, such that $w_2/w_3 \geq 2$. Using the above expression for $L_+(2,4)$, we find

$P[CS] \geq L_+(2,4) \geq 1/(1+4\cdot 2^{-r}+2^{-2r})$ which exceeds 0.95 for $r \geq 7$. In fact $P[CS] \geq 1/(1+4\cdot 2^{-7}+2^{-14}) \cong 0.97$ for any such set of coins.

Now consider a modification of the procedure in which we eliminate "inferior" coins as soon as they fall behind the b^{th} best coin or coins by r heads. Let $N_r^{(b,C)}$ be the *time of first elimination in a c-coin game with coins C*,

$$N_r^{(b,C)} = \inf\{n \geq 1 : X_b^{[n]} - X_c^{[n]} = r\}.$$

Strictly speaking we should retain b in the notation; where b is fixed in the discussion, however, we will simply write $N_r^{(C)}$. If $N_r^{(C)} = N_r^{(b,C)} = n$, we drop from further consideration after toss n any and all coins i satisfying $X_i^{(n)} = X_c^{[n]}$, i.e., we eliminate all coins that have fallen r heads behind the b^{th} best. If more than b coins remain, the procedure continues, starting from the current tallies of the remaining subset of coins $C' \subset C$, and iterates with $N_r^{(C')}$, replacing c by $c' = |C'|$, and so on, until $c-b$ coins have been eliminated. Thereupon we declare the remaining coins as the b best. Let $P_r^{(C)}[i_1 \cdots i_b]$ be the probability that coins $i_1, \ldots, i_b$ are identified as the b best by this procedure with sequential elimination.

Are Theorem 1 and/or Corollary 1 true for the elimination procedure, with $P_r^{(C)}[i_1 \cdots i_b]$ replacing $P_r^*[i_1 \cdots i_b]$? Leu and Levin (1999a) proved that for the sequential elimination procedure, Corollary 1 remains true for any c and any r in the special case $b=1$ although, surprisingly, the stronger Theorem 1 does *not* hold true for all $p_1 \geq \cdots \geq p_c$, even in the special case $b=1$; for a counterexample, see Zybert and Levin (1987). Thus, while Theorem 1 is a "natural" reason for the truth of Corollary 1 in the procedure without elimination, if Corollary 1 does remain true for $b>1$ in the procedure with elimination it must be established by other means, as it was in Leu and Levin (1999a) for $b=1$. Even for the case $r=1$ for $c=4$ coins, where an exact expression for $P[CS]$ is available, it is quite complex and not at all obvious that it satisfies the lower bound (see Levin and Leu (2004)). A rigorous proof of Corollary 1 for the case $b=2$ with $c=3$ coins is already contained in the results of Leu and Levin (1999a) and, more generally, the case $b=c-1$ follows from the results for the procedure without elimination, as explained below in Section 2. We now conjecture that Corollary 1 holds for any c, any $p_1 \geq \cdots \geq p_c$, any r, and for any $1 \leq b < c$. We prove the conjecture in the first non-trivial special case of $b=2$ and $c=4$; see Leu and Levin (2004). A rigorous proof of the conjecture in general for $b>1$ and $c>4$ is still an open problem, although extensive numerical experiments do support the conjecture.

We assume the truth of the general conjecture in the next section, where we point out that the family of sequential procedures with elimination offers a useful connection between identification by selection or by recruitment.

2. Selection vs Recruitment

By a *selection* procedure we mean a method wherein the identified subset of coins becomes known precisely when the procedure terminates. The procedure to identify the best b out of c coins with elimination of inferior coins is a selection procedure because the final choice of b best coins becomes apparent only at the end, after $c-b$ coins have been eliminated. By a *recruitment* procedure we mean a method wherein the identified coins become available sequentially, allowing early utilization of the recruited units. In an industrial testing context, for example, it would generally be desirable to bring the better units "on-line" as soon as they qualify as belonging to the required subset of b best, rather than waiting for testing to be complete. Similarly, for early phase II screening of c potential medical therapies or behavioral interventions, one would generally prefer early recruitment of superior modalities for further definitive testing. This contrasts with the conventional paradigm for late phase II or phase III testing in a clinical trials context, wherein a decision concerning the b best treatments is generally deferred until the end of the trial, while for ethical reasons we prefer to eliminate inferior treatments as soon as we can.

It may be apparent to the reader that the selection procedure with sequential elimination of inferior coins is already a recruitment procedure for the $c-b$ *worst* coins. This is nearly a tautology, capitalizing on the obvious assertion that any procedure for which the probability of identifying the best b out of c coins is bounded from below by $L_+(b,c)$ is, at the same time, a procedure for which the probability of identifying the complementary subset of the worst $c-b$ coins is at least $L_+(b,c)$. For example, if in some context it were desired to recruit the $c-b$ coins with the *smallest* probability of heads, we would simply apply the selection procedure for the b best coins with sequential elimination (i.e., recruitment) of the $c-b$ worst coins.

We make more progress by introducing the parity transformation in which heads become tails and tails become heads and, similarly, adjectives like "best" or "highest" become "worst" or "lowest", and *vice versa*. By this device we arrive directly at a selection procedure for the b worst coins. Let $Y_j^{[n]} = n - X_{c-j+1}^{[n]}$ be the ordered tallies of tails, with $Y_1^{[n]} \geq Y_2^{[n]} \geq \cdots \geq Y_c^{[n]}$, which is also the ordered tallies of "heads" after the parity transformation. We restate the time of first elimination $N_r^{(b,C)}$ as applied to the transformed tallies in original, untransformed notation as follows:

$$\inf\{n \geq 1 : Y_b^{[n]} - Y_c^{[n]} = r\} = \inf\{n \geq 1 : X_1^{[n]} - X_{c-b+1}^{[n]} = r\}$$

at which time we eliminate any coins with transformed tallies $Y_c^{[n]}$, i.e., we recruit any coins with original tallies (of heads) equal to $X_1^{[n]}$.

Note that the original assumed ordering of odds $w_1 \geq \cdots \geq w_c$ reverses after transformation to $w_c^{-1} \geq \cdots \geq w_1^{-1}$. Thus the lower bound formula for the probability of correct selection of b "best" coins transforms invariantly into the following lower bound formula for the probability of correct selection of the b worst coins, which we write as $L_-(b,c)$:

$$\begin{aligned} L_-(b,c|w_1,\ldots,w_c) &= L_+(b,c|w_c^{-1},\ldots,w_1^{-1}) \\ &= \frac{(w_c\cdots w_{c-b+1})^{-r}}{\sum_{[i_1\cdots i_b]}(w_{i_1}\cdots w_{i_b})^{-r}} = \frac{(w_1\cdots w_{c-b})^{r}}{\sum_{[i_1\cdots i_{c-b}]}(w_{i_1}\cdots w_{i_{c-b}})^{r}} \\ &= L_+(c-b,c|w_1,\ldots,w_c). \end{aligned}$$

The penultimate equality follows by multiplying numerator and denominator by $(w_1\cdots w_b)^r$ and noting that the sum in the denominator enumerating all $\binom{c}{b}$ subsets of b indices also enumerates all $\binom{c}{c-b}$ subsets of $c-b$ indices.

Finally, we observe that the equivalence noted above, between selection of best coins and recruitment of worst coins after parity reversal, yields the following procedure for recruiting the $c-b$ best coins: at time

$$N_r^{\dagger(c-b,C)} = inf\{n \geq 1 : X_1^{[n]} - X_{c-b+1}^{[n]} \geq r\},$$

recruit any coins with tallies of heads equal to the best, $X_1^{[n]}$. If fewer than $c-b$ coins have been recruited, continue with the remaining subset of coins $C' \subset C$ at their current tallies, and iterate with $N_r^{\dagger(c'-b,C')}$, $c' = |C'|$, and so on, until $c-b$ coins have been recruited. Note that in the notation for the recruitment times, b remains fixed while c changes as the procedure is iterated. Our general conjecture implies that the probability of correct recruitment for this procedure is bounded from below by the formula $L_+(c-b,c|w_1,\ldots,w_c)$.

We conclude this section with some observations.

(i) In the procedure to select the b best coins with sequential elimination of inferior coins, there is *no claim* that the first coin to be eliminated is the truly worst coin among the c coins. Rather we assert only that at the time of first elimination, the evidence is sufficiently strong to place the eliminated coin among the subset of $c-b$ worst coins with joint probability of correct assertion at least $L_+(b,c)$. Similarly, in recruiting the best $c-b$ coins, we make no claim that the first coin to be recruited is the truly best coin. We merely assert that at each time of recruitment, the evidence is sufficiently strong to place the recruited coins among the subset of $c-b$ best coins, with joint probability of correct assertions at least $L_+(c-b,c)$.

(ii) In the special case of recruiting the best coin using $N_r^{\dagger(1,C)}$, the procedure involves no iteration, because in the case $c-b=1$ one terminates at the time of first (and only) recruitment $N_r^{\dagger(1,C)} = inf\{n \geq 1 : X_1^{[n]} - X_2^{[n]} \geq r\}$. We recognize this rule as the original Levin-Robbins procedure without elimination for selecting the best coin, $M_r^{*(1,C)}$. It is interesting to find this rule as a member of our family of sequential identification procedures after all, as an extreme case of recruitment. Note that, because of this identity, the strong theorem of Levin and Robbins (1981) already demonstrates the validity of the lower bound formula $L_+(1,c)$ for the recruitment procedure $N_r^{\dagger(1,C)}$. Furthermore, recruiting the one best coin is equivalent to selecting the $c-1$ worst coins. Thus by the parity transformation, we also have the validity of the lower bound $L_+(c-1,c) = L_-(1,c)$ for selecting the best $b = c-1$ coins using $N_r^{*(c-1,C)}$. This is the special case $b = c-1$ of Theorem 1 referred to in the previous section.

(iii) In a selection procedure with elimination of inferior coins, as testing continues we gain more and more experience with the apparently better coins (e.g., treatments); this is the essence of selection. The essence of recruitment is that the best units are removed from testing and can be brought into action as soon as they are identified, thus sparing possibly precious lifetime. These and other differences are summarized in Figure 1.

(iv) A thorough evaluation of the operating characteristics of the family of selection and recruitment procedures and comparisons with other sequential selection procedures, notably that of Paulson (1994), lie beyond the scope of this paper and will be presented elsewhere, see Levin and Leu (2007). Instead we point out some characteristic properties of the present procedures through an example. Table 1 gives some simulation results for selection or recruitment of the best $b = 1, 2$, or 3 coins out of $c = 4$ coins. All estimates are based on 100,000 replications. The operating characteristics displayed are:

(1) the actual probability of correct selection (labelled $P[CS]$);
(2) the expected number of rounds, i.e., vectors of coins tossed (labelled $E[N]$);
(3) the expected total number of tosses (labelled $E[T\,]$); for a procedure without elimination, $E[T\,] = 4E[N]$; and
(4) the expected total number of tails (labelled $E[F]$, for failures).

For each value of b, the probability vector $\mathbf{p}$ is on the boundary of a zone of indifference: $\mathbf{p} = (0.2, 0.1, 0.1, 0.1)$ for $b = 1$; $\mathbf{p} = (0.2, 0.2, 0.1, 0.1)$ for $b = 2$; and $\mathbf{p} = (0.2, 0.2, 0.2, 0.1)$ for $b = 3$. The value $r = 5$ was chosen to give probability at least 0.95 of correct selection at these parameters. For $b = 1$ or $b = 3$, the lower bound formula $L_+(b, c) \geq 0.95$, while for $b = 2$, $L_+(2, 4) < 0.95$ but $P[CS] > 0.95$ nevertheless. We chose the same value, $r = 5$, in this case for straightforward comparison.

	BEST (HIGHEST PROB's)	WORST (LOWEST PROB'S)
SELECTION	• Identifies (selects) b best • Waits until procedure ends (the essence of selection) • Eliminates inferior units seq'ly • Tests more with better units • P[correct selection] $\geq L_{+}(b, c)$ • Lowers $E[T]$, $E[F]$ $(b < c/2)$ Lowers $E[N]$ $(b > c/2)$	• Identifies (selects) b worst • Waits unitil procedure ends (the essence of selection) • Eliminates superior units seq'ly • Tests more with worse units • P[correct selection] $\geq L_{-}(b, c)$ • Lowers $E[T]$, $E[S]$ $(b < c/2)$ Lowers $E[T]$ $(b > c/2)$
PECRUITMENT	• Identifies (selects) $c - b$ best • Brings units on line sequentially (the essence of recruitment) • Continues testing inferior units • Spares testing of better units • P[correct recruit'] $\geq L_{+}(c - b, c)$ • Lowers $E[T]$, $E[F]$ $(c - b < c/2)$ Lowers $E[N]$ $(c - b > c/2)$	• Identifies (selects) $c - b$ worst • Brings units on line sequentially (the essence of recruitment) • Continues testing superior units • Spares testing of worse units • P[correct recruit'] $\geq L_{-}(c - b, c)$ • Lowers $E[T]$, $E[S]$ $(c - b < c/2)$ Lowers $E[T]$ $(c - b > c/2)$

Figure 1.

For selecting the best coin ($b = 1$) we find the selection procedure with elimination of inferior coins has noticeably fewer total tosses and failures than the recruitment procedure. As noted above, in this case there is no early recruitment so that there are more tosses and failures than for selection, which can capitalize on early elimination. The situation is reversed for the case $b = 3$. Now the recruitment procedure can take advantage of early recruitment of coins while the selection procedure must continue with no early elimination.

These relations do not depend on the particular parameter values of 0.2 and 0.1 chosen for the illustration. For example, if instead of $\mathbf{p} = (0.2, 0.1, 0.1, 0.1)$

we chose $\mathbf{p} = (0.9, 0.8, 0.8, 0.8)$ for $b = 1$, the selection procedure would be equivalent to the recruitment procedure for identifying the best $b = 3$ coins with $\mathbf{p} = (0.2, 0.2, 0.2, 0.1)$ by virtue of the parity transformation and, likewise, the recruitment procedure for $b = 1$ with $\mathbf{p} = (0.9, 0.8, 0.8, 0.8)$ would be equivalent to the selection procedure for $b = 3$ with $\mathbf{p} = (0.2, 0.2, 0.2, 0.1)$. Consulting the bottom panel in Table 1 allows us to conclude that selection has smaller $E[T]$ and $E[F]$ for $\mathbf{p} = (0.9, 0.8, 0.8, 0.8)$. By similar reasoning, recruitment has smaller $E[T]$ and $E[F]$ for $\mathbf{p} = (0.9, 0.9, 0.9, 0.8)$ than does selection in the case $b = 3$ — reverse parity, and consult the top panel for $b = 1$ with $\mathbf{p} = (0.2, 0.1, 0.1, 0.1)$.

The case $b = 2$ is a little more delicate. The table shows recruitment has slightly fewer expected total tosses and tails than does selection for the given parameter $\mathbf{p} = (0.2, 0.2, 0.1, 0.1)$. At $\mathbf{p} = (0.9, 0.9, 0.8, 0.8)$, however, we would find the opposite, because in that case recruitment would be equivalent to selection of the $b = 2$ best coins with $\mathbf{p} = (0.2, 0.2, 0.1, 0.1)$ by the parity transformation and, similarly, selection at $\mathbf{p} = (0.9, 0.9, 0.8, 0.8)$ would be equivalent to recruitment at $\mathbf{p} = (0.2, 0.2, 0.1, 0.1)$.

Numerical evidence suggests the above remarks generalize to larger values of c. For values of $b < c/2$, selection generally has smaller $E[T]$ and $E[F]$ than recruitment for any $\mathbf{p}$, while the reverse is true for $b > c/2$. For $b = c/2$, recruitment has smaller $E[T]$ and $E[F]$ than selection for $p_1 < 0.5$, while the reverse is true for $p_c > 0.5$.

Table 1. Identifying the best b out of $c = 4$ coins. (All simulation results based on 100,000 replications. The $\pm$ entries are standard errors of the estimate.)

$b=1$ $c=4$ $r=5$	$\mathbf{p}$= (0.2, 0.1, 0.1, 0.1)		$L_+(1,4) = 0.9505$	
Method	$P[CS]$	$E[N]$	$E[T]$	$E[F]$
Selection	0.954 ± 0.0007	65.1 ± 0.14	205.3 ± 0.37	178.5 ± 0.32
Recruitment	0.972 ± 0.0005	65.6 ± 0.13	262.5 ± 0.51	229.7 ± 0.45

$b=2$ $c=4$ $r=5$	$\mathbf{p}$= (0.2, 0.2, 0.1, 0.1)		$L_+(2,4) = 0.9349$	
Method	$P[CS]$	$E[N]$	$E[T]$	$E[F]$
Selection	0.956 ± 0.0006	76.6 ± 0.14	276.9 ± 0.45	234.0 ± 0.39
Recruitment	0.957 ± 0.0006	75.8 ± 0.13	266.2 ± 0.42	228.0 ± 0.37

$b=3$ $c=4$ $r=5$	$\mathbf{p}$= (0.2, 0.2, 0.2, 0.1)		$L_+(3,4) = 0.9505$	
Method	$P[CS]$	$E[N]$	$E[T]$	$E[F]$
Selection	0.974 ± 0.0005	72.3 ± 0.12	289.1 ± 0.49	238.5 ± 0.41
Recruitment	0.957 ± 0.0006	70.4 ± 0.12	209.1 ± 0.31	174.2 ± 0.26

Which procedure to choose in any application depends on the context. For $b < c/2$, selection has the advantage of smaller $E[T]$ and $E[F]$ while, for $b > c/2$ the advantage goes to recruitment. In cases where a preference for selection or recruitment would not enjoy this advantage ($b > c/2$ or $b < c/2$, respectively), one would have to weigh the relative importance of selection or recruitment against the larger values of $E[T]$ and $E[F]$ than would obtain with the other method.

(v) The reader may object that the actual values of $P[CS]$ in Table 1 for selection and recruitment are not exactly equal, and therefore the values of $E[T]$ and $E[F]$ for the two methods are not directly comparable. One way to remove this objection is to truncate the procedure with the higher $P[CS]$ so that it terminates at some maximum number of rounds, such that the $P[CS]$ of the truncated procedure agrees with the $P[CS]$ of the other procedure. (If the procedure does not reach an early decision, at truncation time the b coins in the lead are selected; if there are ties for b^{th} best coins, select at random. For recruitment, at truncation time coins not yet chosen are recruited from those in the lead, with randomization to break ties if necessary.) For example, for $b = 1$, we truncated the recruitment procedure to $N_{max} = 125$ rounds, resulting in a simulated $P[CS]$ of $0.955 \pm .0007$. The corresponding values of $E[T]$ and $E[F]$ were, respectively, $250.5 \pm .41$ and 219.3 ± 0.36, see Table 2. Note that these values are still greater than those of the selection procedure. Thus our substantive conclusions comparing selection with recruitment in comment (iv) do not change.

An interesting final comparison emerges with the proper calibration of $P[CS]$ via truncation. The expected number of rounds, $E[N] = 65.1 \pm 0.14$, for the selection procedure with $b = 1$ is now slightly *greater* than the expected number of rounds for the truncated recruitment procedure, $E[N] = 62.6 \pm 0.10$. The reverse is true for $b = 3$. These differences are not large, but illustrate a general inverse relation between superiority in terms of $E[T]$ and $E[F]$ on the one hand, and $E[N]$ on the other.

3. Proof of Theorem 1

Returning to stopping time $M = M_r^{*(b,C)}$ and the procedure without elimination, we use Wald's change of measure argument to prove Theorem 1. Let $\mathbf{p} = (p_1, \ldots, p_c)$, with $p_1 \geq \cdots \geq p_c$. Let α be any infinite sequence of c-dimensional binary outcome vectors, and let $X_i^{(n)} = X_i^{(n)}(\alpha)$ denote the cumulative number of heads with coin i after n tosses for sequence α. Let

$$f_{\mathbf{p}}^{(n)}(\alpha) = \prod_{i=1}^{c} p_i^{X_i^{(n)}} (1 - p_i)^{n - X_i^{(n)}}$$

Table 2. Comparison of selection with recruitment with equalization of $P[CS]$ via truncation. (All simulation results based on 100,000 replications. The $\pm$ entries are standard errors of the estimate.)

$b = 1$ $c = 4$ $r = 5$ $\mathbf{p}= (0.2, 0.1, 0.1, 0.1)$ $L_+(1, 4) = 0.9505$ (Recruitment procedure is truncated at $N_{\max} = 125$.)				
Method	$P[CS]$	$E[N]$	$E[T]$	$E[F]$
Selection	0.954 ± 0.0007	65.1 ± 0.14	205.3 ± 0.37	178.5 ± 0.32
Recruitment	0.955 ± 0.0007	62.6 ± 0.10	250.5 ± 0.41	219.3 ± 0.36

$b = 2$ $c = 4$ $r = 5$ $\mathbf{p}= (0.2, 0.2, 0.1, 0.1)$ $L_+(2, 4) = 0.9349$				
Method	$P[CS]$	$E[N]$	$E[T]$	$E[F]$
Selection	0.956 ± 0.0006	76.6 ± 0.14	276.9 ± 0.45	234.0 ± 0.39
Recruitment	0.957 ± 0.0006	75.8 ± 0.13	266.2 ± 0.42	228.0 ± 0.37

$b = 3$ $c = 4$ $r = 5$ $\mathbf{p}= (0.2, 0.2, 0.2, 0.1)$ $L_+(3, 4) = 0.9505$ (Selection procedure is truncated at $N_{\max} = 125$.)				
Method	$P[CS]$	$E[N]$	$E[T]$	$E[F]$
Selection	0.956 ± 0.0006	68.8 ± 0.10	275.4 ± 0.39	227.2 ± 0.32
Recruitment	0.957 ± 0.0006	70.4 ± 0.12	209.1 ± 0.31	174.2 ± 0.26

be the probability function for the first n components of sequence α. For any integer ν ($1 \le \nu \le b$), and any two disjoint subsets of integers $I = \{i_1, \ldots, i_\nu\}$ and $J = \{j_1, \ldots, j_\nu\}$ with $1 \le i_1 < \cdots < i_\nu \le c$ and $1 \le j_1 < \cdots < j_\nu \le c$, such that $i_k < j_k$ for all $k = 1, \ldots, \nu$, let $\mathbf{p_{IJ}}$ be the original $\mathbf{p}$ vector but with p_{i_k} and p_{j_k} interchanged for all k. For example, if $b = 2$, $c = 4$, $I = \{1, 2\}$, and $J = \{3, 4\}$, then $\mathbf{p_{IJ}} = (p_3, p_4, p_1, p_2)$; if $I = \{1, 3\}$ and $J = \{2, 4\}$ then $\mathbf{p_{IJ}} = (p_2, p_1, p_4, p_3)$. To take care of the case of overlapping subsets, we extend the notation as follows. For arbitrary subsets I and J with b elements each and $i_k \le j_k$ for all $k = 1, \ldots, b$, let $I' = I \backslash J = \{i'_1, \ldots, i'_\nu\}$, and $J' = J \backslash I = \{j'_1, \ldots, j'_\nu\}$, where $\nu = |I'| = |J'| = b - |I \cap J|$. Then we write $\mathbf{p_{I'J'}}$ to indicate that only the non-overlapping indices in I and J are to be transposed. For example if $I = \{1, 3\}$ and $J = \{2, 3\}$ then $\mathbf{p_{I'J'}} = (p_2, p_1, p_3, p_4)$.

We claim that $i'_k < j'_k$ for $k = 1, \ldots, \nu$. To see this, note that we can obtain I' and J' by eliminating the elements in $I \cap J$ from I and J, respectively, using the following iterative algorithm. Let l and m be the integers such that $i_l = j_m = \max\{n : n \in I \cap J\}$. Since $j_m = i_l \le j_l$, we have $m \le l$. We claim that when we remove i_l from I and j_m from J, $i_k^{(b-1)} \le j_k^{(b-1)}$ for all $i_k^{(b-1)} \in I^{(b-1)} = I - \{i_l\}$ and $j_k^{(b-1)} \in J^{(b-1)} = J - \{j_m\}$. This is true because $i_l = j_m$ is the largest element in $I \cap J$, so that for any $k < m$, $i_k^{(b-1)} = i_k \le j_k = j_k^{(b-1)}$; for k such that $m \le k < l$,

$i_k^{(b-1)} = i_k \le j_k < j_{k+1} = j_k^{(b-1)}$; and for $k \ge l$, $i_k^{(b-1)} = i_{k+1} < j_{k+1} = j_k^{(b-1)}$. The removal of integers in $I \cap J$ may then be iterated with $I^{(b-1)}$ and $J^{(b-1)}$, etc., until we remove all the elements in $I \cap J$ from I and J and end up with I' and J'. It is then clear that the strict inequalities $i'_k < j'_k$ are preserved, because $i'_k \le j'_k$ but $I' \cap J'$ is empty.

The proof of Theorem 1 proceeds by writing $I = [i_1 \cdots i_b]$, $J = [j_1 \cdots j_b]$, and $P_r^*[i_1 \cdots i_b]$ as

$$P_r^*[i_1 \cdots i_b | \mathbf{p}] = \sum_{n=1}^{\infty} \sum_{[\mathrm{M\ =\ n,\ select\ I}]} f_{\mathbf{p}}^{(n)}(\alpha),$$

where the second summation is over all α leading to the event of selection of coins in I as the b best at time $M = n$, so that $X_b^{[n]} - X_{b+1}^{[n]} = r$ with $X_i^{(n)} \ge X_j^{(n)} + r$ for all $i \in I$ and $j \notin I$. Then as noted above, $i'_1 < j'_1, \ldots, i'_\nu < j'_\nu$ and $P_r^*[i_1 \cdots i_b | \mathbf{p}]$ equals

$$\begin{aligned}
&\sum_{n=1}^{\infty} \sum_{[\mathrm{M\ =\ n,\ select\ I}]} f_{\mathbf{p}}^{(n)}(\alpha) \\
&\quad = \sum_{n=1}^{\infty} \sum_{[\mathrm{M\ =\ n,\ select\ I}]} \prod_{i=1}^{c} p_i^{X_i^{(n)}(\alpha)} (1-p_i)^{n - X_i^{(n)}(\alpha)} \\
&\quad = \sum_{n=1}^{\infty} \sum_{[\mathrm{M\ =\ n,\ select\ I}]} \prod_{i=1}^{c} (1-p_i)^n w_i^{X_i^{(n)}(\alpha)} \\
&\quad = \sum_{n=1}^{\infty} \sum_{[\mathrm{M\ =\ n,\ select\ I}]} \prod_{i=1}^{c} (1-p_i)^n \prod_{i \in I \cap J} w_i^{X_i^{(n)}(\alpha)} \prod_{k=1}^{\nu} w_{i'_k}^{X_{i'_k}^{(n)}(\alpha)} w_{j'_k}^{X_{j'_k}^{(n)}(\alpha)}.
\end{aligned}$$

Multiplying and dividing by the factor $\prod_{k=1}^{\nu} w_{j'_k}^{X_{i'_k}^{(n)}(\alpha)} w_{i'_k}^{X_{j'_k}^{(n)}(\alpha)}$ on the right side of the above equation, we have

$$\begin{aligned}
&P_r^*[i_1 \cdots i_b | \mathbf{p}] \\
&\quad = \sum_{n=1}^{\infty} \sum_{[\mathrm{M\ =\ n,\ select\ I}]} \prod_{i=1}^{c} (1-p_i)^n \prod_{i \in I \cap J} w_i^{X_i^{(n)}(\alpha)} \prod_{k=1}^{\nu} w_{j'_k}^{X_{i'_k}^{(n)}(\alpha)} w_{i'_k}^{X_{j'_k}^{(n)}(\alpha)} \\
&\qquad \times \prod_{k=1}^{\nu} \left(\frac{w_{i'_k}}{w_{j'_k}} \right)^{X_{i'_k}^{(n)}(\alpha) - X_{j'_k}^{(n)}(\alpha)} \\
&\quad = \sum_{n=1}^{\infty} \sum_{[\mathrm{M\ =\ n,\ select\ I}]} f_{\mathbf{p}_{\mathbf{I}'\mathbf{J}'}}^{(n)}(\alpha) \prod_{k=1}^{\nu} \left(\frac{w_{i'_k}}{w_{j'_k}} \right)^{X_{i'_k}^{(n)}(\alpha) - X_{j'_k}^{(n)}(\alpha)}.
\end{aligned}$$

The last equation holds because the first three products comprise the probability of sequence α where the probabilities for coins in I' and J' are transposed. Now, on the event that coins $i_1, \ldots, i_b$ are selected as the b best coins at time $M = n$, we have $X^{(n)}_{i'_k}(\alpha) - X^{(n)}_{j'_k}(\alpha) \geq r$ for $i'_k \in I'$ and $j'_k \in J'$, while each pair of matching coins in the overlap portion can be uniquely written as $i_k = j_l \in I \cap J$ for some subscripts k and l with $w_{i_k}/w_{j_l} = 1$. Therefore $\prod_{k=1}^{\nu}(w_{i'_k}/w_{j'_k})^{X^{(n)}_{i'_k}(\alpha)-X^{(n)}_{j'_k}(\alpha)} \geq \prod_{k=1}^{\nu}(w_{i'_k}/w_{j'_k})^r = \prod_{k=1}^{b}(w_{i_k}/w_{j_k})^r$, so that,

$$
\begin{aligned}
P^*_r[i_1 \cdots i_b|\mathbf{p}] &\geq \sum_{n=1}^{\infty} \sum_{\alpha \in [\mathrm{M\ =\ n,\ select\ I}]} f^{(n)}_{\mathbf{p_{I'J'}}}(\alpha) \prod_{k=1}^{b} \left(\frac{w_{i_k}}{w_{j_k}}\right)^r \\
&= \sum_{n=1}^{\infty} \sum_{\alpha' \in [\mathrm{M\ =\ n,\ select\ J}]} f^{(n)}_{\mathbf{p}}(\alpha') \prod_{k=1}^{b} \left(\frac{w_{i_k}}{w_{j_k}}\right)^r \\
&= P^*_r[j_1 \cdots j_b|\mathbf{p}] \prod_{k=1}^{b} \left(\frac{w_{i_k}}{w_{j_k}}\right)^r .
\end{aligned}
$$

The first equality holds because there exists a one-to-one correspondence between sequences $\alpha \in [M = n$ and select $I]$ and $\alpha' \in [M = n$ and select $J]$ which is established by interchanging binary outcomes in position i'_k with those in position j'_k for $k = 1, \ldots, \nu$, such that $f^{(n)}_{\mathbf{p_{I'J'}}}(\alpha) = f^{(n)}_{\mathbf{p}}(\alpha')$. This completes the proof of Theorem 1.

4. An application

Theorem 1.1 leads to an interesting application of the noncentral hypergeometric distribution. Let a, b, c, and d be positive integers satisfying $d \leq a \leq b < c$, and suppose there are a "good" coins with probabilities $p_1 = \cdots = p_a$ and $c - a$ "poor" coins with probabilities $p_{a+1} = \cdots = p_c$, such that $\{p_a/(1-p_a)\}/\{p_{a+1}/(1-p_{a+1})\} = \omega > 1$. Suppose we use procedure $M^{*(b,c)}_r$ to select the "best" b coins. Let the random variable G denote the number of "good" coins appearing among the b coins selected. We are interested in the probability that $G \geq d$ for any integer $d = 1, \ldots, a$. Let Y have the noncentral hypergeometric distribution with indices a, b, c and parameter ω^r, given by

$$
P[Y = y] = \binom{a}{y}\binom{c-a}{b-y}\omega^{ry} \Big/ \sum_u \binom{a}{u}\binom{c-a}{b-u}\omega^{ru},
$$

where the sum extends over integers u satisfying $\max(0, a+b-c) \leq u \leq a$. Then Theorem 1.1 implies that G is stochastically greater than Y, i.e., $P^*_r[G \geq d] \geq P[Y \geq d]$ for all d. The distribution of Y corresponds to the fourfold table in Figure 2 with all margins fixed, and a preference for selecting good coins corresponding to the amplified odds ratio ω^r.

	Good	Poor	
Selected	Y		b
Not Selected			$c-b$
	a	$c-a$	c

Figure 2.

To demonstrate the inequality, we write

$$P_r^*[G \geq d] = \sum_{y=d}^{a} \sum_{[i_{y1}\cdots i_{yb}]} P[i_{y1}\cdots i_{yb}],$$

where the inner summation is over all subsets of b integers with exactly y integers less than or equal to a. Since we are assuming $p_1 = \cdots = p_a$ and $p_{a+1} = \cdots = p_c$, all terms of the form $P_r^*[i_{y1}\cdots i_{yb}]$ are equal to $P_r^*[\{1,\ldots,y,a+1,\ldots,a+b-y\}]$. Thus we can write

$$P_r^*[G \geq d] = \sum_{y=d}^{a} \binom{a}{y}\binom{c-a}{b-y} P_r^*[\{1,\ldots,y,a+1,\ldots,a+b-y\}]$$

or, in odds form,

$$\frac{P_r^*[G \geq d]}{1 - P_r^*[G \geq d]} = \frac{\displaystyle\sum_{y=d}^{a} \binom{a}{y}\binom{c-a}{b-y} P_r^*[\{1,\ldots,y,a+1,\ldots,a+b-y\}]}{\displaystyle\sum_{u=\max(0,a+b-c)}^{d-1} \binom{a}{u}\binom{c-a}{b-u} P_r^*[\{1,\ldots,u,a+1,\ldots,a+b-u\}]}.$$

Then by interchanging $y-d$ good coins with poor coins and applying Theorem 1.1, we have that the numerator is no less than

$$\sum_{y=d}^{a} \binom{a}{y}\binom{c-a}{b-y} \omega^{r(y-d)} P_r^*[\{1,\ldots,d,a+1,\ldots,a+b-d\}]$$

and, by interchanging $(d-1)-u$ good coins with poor coins, the denominator is no greater than

$$\sum_{u=\max(0,a+b-c)}^{d-1} \binom{a}{u}\binom{c-a}{b-u} \omega^{r(u-d+1)} P_r^*[\{1,\ldots,d-1,a+1,\ldots,a+b-d+1\}].$$

This implies

$$
\begin{aligned}
\frac{P_r^*[G \geq d]}{1 - P_r^*[G \geq d]} &\geq \sum_{y=d}^{a} \binom{a}{y}\binom{c-a}{b-y}\omega^{r(y-d)} \Big/ \sum_{u=\max(0,a+b-c)}^{d-1} \binom{a}{u}\binom{c-a}{b-u}\omega^{r(u-d+1)} \\
&\quad \times \frac{P_r^*[\{1,\ldots,d,a+1,\ldots,a+b-d\}]}{P_r^*[\{1,\ldots,d-1,a+1,\ldots,a+b-d+1\}]} \\
&= \sum_{y=d}^{a} \binom{a}{y}\binom{c-a}{b-y}\omega^{ry} \Big/ \sum_{u=\max(0,a+b-c)}^{d-1} \binom{a}{u}\binom{c-a}{b-u}\omega^{ru} \\
&\quad \times \frac{P_r^*[\{1,\ldots,d,a+1,\ldots,a+b-d\}]}{\omega^r \cdot P_r^*[\{1,\ldots,d-1,a+1,\ldots,a+b-d+1\}]} \\
&\geq \sum_{y=d}^{a} \binom{a}{y}\binom{c-a}{b-y}\omega^{ry} \Big/ \sum_{u=\max(0,a+b-c)}^{d-1} \binom{a}{u}\binom{c-a}{b-u}\omega^{ru},
\end{aligned}
$$

where the last inequality holds true because $P_r^*[\{1,\ldots,d,a+1,\ldots,a+b-d\}]/\{\omega^r \cdot P_r^*[\{1,\ldots,d-1,a+1,\ldots,a+b-d+1\}]\} \geq 1$, again by Theorem 1.1. Therefore $P_r^*[G \geq d] \geq P[Y \geq d]$.

We conjecture the same result holds for the procedure $N_r^{(b,c)}$ with elimination of inferior coins.

Acknowledgement

This work was supported in part by NIMH Grant 5-P30-MH43520-16 to the HIV Center for the Clinical and Behavioral Studies.

References

Bechhofer, R. E., Kiefer, J. and Sobel, M. (1968). *Sequential Identification and Ranking Procedures.* University of Chicago Press, Chicago.

Bechhofer, R. E., Santner, T. J. and Goldsman, D. M. (1995). *Design and Analysis of Experiments for Statistical Selection, Screening, and Multiple Comparisons.* Wiley, New York.

Büringer, H., Martin, H., and Schriever, K.-H. (1980). *Nonparametric Sequential Selection Procedures.* Birkhauser, Boston.

Gibbons, J. D., Olkin, I. and Sobel, M. (1977). *Selecting and Ordering Populations: A New Statistical Methodology.* Wiley, Hoboken; corrected, unabridged version (1999), Soc. for Industrial & Applied Math, Philadelphia.

Hoel, D. G. and Mazumdar, M. (1968). An extension of Paulson's selection procedure. *Ann. Math. Statist.* **39**, 2067-2074.

Leu, C. S. and Levin, B. (1999a). On the probability of correct selection in the Levin-Robbins sequential elimination procedure. *Statist. Sinica* **9**, 879-891.

Leu, C. S. and Levin, B. (1999b). Proof of a lower bound formula for the expected reward in the Levin-Robbins sequential elimination procedure. *Sequential Anal.* **18**, 81-105.

Leu, C. S. and Levin, B. (2004). Selecting the best subset of b out of c coins with the Levin-Robbins sequential elimination procedure: Proof of the lower bound formula for the probability of correct selection in the case $b = 2$, $c = 4$. Technical Report #B-91, Department of Biostatistics, Columbia University.

Levin, B. and Leu, C. S. (2004). Formulas for the exact probability of correct selection in the binomial Levin-Robbins sequential selection procedure in the cases $b = 2$, $c = 3$ and $b = 2$, $c = 4$ for $r = 1$. Technical Report #B-92, Department of Biostatistics, Columbia University.

Levin, B. and Leu, C. S. (2007). A comparison of two procedures to select the best binomial population with sequential elimination of inferior populations. *J. Statist. Plann. Inference* **137**, 245-263.

Levin, B. and Robbins, H. (1981). Selecting the highest probability in binomial or multinomial trials. *Proc. Natl. Acad. Sci. USA* **78**, 4663-4666.

Paulson, E. (1994). Sequential procedures for selecting the best one of k Koopman-Darmois populations. *Sequential Anal.* **13**, 207-220.

Zybert, P. and Levin, B. (1987). Selecting the highest of three binomial probabilities. *Proc. Natl. Acad. Sci. USA* **84**, 8180-8184.

The HIV Center for the Clinical and Behavioral Studies, New York State Psychiatric Institute, 1051 Riverside Drive Unit 15, New York, NY 10032, U.S.A.

Department of Biostatistics, Columbia University, Mailman School of Public Health, 722 West 168th Street, Suite 345, New York, NY 10032, U.S.A.

E-mail: CL94@columbia.edu

Department of Biostatistics, Columbia University, Mailman School of Public Health, 722 West 168th Street, Suite 626a, New York, NY 10032, U.S.A.

E-mail: Bruce.Levin@columbia.edu

(Received April 2005; accepted April 2006)