Published in final edited form as:

Seq Anal. 2013 ; 32(4): . doi:10.1080/07474946.2013.843321.

On an Inequality That Implies the Lower Bound Formula for the Probability of Correct Selection in the Levin-Robbins-Leu Family of Sequential Binomial Subset Selection Procedures

Bruce Levin and Cheng-Shiun Leu

Department of Biostatistics, Columbia University, New York, New York, USA

Abstract

We study a key inequality that implies the lower bound formula for the probability of correct selection and other selection-related events of interest in the Levin-Robbins-Leu family of sequential binomial subset selection procedures. We present a strategy for the proof of the key inequality, and a mostly-complete general proof is given. The strategy provides an entirely complete and rigorous proof of the inequality for as many as seven competing populations using computer-assisted symbolic manipulation.

Keywords

Adaptive subset selection; Elimination and recruitment procedures; Lower-bound formulas; Probability of correct selection; Sequential selection; Standardized Muirhead ratios; Subset selection

1. PART I—INTRODUCTION

An intriguing inequality underlies an important property of the Levin-Robbins-Leu (LRL) family of sequential subset selection procedures for binomial populations. This paper presents a proof strategy and a mostly-complete general proof of that inequality. See Levin and Robbins (1981), Leu and Levin (2008a, 1999a, 1999b, 2008b), and Leu et al. (2011) for an overview of the LRL family of sequential subset selection procedures and their development.

The important property just mentioned is that each member of the family admits of a lower bound for the probability of correct subset selection which can be expressed as a simple, explicit function of the true success parameters of the binomial populations. We refer to this as a *lower bound formula*, or LBF, for the probability of correct selection, or $P[cs]$. As discussed in Leu and Levin (2008a), the LBF extends to the probability of several other selection-related events of interest, such as the probability of finding at least x of the y best populations among those in the selected subset. Such LBFs are useful for practical design purposes as discussed below as well as for theoretical purposes. For example, in their landmark monograph on sequential subset selection procedures, Bechhofer et al. (1968 at p. 337) noted the strategic importance such an explicit expression would have had for their own sequential subset selection procedure. Because we established a lower bound formula for the non-adaptive member of the LRL family, though, we were able to prove (for our non-adaptive procedure) the truth of the conjecture that Bechhofer et al. (1968) stated but

did not prove (for their non-adaptive procedure) asserting that the probability of selecting “acceptable” if not entirely “correct” subsets can be made arbitrarily large no matter what the true population parameters are. See Leu and Levin (2008a) for further discussion of these ideas.

Lower bound formulas are useful for a variety of reasons. Most subset selection procedures in the literature only provide a guarantee that the $P[cs]$ is at least P^* for some pre-specified $P^* < 1$ under certain assumptions about the true parameters of the populations under consideration, e.g., that they fall outside of an indifference zone. They are typically silent, however, about the probabilities of other selection-related events of interest like the ones mentioned in the preceding paragraph, and require simulation to determine those probabilities. The LBF yields results for these directly. When the number of competing populations is not too large (say, five or fewer), these lower bounds actually provide quite accurate numerical approximations to the true probabilities. With a LBF, moreover, one can often identify parameter values inside the indifference zone for which the $P[cs]$ nevertheless exceeds P^* . And for theoretical purposes, with a LBF in hand one does not need to identify the procedure’s least favorable configuration (LFC) of parameters. For example, in establishing that their sequential subset selection procedure satisfies $P[cs] \geq P^*$ for all parameter vectors outside of the indifference zone, Bechhofer et al. (1968) needed first to identify the LFC and prove that it minimized the $P[cs]$ over all parameter vectors in the preference zone (the complement of the indifference zone), after which they had to show that $P[cs] \geq P^*$ at the LFC. With a LBF, all one needs to do is identify the parameter vector which minimizes the lower bound formula itself, which is a considerably simpler task. We illustrate this below at equation (2.2). These remarks serve to motivate our efforts in this paper to give a formal derivation and proof of the LBF. We will not, however, give any further consideration to comparisons of the LRL family with other sequential subset selection procedures.

Since the introduction of the LRL family of subset selection procedures, an open problem has been to provide a general rigorous proof of the LBF for the adaptive members of the family. In addition to the non-adaptive member, which we call procedure $\mathcal{N}$, there is procedure $\mathcal{E}$, which allows for sequential elimination of apparently inferior candidate populations with mounting evidence during the course of sampling; procedure $\mathcal{R}$, which allows for sequential recruitment of apparently superior candidates; and procedure $\mathcal{E}/\mathcal{R}$, which allows for both sequential elimination and recruitment of apparently inferior and superior candidates, respectively. We define each of these procedures below. The advantages of an adaptive procedure should be clear. A procedure which eliminates inferior populations as the experiment progresses has great ethical appeal when the application is to find best treatments in the context of a randomized clinical trial. Elimination also results in certain efficiencies: it reduces the expected total number of subjects in the trial and increases the expected proportion of positive outcomes compared to a non-adaptive procedure. In an industrial context, a procedure which recruits superior products as the experiment progresses allows accelerated further development and production of these products even while the experiment continues to select the remaining subset members. The reader is referred to Leu and Levin (2008a and 2008b), Levin and Leu (2007), and Leu et al. (2011) for further discussion of the operating characteristics of the adaptive LRL members together with experimental evidence supporting our contention that the lower bound formula applies to these adaptive members too. Examples illustrating the efficiency and usefulness of these procedures are given in these references as well.

This paper has two parts. In the first we focus on the motivating background for a key algebraic inequality that is studied in detail in the second part. After defining the members of the LRL family and stating the lower bound formula in Section 2, we focus on procedure

$\mathcal{E}$ in Section 3, where we indicate how the key inequality implies the LBF for this procedure's $H[cs]$. Extensions to procedures $\mathcal{R}$ and $\mathcal{E}/\mathcal{R}$ are given at the end of the section. Readers who wish to skip these background sections may proceed directly to the second part, which is self-contained and starts with Section 4, wherein we outline a proof of the key inequality. Our approach relies on two results (Lemmas 4.2 and 4.3) which, while not yet proven in complete generality, do appear to be universally valid. In support of this assertion, we use computer-assisted symbolic algebraic manipulation to provide a complete and rigorous proof of the lemmas—and thus the key inequality—in all cases with up to seven candidate populations. These proofs extend the established LBF results for selecting the one best population from amongst any number, and for selecting the two best populations out of four, using the adaptive procedures discussed in Levin and Robbins (1981), Zybert and Levin (1987), and Leu and Levin (1999a, 1999b, 2004, and 2006). We also present some stronger new findings that reveal “why” the key inequality holds, and which point intriguingly toward a completely general proof.

2. THE LRL FAMILY OF SUBSET SELECTION PROCEDURES AND THEIR LOWER BOUND FORMULA

2.1. The Procedures

Suppose there are c populations and from each one we can observe a sequence of independent and identically distributed Bernoulli observations with success probabilities $p_1, \dots, p_c$. For convenience we shall think of the subset selection problem in this context as a contest between $c \geq 2$ coins, where the goal is to select a subset of size b containing the b best coins ($1 \leq b < c$), where “best” means having the greatest probabilities of success. Here we regard the subset size b as fixed in advance by design, and we assume we have sufficient justification from the context of the application for selecting a subset of such fixed size. This problem should not be confused with Gupta's subset selection problem (see, e.g., Gupta, 1956 and Gupta, 1965), in which the goal is to produce a subset of random size intended to contain the one truly best coin with pre-specified probability. For us, a “correct selection” occurs when the coins in the selected subset have the b truly greatest success probabilities.

Because we shall consider only symmetric procedures, i.e., procedures that are invariant under permutations of the labels of the coins, we assume henceforth that the coins are ordered from best to worst with success probabilities $p_1 \geq \dots \geq p_c$. This will not be known in practice, of course, but for theoretical purposes we may assume this ordering without loss of generality. To avoid trivialities, we also assume that $p_1 < 1$ and $p_c > 0$. Denote the odds on heads versus tails by $w_i = p_i/(1 - p_i) > 0$. The odds are similarly ordered from best to worst, $w_1 \geq \dots \geq w_c$, and it will be most convenient to work in the odds parametrization henceforth.

When coins are tossed in vector-at-a-time fashion, let $X^{(n)} = (X_1^{(n)}, X_2^{(n)}, \dots, X_c^{(n)})$ for $n = 1, 2, \dots$ be the vector that reports the cumulative number of successes observed for each coin after n tosses, and let $X^{[n]} = (X_1^{[n]}, X_2^{[n]}, \dots, X_c^{[n]})$ be the ordered $X^{(n)}$ vector with $X_1^{[n]} \geq X_2^{[n]} \geq \dots \geq X_c^{[n]}$. Procedures $\mathcal{N}$, $\mathcal{E}$, $\mathcal{R}$, and $\mathcal{E}/\mathcal{R}$, are defined as follows.

Procedure $\mathcal{N}$. Choose a positive integer r in advance. Begin tossing the coins vector-at-a-time, and record the success tallies $X^{(n)} = (X_1^{(n)}, X_2^{(n)}, \dots, X_c^{(n)})$. Stop the first time the coin (or coins) with the b^{th} largest tally exceeds the coin (or coins) with the $(b+1)^{\text{st}}$ largest tally by r successes. Declare the subset of b coins with the largest tallies as best and select them.

In symbols, we stop at time $N^{(\mathcal{N})} = N_{b,c,r}^{(\mathcal{N})} = \inf\{n \geq r: X_b^{[n]} - X_{b+1}^{[n]} = r\}$. We omit the proof that $N^{(\mathcal{N})}$ is a stopping time, i.e., that $P[N^{(\mathcal{N})} < \infty] = 1$, which is similar to the standard proof for Wald's sequential probability ratio test; see, e.g., Wald (1947). The procedure “always” (i.e., with probability one) identifies a well-defined subset of b coins with no ties.

Procedure $\mathcal{E}$. Choose a positive integer r in advance. Begin tossing the coins vector-at-a-time, and record the success tallies $X^{(n)} = (X_1^{(n)}, X_2^{(n)}, \dots, X_c^{(n)})$. *Eliminate* apparently inferior coins as soon as they fall r heads behind the coin or coins with the currently held b^{th} largest tally. Iterate with the remaining coins starting from their current tallies, until $c-b$ coins are eliminated. Select the remaining b coins as best.

By “eliminate” we mean that a coin is withdrawn from the competition with no further tosses and is classified as outside the set of b best coins. Note that we do not claim that the first-eliminated coin is equal to the worst coin, merely that it does not qualify to be selected as among the b best. In symbols, we pause at the *time of first elimination*,

$$N^{(\mathcal{E})} = N_{b,c,r}^{(\mathcal{E})} = \inf\{n \geq 1: X_b^{[n]} - X_c^{[n]} = r\},$$

and eliminate any coin (or coins) j with $X_j^{(n)} = X_c^{[n]}$. Upon elimination, replace the number of coins c by the number of coins remaining in the game, namely $c' = c - \#\{j: X_j^{(n)} = X_c^{[n]}\}$, and continue with the c' remaining coins at their previously held tallies until the next elimination time $N_{b,c',r}^{(\mathcal{E})}$, and so on. The procedure always identifies a well-defined subset of b coins with no ties. Strictly speaking we ought to include the specific set of coins in play in the notation for $N^{(\mathcal{E})}$ but here and below we have suppressed that for simplicity.

Procedure $\mathcal{R}$. Choose a positive integer r in advance. Begin tossing the coins vector-at-a-time, and record the success tallies $X^{(n)} = (X_1^{(n)}, X_2^{(n)}, \dots, X_c^{(n)})$. We now *recruit* apparently superior coins as soon as they pull r heads ahead of the coin or coins with the currently held $(b+1)^{\text{st}}$ largest tally. Continue with the remaining coins starting from their current tallies, and iterate until a total of b coins are recruited, selecting them as the b best.

By “recruit” we mean that a coin is withdrawn from the competition with no further tosses and is selected to be amongst the set of b best coins. There is no claim that the first-recruited coin is equal to the best coin, merely that it already qualifies to be selected as among the b best. In symbols, we pause at the *time of first recruitment*,

$$N^{(\mathcal{R})} = N_{b,c,r}^{(\mathcal{R})} = \inf\{n \geq 1: X_1^{[n]} - X_{b+1}^{[n]} = r\},$$

and recruit any coin (or coins) j with $X_j^{(n)} = X_1^{[n]}$. Upon recruitment, replace the number of coins c by the number of coins remaining in the game, namely $c' = c - \#\{j: X_j^{(n)} = X_1^{[n]}\}$, replace the number of coins b to be recruited by $b' = b - \#\{j: X_j^{(n)} = X_1^{[n]}\}$, and continue with the c' remaining coins at their previously held tallies until the next recruitment time, $N_{b',c',r}^{(\mathcal{R})}$, and so on. The procedure always identifies a well-defined subset of b coins with no ties.

Procedure $\mathcal{E}/\mathcal{R}$. Choose a positive integer r in advance. Begin tossing the coins vector-at-a-time, and record the success tallies $X^{(n)} = (X_1^{(n)}, X_2^{(n)}, \dots, X_c^{(n)})$. We now eliminate apparently inferior coins as soon as they fall r heads behind the coin or coins with the currently held b^{th} largest tally, or recruit apparently superior coins as soon as they pull r heads ahead of the coin or coins with the currently held $(b+1)^{\text{st}}$ largest tally, or both. Iterate with the remaining coins starting from their current tallies, until $c-b$ coins are eliminated and b coins are recruited. Select the recruited b coins as best.

In symbols, we pause at the *time of first elimination or recruitment*,

$$N^{(\mathcal{E}/\mathcal{R})} = N_{b,c,r}^{(\mathcal{E}/\mathcal{R})} = \inf\{n \geq 1: X_b^{[n]} - X_c^{[n]} = r\} \wedge \inf\{n \geq 1: X_1^{[n]} - X_{b+1}^{[n]} = r\},$$

and eliminate all coins i with $X_i^{(n)} = X_c^{[n]}$ if $X_b^{[n]} - X_c^{[n]} = r$ and/or recruit all coins i with $X_i^{(n)} = X_1^{[n]}$ if $X_1^{[n]} - X_{b+1}^{[n]} = r$. If fewer than b coins are recruited and/or fewer than $c-b$ coins are eliminated, the procedure continues, starting from the current tallies of the remaining

subset of coins, and iterating with $N_{b',c',r}^{(\mathcal{E}/\mathcal{R})}$, where

$c' = c - \#\{j: X_j^{(n)} = X_c^{[n]}\} - \#\{j: X_j^{(n)} = X_1^{[n]}\}$ replaces c and $b' = b - \#\{j: X_j^{(n)} = X_1^{[n]}\}$ replaces b . At any time in the iteration, let b represent generically b minus the total number of coins recruited to that time, and let c represent generically c minus the total number of coins either eliminated or recruited to that time. Stop whenever there is a simultaneous recruitment of b coins and elimination of $c-b$ coins, at which point a total of b coins will have been recruited and $c-b$ coins eliminated. Upon stopping we declare the subset of recruited coins as the b best. The procedure always identifies a well-defined subset of b coins with no ties.

In what follows it will be convenient to use the notation (b) to denote a generic b -tuple of integers, $(b) = (i_1, \dots, i_b)$ with $1 \leq i_1 < \dots < i_b \leq c$. For the given odds parameters $w_1 \dots w_c$ and any b -tuple (b) , we define the symbol $w_{(b)}^r$ to be $w_{(b)}^r = w_{i_1}^r \dots w_{i_b}^r$. Also let $(b)^*$ denote the b -tuple of coins selected by any of the above subset selection procedures. Given our ordering convention, a correct selection is the event $[(b)^* = (1, \dots, b) = (cs)]$.

2.2. Lower Bound Formula for the $P[cs]$ of Procedure $\mathcal{N}$

For any number of coins c , any fixed subset size b ($1 \leq b < c$), and any vector $w = (w_1, \dots, w_c)$ of true success odds with $w_1 \dots w_c$, the probability of correct selection with procedure $\mathcal{N}$ is bounded from below by

$$P_w[cs] \geq (w_1 \dots w_b)^r / \sum_{(b)} w_{(b)}^r, \quad (2.1)$$

where the sum in the denominator is taken over all $\binom{c}{b}$ possible b -tuples.

See Leu and Levin (2008a) for the proof of this lower bound formula and extensions of it for other selection-related events of interest. The LBF allows us to choose r in designing selection experiments under the indifference zone/preference zone approach of Bechhofer et al. (1968), as follows. Given a pre-specified odds ratio > 1 , and a pre-specified probability value $P^* < 1$ but arbitrarily close to 1, we shall require that $P_w[cs] \geq P^*$ whenever w falls into

the preference zone $\{w: w_b/w_{b+1} \geq g\}$. It is easy to see that the right-hand side of (2.1) is minimized for any w in the preference zone at $g \cdot (1, \dots, 1, \dots, 1)$ for any positive constant g . Thus for any w in the preference zone we have

$$P_w[cs] \geq \theta^{rb} / \sum_{i=0}^{b \wedge (c-b)} \binom{b}{i} \binom{c-b}{i} \theta^{r(b-i)}. \quad (2.2)$$

It follows that we can choose $r = r(b, c, \theta, P^*)$ such that the right hand side of (2.2) is at least P^* for any pre-specified $P^* < 1$, from which the $P[cs]$ guarantee follows.

Unfortunately, the method of proof used to establish (2.1) for procedure $\mathcal{N}$ does not work for the adaptive LRL procedures. We have proofs of (2.1) for various special cases, e.g., for selecting the single best coin ($b=1$) from among any number of coins $c \geq 1$ using the elimination procedure $\mathcal{E}$ (see Leu and Levin, 1999a) and for selecting $b=1, 2$, or 3 coins from up to $c=4$ coins using any of the adaptive procedures $\mathcal{E}$, $\mathcal{R}$, or $\mathcal{E}/\mathcal{R}$ (see Leu and Levin, 2004 and 2006). We next derive the LBF for the adaptive procedures in general. We first focus on procedure $\mathcal{E}$ and derive the key algebraic inequality required to prove the LBF for any $c \geq 2$ and any subset size $b < c$. We then extend the derivation to procedures $\mathcal{R}$ and $\mathcal{E}/\mathcal{R}$.

3. DERIVATION OF THE LOWER BOUND GIVEN A KEY ALGEBRAIC INEQUALITY

3.1. The Conditional Form of the LBF

The LBF stated in (2.1) is actually a special case of a much stronger assertion. To state it, we define the *configuration of tallies at time n* to be the c -vector

$(X_1^{(n)} - X_c^{[n]}, \dots, X_{c-1}^{(n)} - X_c^{[n]}, X_c^{[n]} - X_c^{[n]})$ and the *ordered configuration of tallies at time n* to be the c -vector $(X_1^{[n]} - X_c^{[n]}, \dots, X_{c-1}^{[n]} - X_c^{[n]}, 0)$. For procedure $\mathcal{E}$ we define the *b -depth*

of any configuration of tallies to be the quantity $X_b^{[n]} - X_c^{[n]}$, i.e., the b^{th} component of the ordered configuration. Also for any ordered configuration a , let S_a denote the set of all *distinct* rearrangements of a . We refer to S_a as a *symmetric subset of configurations*. We claim the LBF applies to the following conditional probabilities of correct selection.

Theorem 3.1—For any ordered configuration d of b -depth s ($0 \leq s \leq r$), the conditional $P[cs]$, given that the sample path $X^{(n)} = (X_1^{(n)}, X_2^{(n)}, \dots, X_c^{(n)})$ visits a member of the symmetric subset of configurations S_d at the **first** time that $X^{(n)}$ reaches any configuration of b -depth s , is bounded from below by the LBF.

We write the assertion of Theorem 3.1 as

$$P_w[cs|S_d] \geq (w_1 \cdots w_b)^r / \sum_{(b)} w_{(b)}^r \quad (3.1)$$

and refer to it as the “conditional $P[cs]$ given S_d ”, so as to abbreviate the longer phrase that specifies the conditioning event of the statement. Inequality (2.1) is just the special case where the configuration is in the singleton symmetric subset of configurations $S_{(0, \dots, 0)}$ at b -depth 0 at time $n=0$ or, for that matter, at any other time $n>0$ if the random walk returns to the origin.

The proof of Theorem 3.1 proceeds by first demonstrating (3.1) for any ordered configuration of b -depth r , and then showing that that implies (3.1) holds for all ordered configurations of any b -depth. Note that the first assertion again implies the unconditional LBF since then,

$$P_w[cs] = \sum_a P_w[cs|S_a] P_w[S_a] \geq \frac{(w_1 \cdots w_b)^r}{\sum_{(b)} w_{(b)}^r} \left\{ \sum_a P_w[S_a] \right\} = \frac{(w_1 \cdots w_b)^r}{\sum_{(b)} w_{(b)}^r},$$

where the summation on the left is over all distinct ordered configurations of b -depth r , for which we use the particular symbol a . The terms $P_w[S_a]$ denote the probability of visiting a permutation of a the first time $X^{(n)}$ reaches b -depth r . We do require the full generality of (3.1), however, because in order to establish the first assertion we will need to use the LBF for configurations d of other b -depths in an inductive manner.

To establish the first assertion, we argue by induction on the number of coins c . The case $c=2$ with $b=1$ is trivial since (3.1) reduces to an exact equality given by the classic gambler's ruin problem, $P_w[cs|S_d] = w_1^r / (w_1^r + w_2^r) = P_w[cs]$. The conditioning is irrelevant here because any path starting at configuration $(0,0)$ must pass through the configuration $d=(s,0)$ or $(0,s)$ for any $0 < s < r$ before reaching the boundary configurations $(r,0)$ or $(0,r)$. Also, for the inductive argument, let us extend (3.1) to the degenerate case $b=c$, wherein the exact conditional $P[cs]$ given passage through any symmetric set of configurations and the LBF are clearly both equal to one.

So assume that we have proven (3.1) for up to $c-1$ coins with any $1 \leq b \leq c-2$, and now consider the case of selecting a subset of $b \leq c-1$ best coins from among c coins with labels in the set $C = \{1, \dots, c\}$. Let a be any ordered configuration of b -depth r , which represents (in ordered form) the configuration of a possible set of success tallies at the time of first elimination $N^{(\mathcal{E})}$ in procedure $\mathcal{E}$. Note that $a_b = r$ and if k coins, say, are eliminated at this time, then the last k components of a equal zero, where $1 \leq k \leq c-b$. We need to keep track of the various subsets of k coins that might be eliminated at the time of first elimination for the given ordered configuration a , and we will denote such subsets by K . Since there cannot be a correct selection if any of the first b coins are eliminated, in calculating the conditional probability of correct selection we need only consider subsets such that $K \subseteq C \setminus B$ with $|K| = k$, where we have written $B = \{1, \dots, b\}$. Then the conditional probability of correct selection given passage through S_a at the time of first elimination is

$$P_w[cs|S_a] = \frac{\sum_{u \in S_a} P_w[cs|u] P_w[u]}{P_w[S_a]} = \sum_{\substack{K \subseteq C \setminus B \\ |K|=k}} \sum_{\substack{u \in S_a \\ u_K=0}} \left\{ \frac{P_w[cs|u] P_w[u]}{\sum_{\substack{u \in S_a \\ u_K=0}} P_w[u]} \right\} \cdot \left\{ \frac{\sum_{\substack{u \in S_a \\ u_K=0}} P_w[u]}{\sum_{u \in S_a} P_w[u]} \right\}, \quad (3.2)$$

where $P_w[cs|u]$ denotes the conditional probability of correct selection given $X^{(n)} = u \in S_a$ at time $n = N^{(\mathcal{E})}$ and where, for a given subset K of eliminated coins, $u_K=0$ indicates that the components of configuration u are equal to zero in the positions specified by K . We recognize the term inside the first pair of braces,

$$\sum_{\substack{u \in S_a \\ u_K=0}} \frac{P_w[cs|u]P_w[u]}{\sum_{\substack{u \in S_a \\ u_K=0}} P_w[u]} = P_{w'}[cs|S_{a'}], \quad (3.3)$$

as the conditional $P[cs]$ in the *reduced* game which continues after the k coins in K have been eliminated, where $w = w(C \setminus K)$ denotes the corresponding reduced vector of odds of dimension $c-k$ and $a = a(C \setminus K)$ denotes the reduced ordered configuration of the remaining success tallies. For example, consider running procedure $\mathcal{E}$ with $b=2$, $c=4$, and $r=3$, and suppose at time $n = N^{\mathcal{E}}$ we observe $X^{(n)} = (11, 12, 8, 9)$. These tallies correspond to configuration $u = (3, 4, 0, 1)$ which is a permutation $u \in S_a$ of the ordered configuration $a = (4, 3, 1, 0)$. In this case $k=1$, $K=\{3\}$, and the reduced ordered configuration is $a' = (3, 2, 0)$ of b -depth two in the reduced game with $c=3$ coins in the set $C' = C \setminus K = \{1, 2, 4\}$, having eliminated coin 3, with odds $w = (w_1, w_2, w_4)$. The sum in (3.2) also requires consideration of $K=\{4\}$ in which case $C' = C \setminus K = \{1, 2, 3\}$ when coin 4 is eliminated, with $a' = (3, 2, 0)$ again and odds $w = (w_1, w_2, w_3)$.

The key idea here is that because we consider *all* possible permutations of a for a given set of eliminated coins in (3.3), we automatically condition in the reduced game on the sample path visiting *any* of the possible permutations of the reduced ordered configuration a at b -depth $s=r-a_{c-k}$. By the inductive hypothesis, the conditional $P[cs]$ in (3.3) is at least as large as its LBF in the reduced game, so

$$P_w[cs|S_a] \geq \sum_{\substack{k \subseteq C \setminus B \\ |K|=k}} \frac{(w_1 \cdots w_b)^r}{\sum_{\substack{(b) \subseteq C \setminus K}} w_{(b)}^r} \cdot \frac{\sum_{\substack{u \in S_a \\ u_K=0}} P_w[u]}{\sum_{u \in S_a} P_w[u]}. \quad (3.4)$$

and we are trying to show that the right-hand side of (3.4) is no less than

$$(w_1 \cdots w_b)^r / \sum_{(b) \subseteq C} w_{(b)}^r. \text{ Thus it would suffice to show that}$$

$$\sum_{\substack{k \subseteq C \setminus B \\ |K|=k}} \frac{\sum_{\substack{u \in S_a \\ u_K=0}} P_w[u]}{\sum_{\substack{(b) \subseteq C \setminus K}} w_{(b)}^r} \geq \frac{\sum_{u \in S_a} P_w[u]}{\sum_{(b) \subseteq C} w_{(b)}^r}. \quad (3.5)$$

Note that we need not be concerned that procedure $\mathcal{E}$ terminates at the time of first elimination in the case $b = c-1$. As mentioned above, the conditional $P[cs]$ in (3.3) equals unity and so does the LBF for the reduced game of selecting $b=c-1$ coins out of $c-1$ coins, so (3.4) is still valid and (3.5) still suffices. Similarly, we need not be concerned that the procedure terminates at time $N^{\mathcal{E}}$ in the case $k = c-b$. The conditional $P[cs]$ in (3.3) again equals unity, and the LBF in the reduced game also equals unity since

$\sum_{(b) \subseteq B=C \setminus K} w' (C \setminus K)_{(b)}^r = (w_1 \cdots w_b)^r$. So in this case too, (3.4) is still valid and (3.5) still suffices.

Some simplifications of (3.5) will be helpful. In Leu and Levin (1999a) we showed by Wald's change of measure argument applied at the time of first elimination that if $u \in S_a$ is a permutation of a , then $P_w[u]/w^u = P_w[a]/w^a$, where we have used the abbreviation $w^u = w_1^{u_1} \cdots w_c^{u_c}$. Thus dividing both sides of (3.5) by $P_w[a]/w^a$ we have the equivalent inequality

$$\sum_{\substack{k \subseteq C \setminus B \\ |K|=k}} \frac{\sum_{\substack{u \in S_a \\ u_K=0}} w^u}{\sum_{(b) \subseteq C \setminus K} w_{(b)}^r} \geq \frac{\sum_{u \in S_a} w^u}{\sum_{(b) \subseteq C} w_{(b)}^r}. \quad (3.6)$$

We prefer to keep the exponents fixed and permute the order of the w_j so to that end we introduce the following notation. Let $\wp_C$ denote the permutation group over C under composition of permutations from right to left. If $D \subset C$ is any proper subset of C , let $\wp_D$ denote the subgroup of $\wp_C$ permuting elements of D but leaving elements of $C \setminus D$ fixed. If $\pi = (\pi(1), \dots, \pi(c)) \in \wp_C$ is a permutation of $\{1, \dots, c\}$, we think of π as acting on the vector w by making the i^{th} component of the result $w(\pi(i))$ and we write the resulting vector $w^\pi = (w_{\pi(1)}, \dots, w_{\pi(c)})$. For $\pi \in \wp_C$, we also define the symbol $w_\pi^a = w_{\pi(1)}^{a_1} \cdots w_{\pi(c)}^{a_c}$. For example, if $c=3$, $a=(4,2,0)$, and $\pi = (1,3,2)$ and $\pi = (3,2,1)$ are two permutations in $\wp_C$, then $w^\pi = (w_3, w_2, w_1)$, $w^\pi = (w_3, w_1, w_2)$, and with $\pi = \text{id}$, $w_\pi^a = w_3^4 w_1^2$. With this notation, then, it is easy to see that if u is a permutation of a , say $u = (a)$ where $\pi \in \wp_C$, then

$w^u = w^{\pi(a)} = w_{\pi^{-1}}^a$. Consequently we can re-express $\sum_{u \in S_a} w^u$ as

$$\sum_{u \in S_a} w^u = \sum_{\pi \in \wp_C} w_{\pi^{-1}}^a / \{m(a)k!\} = \sum_{\pi^{-1} \in \wp_C} w_{\pi^{-1}}^a / \{m(a)k!\} = \sum_{\pi \in \wp_C} w_\pi^a / \{m(a)k!\}, \quad (3.7)$$

where we have used the fact that $\wp_C$ is a group to replace sums over all permutations with sums over all permutation inverses and then changed variables back to permutations. The divisor $m(a)$ is the product of the factorials of the frequencies of the distinct non-zero components of a , which we use to adjust the enumeration over all permutations to agree with the enumeration of all distinct arrangements of a in S_a . The final term $k!$ similarly adjusts for the k -fold multiplicity of the zero components of a .

We wish to apply a similar transformation to the numerators on the left-hand side of (3.6). To that end, let π_K be the permutation that sends K into $Z = \{c-k+1, \dots, c\}$ by transposing the q^{th} element of $K \setminus Z$ with the q^{th} element of $Z \setminus K$ for $q=1, \dots, k-|K \setminus Z|$, with π_K equal to the identity permutation if $K=Z$. Then we can generate all rearrangements $u \in S_a$ for which $u_K = 0$ by applying σ_K^{-1} to the subscripts of a first, sending the last k zero components of a to positions in K , and then following by a permutation $\pi \in \wp_{C \setminus K}$ which leaves the positions with zeros fixed. It follows that

$$\sum_{\substack{u \in S_a \\ u_K=0}} w^u = \sum_{\tau \in \wp_{C \setminus K}} w^{\tau \sigma_K^{-1}(a)} / \{m(a)\} = \sum_{\tau \in \wp_{C \setminus K}} w_{\sigma_K \tau^{-1}}^a / \{m(a)\} = \sum_{\tau^{-1} \in \wp_{C \setminus K}} w_{\sigma_K \tau^{-1}}^a / \{m(a)\} = \sum_{\pi \in \sigma_K \wp_{C \setminus K}} w_{\pi}^a / \{m(a)\}. \quad (3.8)$$

No adjustment by $k!$ appears here because the subgroup $\wp_{C \setminus K}$ already leaves the positions in K fixed. The final summation is over all permutations in left cosets of the form $\sigma_K \wp_{C \setminus K}$. As specified above, permutations in such cosets are understood to apply an element of $\wp_{C \setminus K}$ first (which leaves elements in K fixed), followed by σ_K when permuting subscripts $1, \dots, c$ of w .

Finally, we re-express the denominators of the lower bound formulas in (3.6) similarly:

$$\sum_{(b)} w_{(b)}^r = \sum_{\pi \in \wp_C} w_{\pi}^{a^*} / \{b!(c-b)!\} \quad \text{and} \quad \sum_{(b) \subseteq C \setminus K} w_{(b)}^r = \sum_{\pi \in \sigma_K \wp_{C \setminus K}} w_{\pi}^{a^*} / \{b!(c-b-k)!\} \quad (3.9)$$

where a^* denotes the distinguished configuration $a^* = (r, \dots, r, 0, \dots, 0)$ with b components equal to r and $c-b$ components equal to 0. Therefore, upon substitution of (3.7), (3.8), and (3.9) into (3.6), we arrive at our **key algebraic inequality** which suffices to prove (3.1):

$$\sum_{\substack{K \subseteq C \setminus B \\ |K|=k}} \frac{\sum_{\pi \in \sigma_K \wp_{C \setminus K}} w_{\pi}^a}{\sum_{\pi \in \sigma_K \wp_{C \setminus K}} w_{\pi}^{a^*}} > \frac{\sum_{\pi \in \wp_C} w_{\pi}^a}{\sum_{\pi \in \wp_C} w_{\pi}^{a^*}} \binom{c-b}{k}. \quad (3.10)$$

To conclude the derivation we must demonstrate the second assertion made at the beginning of this section, namely, that if (3.1) holds for any ordered configuration a of b -depth r , then it holds for all ordered configurations d of any b -depth s . For this we use the same argument as was used in Leu and Levin (1999a), which we briefly reprise here. There we considered the case $b=1$, but the same development applies *mutatis mutandis* upon replacing “depth s ” by “ b -depth s ”. The development shows that for any s with $0 \leq s < r$ and any ordered configuration d of b -depth s , the conditional probability that procedure $\mathcal{E}$ will ultimately select *any* b -tuple (b) , given that the sample path visits S_d at the first time it reaches b -depth s , can be expressed as

$$P_w[(b)|S_d] = \sum_a W_{a,d} P[(b)|S_a].$$

The summation is over all ordered configurations a of b -depth r , the probabilities of selecting (b) on the right-hand side are conditional on visiting a permutation of a at the first time the sample path reaches b -depth r , and where $W_{a,d}$ are positive weights *that do not depend on* (b) . If now we assume that (3.1) holds for any configuration of b -depth r , we can write (3.1) as

$$\frac{P_w[cs|S_a]}{\sum_{(b) \neq (cs)} P_w[(b)|S_a]} \geq \frac{w_{(cs)}^r}{\sum_{(b) \neq (cs)} w_{(b)}^r}$$

with $(cs) = (1, \dots, b)$, or we can write (3.1) equivalently as

$$P_w[cs|S_a] \geq \sum_{(b) \neq (cs)} P_w[(b)|S_a] \cdot \frac{w_{(cs)}^r}{\sum_{(b) \neq (cs)} w_{(b)}^r}.$$

Multiply both sides by $W_{a,d}$, sum over all ordered configurations a of b -depth r , and interchange the order of summation on the right-hand side to find that

$$P_w[cs|S_d] \geq \sum_{(b) \neq (cs)} P_w[(b)|S_d] \cdot \frac{w_{(cs)}^r}{\sum_{(b) \neq (cs)} w_{(b)}^r},$$

which is equivalent to (3.1). So (3.1) holds for all ordered configurations of any b -depth, and since b was arbitrary, the inductive step is complete. This demonstrates the sufficiency of (3.10) to establish Theorem 3.1 and the LBF (3.1).

3.2. Extension to Procedures $\mathcal{R}$ and $\mathcal{E}/\mathcal{R}$

We now extend the derivation of the lower bound formula to the other two adaptive LRL subset selection procedures. As discussed in Leu and Levin (2008b), procedure $\mathcal{R}$ is isomorphic to procedure $\mathcal{E}$ under the *parity transformation* wherein binary outcomes are exchanged with their complements. That is, if we call a success a “failure” and a failure a “success”, then running procedure $\mathcal{R}$ to select the b best coins with sequential recruitment of superior coins is identical to running procedure $\mathcal{E}$ to select the $b = c-b$ “best” coins with sequential elimination of “inferior” coins, where the eliminated coins comprise the selected subset for the original recruitment procedure. The design constants and population parameters transform as follows. Subset size b in $\mathcal{R}$ becomes $b = c-b$ in the isomorphic $\mathcal{E}$ procedure; subset B becomes $B = C \setminus B$ and its complement $C \setminus B$ becomes $C \setminus B = B$. The parity transformation inverts the odds on “success” for each coin, so upon re-labelling the coins in reverse order, the transformed parameters are $w_j = 1/w_{c+1-j}$ with $w_1 \cdots w_c$. Note that the lower bound formula is invariant under the parity transformation, because

$$\frac{(w_1' \cdots w_b')^r}{\sum_{(b') \subseteq C} w_{(b')}^r} = \frac{(w_c^{-1} \cdots w_{b+1}^{-1})^r}{\sum_{(c-b) \subseteq C} w_{(c-b)}^{-r}} \cdot \frac{(w_1 \cdots w_c)^r}{(w_1 \cdots w_c)^r} = \frac{(w_1 \cdots w_b)^r}{\sum_{(b) \subseteq C} w_{(b)}^r}.$$

To derive the LBF for procedure $\mathcal{R}$ we will apply inequality (3.10) to the isomorphic $\mathcal{E}$ procedure. To set up for that, we note the following. An ordered configuration a of “ $\mathcal{R}$ -form” with $a_1 = a_{b+1} + r$ (corresponding to the success tallies at the time of first recruitment in procedure $\mathcal{R}$) becomes the ordered configuration $a = (a_1, \dots, a_j, \dots, 0)$ with $a_j = a_1 - a_{c+1-j}$. Note that $a_1 = a_1$, $a_c = 0$, and $a_b = a_1 - a_{c+1-(c-b)} = a_1 - a_{b+1} = r$, so a is of proper “ $\mathcal{E}$ -form”. Distinguished configuration a^* becomes $a^* = (r, \dots, r, 0, \dots, 0)$ with $b = c-b$ components equal to r and $c-b = b$ components equal to 0. Any subset $H \subseteq B$ of size $|H|=h$ ($h=1, \dots, b$) corresponds to a subset $K \subseteq C \setminus B$ of size $|K|=k=h$ and the transpositions H that move elements of H to positions $1, \dots, h$ correspond to the transpositions that move elements of K to $c, \dots, c-k+1$, respectively. We can thus apply inequality (3.10) for the

isomorphic $\mathcal{E}$ game with design constants a, b, c, B, C , and parameter w to deduce the following inequality, which implies the LBF (2.1) or (3.1) for procedure $\mathcal{R}$:

$$\sum_{\substack{H \subseteq B \\ |H|=h}} \frac{\sum_{\pi \in \sigma_H \wp_{C \setminus H}} w_\pi^a}{\sum_{\pi \in \sigma_H \wp_{C \setminus H}} w_\pi^{a^*}} > \frac{\sum_{\pi \in \wp_C} w_\pi^a}{\sum_{\pi \in \wp_C} w_\pi^{a^*}} \binom{b}{h}. \quad (3.11)$$

To go from (3.10) to (3.11) we used the fact that if $K \wp_C |K$ then

$w_\pi^{a'} = w_\pi^a \cdot (w_1 \cdots w_c)^{-a_1}$ for some unique $H \wp_C |H$; the constant $(w_1 \cdots w_c)^{-a_1}$ appears on both sides of the inequality and therefore cancels. Ditto for $(w_1 \cdots w_c)^{-r}$ in the denominators. The equivalent inequality corresponding to (3.6) is

$$\sum_{\substack{H \subseteq B \\ |H|=h}} \frac{\sum_{\substack{u \in S_a \\ u_H=a_1}} w^u}{\sum_{\substack{H \subseteq B \subseteq C \\ |H|=h}} w_{(b)}^r} > \frac{\sum_{u \in S_a} w^u}{\sum_{(b) \subseteq C} w_{(b)}^r}, \quad (3.12)$$

which follows from the analogue of (3.4) upon replacement of the conditional $P[cs]$ in the reduced game after recruitment of coins in H by the reduced-game LBF:

$$\frac{w_{B \setminus H}^r}{\sum_{(b-h) \subseteq C \setminus H} w_{(b-h)}^r} = \frac{w_{B \setminus H}^r}{\sum_{(b-h) \subseteq C \setminus H} w_{(b-h)}^r} \cdot \frac{w_H^r}{w_H^r} = \frac{(w_1 \cdots w_b)^r}{\sum_{H \subseteq (b) \subseteq C} w_{(b)}^r}.$$

Next consider procedure $\mathcal{E} / \mathcal{R}$. Any configuration a at the time of first elimination or recruitment or both will be either of $\mathcal{E}$ -form or $\mathcal{R}$ -form, or “ $\mathcal{E} / \mathcal{R}$ -form” with

$$a_1 = \cdots = a_h = a_{b+1} + r > a_{h+1} \geq \cdots \geq a_b = r \geq a_{b+1} \geq \cdots \geq a_{c-k} > a_{c-k+1} = \cdots = a_c = 0$$

for $1 \leq h \leq b$ and $1 \leq k \leq c-b$. We have already covered the case of $\mathcal{E}$ -form and $\mathcal{R}$ -form configurations, so let a be of $\mathcal{E} / \mathcal{R}$ -form. The analogue of (3.4) is now

$$P_w[cs|S_a] \geq \sum_{\substack{H \subseteq B \\ |H|=h}} \sum_{\substack{K \subseteq C \setminus B \\ |K|=k}} \frac{\sum_{\substack{u \in S_a \\ u_H=a_1, u_K=0}} P_w[u]}{\sum_{(b-h) \subseteq C \setminus (H \cup K)} w_{(b-h)}^r} \cdot \frac{\sum_{u \in S_a} P_w[u]}{\sum_{u \in S_a} P_w[u]}. \quad (3.13)$$

Multiply numerator and denominator of the LBF for the reduced game in the first factor by w_H^r to write it as

$$\frac{w_B^r}{w_H^r \sum_{(b-h) \subseteq C \setminus (H \cup K)} w_{(b-h)}^r} = \frac{w_B^r}{\sum_{H \subseteq (b)C \setminus K} w_{(b)}^r}.$$

Then the analogue of (3.6) is the inequality

$$\sum_{\substack{H \subseteq B \\ |H|=h}} \sum_{\substack{K \subseteq C \setminus B \\ |K|=k}} \frac{\sum_{\substack{u \in S_a \\ u_H=a_1, u_K=0}} w^u}{\sum_{H \subseteq (b)C \setminus K} w_{(b)}^r} \geq \frac{\sum_{u \in S_a} w^u}{\sum_{(b) \subseteq C} w_{(b)}^r}. \quad (3.14)$$

Now, for a given fixed subset H , write $b = b-h$, $c = c-h$, $B = B \setminus H$, $C = C \setminus H$, and let w be the c -vector $w_{C \setminus H}$ with components w_j for $j \in C \setminus H$, and similarly let a be the ordered configuration $a = (a_{h+1}, \dots, a_c)$ with $a_b = a_{h+b-h} = a_b = r$. (Note that our use of the prime notation here differs from that in the preceding discussion of procedure $\mathcal{R}$.) Then the inner summation on the left-hand side of (3.14) can be written

$$\begin{aligned} & \sum_{\substack{K \subseteq C \setminus B \\ |K|=k}} \frac{\sum_{\substack{u \in S_a \\ u_H=a_1, u_K=0}} w^u}{\sum_{H \subseteq (b)C \setminus K} w_{(b)}^r} \\ &= \sum_{\substack{K \subseteq C \setminus B \\ |K|=k}} \frac{w_H^{a_1} \sum_{\substack{u' \in S_{a'} \\ u'_K=0}} w'^{u'}}{w_H^r \sum_{(b') \subseteq C' \setminus K} w'^r_{(b')}} \quad (3.15) \\ &= w_H^{a_{b+1}} \sum_{\substack{K \subseteq C' \setminus B' \\ |K|=k}} \frac{\sum_{\substack{u' \in S_{a'} \\ u'_K=0}} w'^{u'}}{\sum_{(b') \subseteq C' \setminus K} w'^r_{(b')}} \geq w_H^{a_{b+1}} \frac{\sum_{u' \in S_{a'}} w'^{u'}}{\sum_{(b') \subseteq C'} w'^r_{(b')}} \end{aligned}$$

where the inequality is (3.6) applied to the case a, b, c, B, C, w . But upon multiplying numerator and denominator on the right-hand side of (3.15) by w_H^r , we find it equals

$$\frac{w_H^{a_1} \sum_{u' \in S_{a_1}} w'^{u'}}{\sum_{H \subseteq (b) \subseteq C} w_{(b)}^r} = \frac{\sum_{u \in S_a} w^u}{\sum_{H \subseteq (b) \subseteq C} w_{(b)}^r}.$$

Thus summing (3.15) over all $H \subseteq B$ with $|H|=h$ and applying (3.12) we obtain (3.14). The corresponding inequality in terms of permutation cosets is

$$\sum_{\substack{H \subseteq B \\ |H|=h}} \sum_{\substack{K \subseteq C \setminus B \\ |K|=k}} \frac{\sum_{\pi \in \sigma_H \sigma_K \mathcal{P}_{C \setminus (H \cup K)}} w_\pi^a}{\sum_{\pi \in \sigma_H \sigma_K \mathcal{P}_{C \setminus (H \cup K)}} w_\pi^{a^*}} \geq \frac{\sum_{\pi \in \mathcal{P}_C} w_\pi^a}{\sum_{\pi \in \mathcal{P}_C} w_\pi^{a^*}} \binom{b}{h} \binom{c-b}{k}. \quad (3.16)$$

This implies the LBF for procedure $\mathcal{E}/\mathcal{R}$.

Thus the LBF for each adaptive LRL procedure follows from inequality (3.10), to which we now turn. Note that it suffices to demonstrate (3.10) when each odds parameter is greater than or equal to one. For if $w = (w_1, \dots, w_c)$ is the original vector of success odds of arbitrary magnitude, let $\bar{w} = (\bar{w}_1, \dots, \bar{w}_c)$ denote the vector of scaled odds with $\bar{w}_j = w_j / w_0$ for any w_0 with $0 < w_0 \leq w_c$. Then if (3.10) holds when evaluated at $\bar{w}$ we can multiply both sides of the inequality by $w_0^{a_1 + \dots + a_c - br}$ to obtain (3.10) evaluated at w . Henceforth we assume that $w_1 \leq \dots \leq w_c = 1$ without loss of generality.

4. PART II—ESTABLISHING THE KEY ALGEBRAIC INEQUALITY

We re-introduce the notation used in Part I for readers who may have skipped the motivating background presented there.

Let b, c , and r be integers with $1 \leq b < c-1$ and $r \geq 1$. Let $a = (a_1, \dots, a_c)$ satisfy

$$a_1 \geq \dots \geq a_b = r \geq a_{b+1} \geq \dots \geq a_c = 0. \quad (4.1)$$

Although interest for our application restricts the components of a to non-negative integers, here we allow them to be non-negative real numbers. Let $a^* = (r, \dots, r, 0, \dots, 0)$ with b components equal to r , and let $w = (w_1, \dots, w_c)$ be a vector of r real numbers satisfying $w_1 \leq \dots \leq w_c = 1$. To avoid trivialities we assume henceforth that the components of w are not all equal, i.e., $w_1 < w_c$. Let $B = \{1, \dots, b\}$ and $C = \{1, \dots, c\}$, and let $\mathcal{P}_C$ denote the permutation group over C under composition of permutations from right to left. If $D \subseteq C$ is any proper subset of C , let $\mathcal{P}_D$ denote the subgroup of $\mathcal{P}_C$ permuting elements of D but leaving elements of $C \setminus D$ fixed. If $\pi = (\pi(1), \dots, \pi(c)) \in \mathcal{P}_C$ is a permutation of $\{1, \dots, c\}$, we think of π as acting on the vector w by making the i^{th} component of the result $w_{\pi(i)}$ and we write the resulting vector $w^\pi = (w_{\pi(1)}, \dots, w_{\pi(c)})$. For $\pi \in \mathcal{P}_C$ we also define the symbol $w_\pi^a = w_{\pi(1)}^{a_1} \dots w_{\pi(c)}^{a_c}$. For example, if $c=3$, $a=(4,2,0)$, and $\pi = (1,3,2)$ and $\pi = (3,2,1)$ are two permutations in $\mathcal{P}_3$, then $w^\pi = (w_3, w_2, w_1)$, $w^\pi = (w_3, w_1, w_2)$, and with $\pi = \text{id}$, $w_\pi^a = w_3^4 w_1^2$.

We recognize w_π^a as elements of Muirhead's theorem—see, e.g., Muirhead (1903), Hardy et al. (1934), and Marshall et al. (2011)—which states in one form that for all $w_1 \cdots w_c \geq 1$,

$$\sum_{\pi \in \wp_C} w_\pi^a \geq \sum_{\pi \in \wp_C} w_\pi^{a'} \text{ if and only if } a \text{ weakly submajorizes } a', \text{ written } a \succ_w a', \text{ i.e.,}$$

$a_1 + \cdots + a_i \geq a'_1 + \cdots + a'_i$ for $i=1, \dots, c$. This paper presents an inequality involving

standardized Muirhead ratios, $\sum_{\pi \in \mathfrak{Z}} w_\pi^a / \sum_{\pi \in \mathfrak{Z}} w_\pi^{a^*}$ for $a \succ_w a^*$, when averaged over certain subsets $\mathfrak{Z} \subseteq \wp_C$.

Let $k \geq 1$ denote the number of zeros in a , such that $a_{c-k} > 0$ but $a_{c-k+1} = \cdots = a_c = 0$, and let

K denote any of the $\binom{c-b}{k}$ subsets of $C \setminus B = \{b+1, \dots, c\}$ with $|K| = k$ elements. Let π_K be the permutation that sends K into $Z = \{c-k+1, \dots, c\}$ by transposing the q^{th} element of $K \setminus Z$ with the q^{th} element of $Z \setminus K$ for $q=1, \dots, k-|K \cap Z|$, with π_K equal to the identity permutation if $K=Z$. We are interested in left cosets of the form $\pi_K \wp_{C \setminus K}$. As specified above, permutations in such cosets are understood to apply an element of $\wp_{C \setminus K}$ first (which leaves elements in K fixed), followed by π_K . Thus if $j \in K$ and $\pi \in \pi_K \wp_{C \setminus K}$, then w_j does not “appear” in the expression w_π^a , because j will be among the final k components of π so that the exponent of w_j in w_π^a will be zero.

In this part we study the following function:

$$F(a) = \sum_{\substack{K \subseteq C \setminus B \\ |K|=k}} \left\{ \frac{\sum_{\pi \in \sigma_K \wp_{C \setminus K}} w_\pi^a}{\sum_{\pi \in \sigma_K \wp_{C \setminus K}} w_\pi^{a^*}} - \frac{\sum_{\pi \in \wp_C} w_\pi^a}{\sum_{\pi \in \wp_C} w_\pi^{a^*}} \right\} = \left\{ \sum_{\substack{K \subseteq C \setminus B \\ |K|=k}} \frac{\sum_{\pi \in \sigma_K \wp_{C \setminus K}} w_\pi^a}{\sum_{\pi \in \sigma_K \wp_{C \setminus K}} w_\pi^{a^*}} \right\} - \frac{\sum_{\pi \in \wp_C} w_\pi^a}{\sum_{\pi \in \wp_C} w_\pi^{a^*}} \binom{c-b}{k}. \quad (4.2)$$

The assertion $F(a) > 0$ means that the average of standardized Muirhead ratios over permutation subsets of the form $\mathfrak{Z} = \pi_K \wp_{C \setminus K} \subseteq \wp_C$ as K varies over subsets of size k contained in $C \setminus B$ is bounded from below by the standardized Muirhead ratio over the full permutation group $\wp_C$.

Example 4.1. ($b=1, c=3$)—For given positive integer r , let $a = (r, s, 0)$, for integer s with $0 < s < r$, and $a^* = (r, 0, 0)$. There are two subsets of size $k=1$ corresponding to possible zero positions in $C \setminus B$, namely $K_1 = \{3\}$ and $K_2 = \{2\}$. Then π_{K_1} is the identity permutation and π_{K_2} is the transposition $(1, 3, 2)$, with respective $\wp_{C \setminus K}$ subgroups $\{(1, 2, 3), (2, 1, 3)\}$ and $\{(1, 2, 3), (3, 2, 1)\}$. Thus $\pi_{K_1} \wp_{C \setminus K_1} = \{(1, 2, 3), (2, 1, 3)\}$ and $\pi_{K_2} \wp_{C \setminus K_2} = \{(1, 3, 2), (3, 1, 2)\}$. Then for any $w_1 = w_2 = w_3 = 1$,

$$F(r, s, 0) = \frac{w_1^r w_2^s + w_2^r w_1^s}{w_1^r + w_2^r} + \frac{w_1^r w_3^s + w_3^r w_1^s}{w_1^r + w_3^r} - \frac{w_1^r w_2^s + w_1^r w_3^s + w_2^r w_1^s + w_2^r w_3^s + w_3^r w_1^s + w_3^r w_2^s}{w_1^r + w_2^r + w_3^r},$$

where we have cancelled a common factor of 2 from numerator and denominator in the last term.

Example 4.2. ($b=2, c=4$)—For given positive integer r , let $a = (q, r, s, 0)$, for integers q and s with $q < r < s > 0$. Now $a^* = (r, r, 0, 0)$. The overall permutation group $\wp_C$ has $4! = 24$

elements. Again choose $k=1$ with two subsets corresponding to possible zero positions in $C \setminus B$, namely $K_1=\{4\}$ and $K_2=\{3\}$. Then K_1 is the identity permutation and $K_2=(1,2,4,3)$. The respective $\wp_{C \setminus K}$ subgroups are $\wp_{C \setminus K_1} = \{(1,2,3,4), (1,3,2,4), (2,1,3,4), (2,3,1,4), (3,1,2,4), (3,2,1,4)\} = K_1 \wp_{C \setminus K_1}$ and $\wp_{C \setminus K_2} = \{(1,2,3,4), (1,4,3,2), (2,1,3,4), (2,4,3,1), (4,1,3,2), (4,2,3,1)\}$, so that $K_2 \wp_{C \setminus K_2} = \{(1,2,4,3), (1,4,2,3), (2,1,4,3), (2,4,1,3), (4,1,2,3), (4,2,1,3)\}$. Then for any $w_1 \ w_2 \ w_3 \ w_4 \ 1$, (4.2) becomes, after dividing out common factors of $\binom{c-b}{k} / 4 = \frac{1}{2}$ and multiplying through by 2,

$$2 \cdot F(q, r, s, 0) = \left\{ \frac{w_1^q w_2^r w_3^s + w_1^q w_3^r w_2^s + w_2^q w_1^r w_3^s + w_2^q w_3^r w_1^s + w_3^q w_1^r w_2^s + w_3^q w_2^r w_1^s}{w_1^r w_2^r + w_1^r w_3^r + w_2^r w_3^r} + \left\{ \frac{w_1^q w_2^r w_4^s + w_1^q w_4^r w_2^s + w_2^q w_1^r w_4^s + w_2^q w_4^r w_1^s + w_4^q w_1^r w_2^s + w_4^q w_2^r w_1^s}{w_1^r w_2^r + w_1^r w_4^r + w_2^r w_4^r} \right\} \right. \\ \left. - \frac{\begin{pmatrix} w_1^q w_2^r w_3^s + w_1^q w_3^r w_2^s + w_1^q w_2^r w_4^s + w_1^q w_4^r w_2^s + w_1^q w_3^r w_4^s + w_1^q w_4^r w_3^s \\ + w_2^q w_1^r w_3^s + w_2^q w_3^r w_1^s + w_2^q w_1^r w_4^s + w_2^q w_4^r w_1^s + w_2^q w_3^r w_4^s + w_2^q w_4^r w_3^s \\ + w_3^q w_1^r w_2^s + w_3^q w_2^r w_1^s + w_3^q w_1^r w_4^s + w_3^q w_4^r w_1^s + w_3^q w_2^r w_4^s + w_3^q w_4^r w_2^s \\ + w_4^q w_1^r w_2^s + w_4^q w_2^r w_1^s + w_4^q w_1^r w_3^s + w_4^q w_3^r w_1^s + w_4^q w_2^r w_3^s + w_4^q w_3^r w_2^s \end{pmatrix}}{w_1^r w_2^r + w_1^r w_3^r + w_1^r w_4^r + w_2^r w_3^r + w_2^r w_4^r + w_3^r w_4^r} \right\}.$$

We now state the main inequality we wish to prove:

Theorem 4.1—For any $w_1 \cdots w_c \ 1$ not all equal, and k with $1 \leq k \leq c-b$, and any configuration a subject to (4.1) with $a \geq a^*$, we have $F(a) > 0$. Thus the average of standardized Muirhead ratios over subsets $K \subseteq C \setminus B$ for $K \subseteq C \setminus B$ with $|K|=k$ is bounded from below by the standardized Muirhead ratio over $\wp_C$, i.e.,

$$\binom{c-b}{k}^{-1} \left\{ \sum_{\substack{K \subseteq C \setminus B \\ |K|=k}} \frac{\sum_{\pi \in \sigma_K \wp_{C \setminus K}} w_\pi^a}{\sum_{\pi \in \sigma_K \wp_{C \setminus K}} w_\pi^{a^*}} \right\} > \frac{\sum_{\pi \in \wp_C} w_\pi^a}{\sum_{\pi \in \wp_C} w_\pi^{a^*}}.$$

Before beginning the formal proof we discuss some preliminaries. For any i in the set $I = \{1, \dots, b-1\} \cup \{b+1, \dots, c-k\}$ and $j = 1, \dots, c$, define

$$F_j^{(i)}(a) = \left\{ \sum_{\substack{K \subseteq C \setminus B \\ |K|=k}} \frac{\sum_{\substack{\pi \in \sigma_K \wp_{C \setminus K} \\ \pi(i)=j}} w_\pi^a}{\sum_{\pi \in \sigma_K \wp_{C \setminus K}} w_\pi^{a^*}} \right\} - \frac{\sum_{\pi \in \wp_C} w_\pi^a}{\sum_{\pi \in \wp_C} w_\pi^{a^*}} \binom{c-b}{k}, \quad (4.3)$$

and for $h = 1, \dots, c$, define the cumulative sum functions,

$$S_h^{(i)}(a) = \sum_{j=1}^h F_j^{(i)}(a).$$

We note the following properties of $F_j^{(i)}(a)$ and $S_h^{(i)}(a)$.

- A. $F(a) = S_c^{(i)}(a) = \sum_{j=1}^c F_j^{(i)}(a)$ for any $i \in I$.
- B. Factoring out the common term $w_j^{a_i}$, the remaining factor $F_j^{(i)}(a)/w_j^{a_i}$ does not depend on a_i .
- C. If for any $i, i' \in I$ $a_i = a_{i'}$ then $F_j^{(i)}(a) = F_j^{(i')}(a)$ for $j=1, \dots, c$. Thus marking positions i and i' with an x and fixing components of a in other positions, the functions $F_j^{(i)}(\dots x \dots x \dots)$ and $F_j^{(i')}(\dots x \dots x \dots)$ are identical functions of x for each $j=1, \dots, c$. Likewise, the functions $S_h^{(i)}(\dots x \dots x \dots)$ and $S_h^{(i')}(\dots x \dots x \dots)$ are identical functions of x for each $h=1, \dots, c$.

Remark 4.1—For each $i \in I$, $F_j^{(i)}(a)$ groups together all terms in (4.2) with the common factor $w_j^{a_i}$. Note that in the definition of $F_j^{(i)}(a)$ for $j > b$, if any subset K contains j , then the set of permutations $\sigma \in \mathcal{P}_{C \setminus K}$ such that $\sigma(i) = j$ is empty because, as mentioned above, any $j \in K$ would appear in $\sigma = (\dots (1), \dots, (c))$ in one of the positions $c-k+1, \dots, c$ and would thus not be available as a subscript of w with exponent a_i in w_π^a since $i \leq c-k$. Thus subsets K which contain j do not contribute to the term in braces in (4.3), which might even be written as a summation over $K \subseteq C \setminus B \setminus \{j\}$ with $|K|=k$. Consequently even if $F_j^{(i)}(a)$ were positive for $j \leq b$, $F_j^{(i)}(a)$ may be negative for some (though not necessarily all) $j > b$. To demonstrate positivity of $F(a)$, then, we need to work with the cumulative sums $S_h^{(i)}(a)$.

Definition 4.1—A sequence of real numbers $\{m_1, \dots, m_h\}$ has the *Monotone Sign Property* (MSP) if for some index μ ($1 \leq \mu \leq h$), $m_j > 0$ for $j \leq \mu$ and, if $\mu < h$, $m_j < 0$ for $j > \mu$. Thus a sequence with the MSP begins with one or more strictly positive terms, possibly followed only by strictly negative terms, i.e., with an unambiguous sign pattern $+\dots+ - \dots -$.

Lemma 4.1—Let $m_1, \dots, m_h$ be any set of h ordered real numbers with $m_1 > 0$ and let $\{m_1, \dots, m_h\}$ have the MSP. Let μ denote the index of the last positive member of the sequence and let μ denote the index of the last positive member of the m sequence, i.e., $\mu = \max\{j : m_j > 0\}$ and $\mu = \max\{j : m_j > 0\}$. Set $d = \min(\mu, h)$.

Then

$$\text{i.} \quad \sum_{j=1}^h \ell_j m_j \geq \left(\sum_{j=1}^{\lambda} m_j \right) \ell_d. \quad (4.4)$$

ii.

If, furthermore, $\sum_{j=1}^h m_j \geq 0$ and $\mu > \mu_h$ i.e., not all of the $\mu_1, \dots, \mu_h$ are equal, then the left-hand side of (4.4) is strictly positive, $\sum_{j=1}^h \ell_j m_j > 0$, even if the right-hand side of (4.4) equals zero.

Proof: By the assumptions, $\mu_1 \leq \mu_h$ and $1 \leq \mu \leq h$. If $\mu = \mu_h$, then with only non-zero μ_j 's,

$$\sum_{j=1}^h \ell_j m_j = \sum_{j=1}^{\lambda} \ell_j m_j \geq \left(\sum_{j=1}^{\lambda} m_j \right) \ell_{\lambda} > 0$$

where the last inequality is strict because $m_1, \dots, m_{\lambda}$ are all positive. If $\mu > \mu_h$, then since the m_j 's are positive through m_{μ} and negative thereafter,

$$\sum_{j=1}^{\mu} \ell_j m_j \geq \left(\sum_{j=1}^{\mu} m_j \right) \ell_{\mu} \text{ and } \sum_{j=\mu+1}^h \ell_j m_j = \sum_{j=\mu+1}^{\lambda} \ell_j m_j \geq \left(\sum_{j=\mu+1}^{\lambda} m_j \right) \ell_{\mu+1} \geq \left(\sum_{j=\mu+1}^{\lambda} m_j \right) \ell_{\mu}. \quad (4.5)$$

The last inequality in (4.5) holds because $\mu > \mu_{\mu+1} > 0$ since $\mu+1 \leq h$ and because the sum on the right is strictly negative. Adding the two inequalities in (4.5) yields

$$\sum_{j=1}^h \ell_j m_j \geq \left(\sum_{j=1}^{\lambda} m_j \right) \ell_{\mu}. \text{ Combining the two cases } \mu \leq \mu_h \text{ and } \mu > \mu_h, \text{ we have (4.4). For}$$

assertion (ii), if $\sum_{j=1}^h m_j \geq 0$, then $\sum_{j=1}^{h'} m_j > 0$ for all $h' < h$ by the MSP. (For if not, and

$$\sum_{j=1}^{h'} m_j \leq 0 \text{ for some } h' < h, \text{ then } \mu < h' < h, \text{ so } m_{h'+1} < 0, \dots, m_h < 0, \text{ in which case}$$

$$\sum_{j=1}^h m_j = \sum_{j=1}^{h'} m_j + \sum_{j=h'+1}^h m_j < 0, \text{ contradiction.) Therefore, the only possible case in which}$$

$\left(\sum_{j=1}^{\lambda} m_j \right) \ell_{\mu}$ might not be strictly greater than zero would be $\mu = h > \mu_h$ with $\sum_{j=1}^h m_j = 0$. But in that case, at least one of the inequalities in (4.5) must be strict, because at least two of the μ_j , all of which are positive, must differ by assumption. Hence even in this case the left-hand side of (4.4) is strictly positive. This proves Lemma 4.1.

The next lemma concerns the behavior of $F_j^{(i)}(a)$ and $S_h^{(i)}(a)$ ($j, h = 1, \dots, c$) for a in a neighborhood of a^* . To avoid any confusion about the number of zero components in the vector a as we let it vary as an argument of the functions we study, and because we wish to keep k fixed, we shall, for any continuous function $G(a)$, write $G(a^+) = \lim_{a \rightarrow a^*} G(a)$, where the limit is taken through any weakly submajorized sequence of a 's satisfying (4.1) such that $a_i = r$ for $i \in I$, $i < b$ and $a_i = 0$ for $i \in I$, $i > b$. We shall also use notation such as $a^+ = (r, \dots, r, 0^+, \dots, 0^+, 0, \dots, 0)$ with first b components equal to r , the next $c-b-k$ components positive but

arbitrarily close to zero, and the final k components exactly equal to zero. Note that by continuity, $G(a^+) = G(a^*)$.

Lemma 4.2—For each $i \in I$, the sequence $\{F_j^{(i)}(a^+), j=1, \dots, c\}$ possesses the Monotone Sign Property. In fact, for each $i \in I$, we have $F_j^{(i)}(a^+) > 0$ for $j=1, \dots, b$ and $F_j^{(i)}(a^+) < 0$ for $j=b+1, \dots, c$ where the inequalities are strict.

The proof of this lemma is technical and is discussed in Levin and Leu (2013), where its validity is demonstrated rigorously in *certain* cases for any number of coins c , and in *all* cases with $c \leq 7$ by computer-assisted symbolic manipulation. The time-complexity of the computational approach grows exponentially with c , so we stopped after $c=7$, which in certain cases already involved manipulating billions of terms. Though a completely general proof that the computational algorithm works for $c>7$ is an open problem, we conjecture in Levin and Leu (2013) that it can work in all cases.

Corollary 4.1— $S_h^{(i)}(a^+) > 0$ for $h = 1, \dots, c-1$.

This is because $S_c^{(i)}(a^+) = F(a^+) = F(a^*) = 0$ for any $i \in I$. As mentioned in the proof of Lemma 4.1, the MSP implies that $S_h^{(i)}(a^+) > 0$ for $h = 1, \dots, c-1$.

Corollary 4.2—For $h=1, \dots, c-1$, $\sum_{j=1}^h F_j^{(i)}(a^+) \log w_j \geq S_\lambda^{(i)}(a^+) \log w_d > 0$ where $\lambda = h$ and $d = \min(b, h)$; for $h=c$, $\sum_{j=1}^c F_j^{(i)}(a^+) \log w_j > 0$.

For $h < c$, apply (4.4) with $j = \log w_j$ and $m_j = F_j^{(i)}(a^+)$. For $h=c$, apply part (ii) of Lemma 4.1, since $\sum_{j=1}^c F_j^{(i)}(a^+) = S_c^{(i)}(a^+) = 0$ and we have assumed that not all of the $w_1, \dots, w_c$ are equal.

In particular, $F(a)$ has positive first and second partial derivatives with respect to a_i at a^+ since for any given $i \in I$,

$$\frac{\partial F(a^+)}{\partial a_i} = \sum_{j=1}^c \frac{\partial F_j^{(i)}(a^+)}{\partial a_i} = \sum_{j=1}^c F_j^{(i)}(a^+) \log w_j$$

and, using $j = (\log w_j)^2$,

$$\frac{\partial^2 F(a^+)}{\partial a_i^2} = \sum_{j=1}^c \frac{\partial^2 F_j^{(i)}(a^+)}{\partial a_i^2} = \sum_{j=1}^c F_j^{(i)}(a^+) (\log w_j)^2.$$

Remark 4.2—Corollary 4.2 shows that $F(a)$ is an increasing, convex function of each component a_i for $i \in I$ for all a in a sufficiently small neighborhood of a^* . This is because the inequalities in Lemma 4.2 are strict, so by continuity there is a neighborhood of a^+ satisfying

(4.1) in which strict inequalities still hold. Theorem 4.1 would follow if we knew that $F(a)$ were a convex function for all a . Alas, the MSP for $\{F_j^{(i)}(a): j=1, \dots, c\}$ does not hold globally for all a , so global convexity does not follow directly. Instead, we need a different approach to prove Theorem 4.1, by which certain convexity results will be established along the way. See also Remark 4.3 below.

Proof of Theorem 4.1—We proceed in several steps. First, we fix the arguments of $S_1^{(c-k)}$ in positions $1, \dots, b$ at value r and in positions $b+1, \dots, c-k$ at a common value x and demonstrate that the resulting function of x is positive and increasing. We call this “Step $c-k$, $h=1$.” We then extend this result inductively to the other cumulative sums $S_h^{(c-k)}$ for $h>1$ at the same arguments, calling this Step $c-k$, $h>1$. Next, we consider $S_1^{(c-k-1)}$ at arguments of the form $(r, \dots, r, x, \dots, x, a_{c-k}, 0, \dots, 0)$ for the given component a_{c-k} and deduce that the resulting function of x is also positive and increasing (Step $c-k-1$, $h=1$) and again extend the result to $S_h^{(c-k-1)}$ (Step $c-k-1$, $h>1$). We continue working toward the left until we have shown that for the given configuration components $a_{b+1}, \dots, a_{c-k}$ the cumulative sums $S_h^{(b+1)}(r, \dots, r, a_{b+1}, \dots, a_{c-k}, 0, \dots, 0)$ are all positive. We require a special argument for Step $b-1$, $h=1$ which considers $S_1^{(b-1)}(x, \dots, x, r, a_{b+1}, \dots, a_{c-k}, 0, \dots, 0)$ as a function of x , but then, with the same approach, we will be able to conclude that $S_h^{(1)}(a)>0$ for all $h=1, \dots, c$, the last case of which demonstrates $F(a)>0$. We now provide the specific arguments.

Step $c-k$, $h=1$ —For $i = b+1, \dots, c-k$, consider the functions

$$G_1^{(i)}(x) = S_1^{(i)}(r, \dots, r, x, \dots, x, 0, \dots, 0)$$

with identical argument x in components $b+1$ through $c-k$. As noted in Property (C) above, the functions $G_1^{(i)}(x)$ are identical functions of x , so we may just write $G_1(x)$, understanding that $G_1(x) = G_1^{(b+1)}(x) = \dots = G_1^{(c-k)}(x)$. Now $G_1(0^+) = S_1^{(i)}(a^+) = F_1^{(i)}(a^+) = F_1^{(i)}(a^*) > 0$ for each i by continuity and Lemma 4.2. It follows that $dG_1(0^+)/dx > 0$, because

$$\frac{dG_1(0^+)}{dx} = \sum_{i=b+1}^{c-k} \frac{\partial S_1^{(i)}(a^+)}{\partial a_i} = \sum_{i=b+1}^{c-k} \frac{\partial F_1^{(i)}(a^+)}{\partial a_i} = \sum_{i=b+1}^{c-k} F_1^{(i)}(a^+) \log w_1 = (c-b-k)(\log w_1) F_1^{(i)}(a^+) > 0$$

for any i with $i > b$. The first equality holds because, while $\frac{\partial S_1^{(i)}(a^+)}{\partial a_i}$ generally differ for different i , they agree when evaluated at $a_{b+1} = \dots = a_{c-k} = x$. Letting x approach zero establishes the identity. Thus there exists a $\delta_1 > 0$ such that $G_1(x) > G_1(0^+)$ for all $0 < x < \delta_1$. We claim that in fact $G_1(x) > G_1(0^+)$ for all x with $0 < x < r$. This we prove by contradiction. For suppose the set $M_1 = \{x < r: G_1(x) = G_1(0^+)\}$ is non-empty. Then $x_1^* = \inf M_1 < r$. Since $G_1(x) > G_1(0^+)$ for $x < x_1^*$ and G_1 is continuous, we have $G_1(x_1^*) = G_1(0^+)$. By Rolle's theorem there exists an x_1^{**} with $0 < x_1^{**} < x_1^*$ such that $dG_1(x_1^{**})/dx = 0$. For our contradiction we show that $dG_1(x_1^{**})/dx > 0$. Note that since $x_1^{**} < x_1^*$ and x_1^* is the greatest lower bound of M_1 , $G_1(x_1^{**}) > G_1(0^+) > 0$. Thus

$$\frac{dG_1(x_1^{**})}{dx} = (c-b-k)(\log w_1)G_1(x_1^{**}) > 0.$$

We conclude that $G_1(x) > G_1(0^+)$ for all $x \in (0, r]$.

Step c – k, part h—Next we argue by induction on h that the cumulative sum functions

$$G_h(x) = S_h^{(i)}(r, \dots, r, x, \dots, x, 0, \dots, 0)$$

are positive increasing functions of x for $0 < x \leq r$ for $h=1, \dots, c$ and $i=b+1, \dots, c-k$. The case $h=1$ was proved in the preceding part, so for $h>1$, assume we have that $G_h(x) > 0$ for $0 < x \leq r$ for all $h = 1, \dots, h-1$. Repeat the argument starting at a^+ . By Corollary 4.1, $G_h(0^+) > 0$ and by Corollary 4.2, for any $i \in I$ with $i > b$,

$$\frac{dG_h(0^+)}{dx} = \sum_{i=b+1}^{c-k} \frac{\partial S_h^{(i)}(a^+)}{\partial a_i} = \sum_{i=b+1}^{c-k} \sum_{j=1}^h F_j^{(i)}(a^+) \log w_j > 0.$$

As before, there exists a $\delta_h > 0$ such that $G_h(x) > 0$ for all $0 < x < \delta_h$. We claim $G_h(x) > 0$ for all $0 < x \leq r$. For if $M_h = \{x \in (0, \delta_h) : G_h(x) = G_h(0^+)\}$, then $x_h^{**} = \inf M_h < \delta_h$ and $G_h(x_h^{**}) = G_h(0^+)$, so there exists an x_h^{**} with $0 < x_h^{**} < \delta_h$ and $dG_h(x_h^{**})/dx = 0$ at which $G_h(x_h^{**}) > G_h(0^+) > 0$. We produce a contradiction by showing that $dG_h(x_h^{**})/dx > 0$. We again write

$$\frac{dG_h(x_h^{**})}{dx} = \sum_{i=b+1}^{c-k} \frac{\partial S_h^{(i)}(r, \dots, r, x_h^{**}, \dots, x_h^{**}, 0, \dots, 0)}{\partial a_i} = \sum_{i=b+1}^{c-k} \sum_{j=1}^h F_j^{(i)}(r, \dots, r, x_h^{**}, \dots, x_h^{**}, 0, \dots, 0) \log w_j \quad (4.6)$$

but note now that *we cannot rely on the MSP to show positivity* of (4.6), because the MSP may no longer hold at the vector argument $(r, \dots, r, x_h^{**}, \dots, x_h^{**}, 0, \dots, 0)$. Instead we use the positivity of the cumulative sums provided by the inductive hypothesis. For each $i \in I$, $i > b$ we have at that argument,

$$\begin{aligned} \sum_{j=1}^h F_j^{(i)}(\cdot) \log w_j &= S_1^{(i)}(\cdot) \log w_1 + \sum_{j=2}^h F_j^{(i)}(\cdot) \log w_j \\ &\geq S_1^{(i)}(\cdot) \log w_2 + \sum_{j=2}^h F_j^{(i)}(\cdot) \log w_j \\ &= S_2^{(i)}(\cdot) \log w_2 + \sum_{j=3}^h F_j^{(i)}(\cdot) \log w_j \\ &\vdots \\ &= S_{d-1}^{(i)}(\cdot) \log w_{d-1} + F_d^{(i)}(\cdot) \log w_d \\ &\geq S_d^{(i)}(\cdot) \log w_d = G_d(x_h^{**}) \log w_d \end{aligned} \quad (4.7)$$

where now $d = \min(i, h)$. If $d = h$ we have $G_d(x_h^{**}) = G_h(x_h^{**}) > 0$ as mentioned above (4.6), and if $d < h$ then $G_d(x_h^{**}) > 0$ by the inductive hypothesis. Therefore

$$\frac{dG_h(x_h^{**})}{dx} = (c-b-k)(\log w_d)G_d(x_h^{**}) > 0,$$

contradiction. We conclude that $G_h(x) = S_h^{(i)}(r, \dots, r, x, \dots, x, 0, \dots, 0) > 0$ for all $0 < x \leq r$, $h=1, \dots, c$, and $i=b+1, \dots, c-k$. In the sequel we shall repeatedly use the argument embodied in (4.7), which is enabled by the positivity of the cumulative sums $S_h^{(i)}$.

Step $c-k-1$, $h=1$ through Step $b+1$, $h>1$ —We now extend the result by fixing components starting on the right and working toward the left. For $q = c-k, \dots, b+2$, fix components $a_q, \dots, a_{c-k}$ at their values in the given configuration a , and for any $x \in [a_q, r]$, consider the functions

$$G_{h,q}(x) = S_h^{(i)}(r, \dots, r, x, \dots, x, a_q, \dots, a_{c-k}, 0, \dots, 0),$$

uniquely defined for any $i = b+1, \dots, q-1$ with identical arguments x in those positions. The preceding step established that for $q = c-k$,

$$G_{h,c-k}(a_{c-k}) = S_h^{(i)}(r, \dots, r, a_{c-k}, \dots, a_{c-k}, a_{c-k}, 0, \dots, 0) = G_h(a_{c-k}) > 0$$

for each $h=1, \dots, c$. This allows us to use the same argument for $G_{h,c-k}$ as in Step $c-k$, starting with $h=1$, using $x = a_{c-k}$ instead of 0^+ to begin the induction on h , and concluding finally that $G_{h,c-k}(x)$ is a positive, increasing function of x for $a_{c-k} \leq x \leq r$. The argument is then repeated for successively smaller values of $q = c-k-1, \dots, b+2$. At stage q , one begins with $x = a_q$ for which $G_{h,q}$ is positive by the preceding step. The key step in demonstrating that $dG_{h,q}(x)/dx > 0$ is to argue analogously to (4.6) and (4.7), relying on the positivity of the cumulative sums for preceding values of $h=1, \dots, h-1$ in

$G_{h',q}(x) = S_{h'}^{(i)}(r, \dots, r, x, \dots, x, a_q, \dots, a_{c-k}, 0, \dots, 0)$ for $a_q \leq x \leq r$. Thus we have shown that for an arbitrary set of components $a_{b+1}, \dots, a_{c-k}$ the cumulative sums

$S_h^{(b+1)}(r, \dots, r, a_{b+1}, \dots, a_{c-k}, 0, \dots, 0)$ are positive for $h=1, \dots, c$. Note that this is all that is required to prove Theorem 4.1 in the special case $b=1$. Thus we assume $b>1$ for the remainder of the proof.

Step $b-1$, $h=1$ —We continue working toward the left with the same approach. Now we redefine the functions

$$G_h(x) = S_h^{(i)}(x, \dots, x, r, a_{b+1}, \dots, a_{c-k}, 0, \dots, 0)$$

for any $x \leq r$, $i = 1, \dots, b-1$, and given $a_{b+1}, \dots, a_{c-k}$. We wish to assert that

$$G_1(r) = F_1^{(b-1)}(r, \dots, r, a_{b+1}, \dots, a_{c-k}, 0, \dots, 0) > 0, \quad (4.8)$$

but (4.8) requires proof, because unlike previous instances where setting $x = a_i$ yielded a value of a function with superscript i previously considered, we have not previously considered values of $i < b$.

To establish (4.8) we use a method similar to that used so far, writing for any i with $b+1 \leq i \leq c-k$,

$$F_1^{(b-1)}(a) = \sum_{j'=2}^c F_{1,j'}^{(b-1,i')}(a), \quad (4.9)$$

where for any $i, i' \in I$ with $i \leq i'$, any $1 \leq j \leq j' \leq c$, and any a , $F_{j,j'}^{(i,i')}(a)$ is defined as

$$F_{j,j'}^{(i,i')}(a) = \left\{ \sum_{\substack{K \subseteq C \setminus B \\ |K|=k}} \frac{\sum_{\substack{\pi \in \sigma_K \wp_{C \setminus K} \\ \pi(i)=j, \pi(i')=j'}} w_{\pi}^a}{\sum_{\pi \in \sigma_K \wp_{C \setminus K}} w_{\pi}^{a^*}} \right\} - \frac{\sum_{\substack{\pi \in \wp_C \\ \pi(i)=j, \pi(i')=j'}} w_{\pi}^a}{\sum_{\pi \in \wp_C} w_{\pi}^{a^*}} \binom{c-b}{k}. \quad (4.10)$$

Properties of $F_{j,j'}^{(i,i')}(a)$ are stated in the following Lemma 4.3. Again, the technical proof is discussed in Levin and Leu (2013), where Lemma 4.3 is demonstrated in all cases with up to $c=7$ coins by computer-assisted symbolic manipulation.

Lemma 4.3—For any $i=1, \dots, b-1$ and $i'=b+1, \dots, c-k$, the sequence

$$\{F_{1,j'}^{(i,i')}(a^+) : j'=2, \dots, c\} \text{ has positive cumulative sums, } S_{1,h'}^{(i,i')}(a^+) = \sum_{j'=2}^{h'} F_{1,j'}^{(i,i')}(a^+) > 0, \text{ for } h=2, \dots, c-1.$$

So for $i=b+1, \dots, c-k$, we consider the partial derivative of $F_1^{(b-1)}(a)$ at $a=a^+$ with respect to

$$a_i, \text{ which is } \frac{\partial F_1^{(b-1)}(a^+)}{\partial a_{i'}} = \sum_{j'=2}^c F_{1,j'}^{(b-1,i')}(a^+) \log w_{j'}.$$

By the argument used in (4.7) we have that $\frac{\partial F_1^{(b-1)}(a^+)}{\partial a_{i'}} \geq (\log w_d) \sum_{j'=2}^d F_{1,j'}^{(b-1,i')}(a^+)$ where $d = \min(i, c)$. If $d < c$ this expression is positive by Lemma 4.3, and if $d = c$ the expression is positive because

$$\sum_{j'=2}^c F_{1,j'}^{(b-1,i')}(a^+) = F_1^{(b-1)}(a^+) > 0 \quad \text{by Lemma 4.2. Then by the same approach used thus far,}$$

we conclude that $F_1^{(b-1)}(a) > 0$ for any argument of the form $a = (r, \dots, r, a_{b+1}, \dots, a_{c-k}, 0, \dots, 0)$.

Therefore $G_1(r) = F_1^{(b-1)}(a) > 0$ as asserted.

Resuming the main proof, we have

$$\frac{dG_1(r)}{dx} = (b-1)(\log w_1)F_1^{(b-1)}(r, \dots, r, a_{b+1}, \dots, a_{c-k}, 0, \dots, 0) > 0$$

and then $G_1(x) > 0$ for all $x \succ r$, using in the proof by contradiction the set $M_1 = \{x \succ r: G_1(x) \leq 0\}$.

Step $b-1$, $h > 1$ —By induction on h as in (4.6) and (4.7) we have $G_h(x) > 0$ for all $x \succ r$, again using proof by contradiction with the set $M_h = \{x \succ r: G_h(x) \leq 0\}$.

Step $b-1$, $h=1$ through Step 1, $h > 1$ —Finally, for $q = b-1, \dots, 2$ we consider the functions

$$G_{h,q}(x) = S_h^{(i)}(x, \dots, x, a_q, \dots, a_{b-1}, r, a_{b+1}, \dots, a_{c-k}, 0, \dots, 0)$$

with identical arguments $x \succ a_q$ in the first $q-1$ positions, uniquely defined for $i = 1, \dots, q-1$.

By the same arguments as above, we conclude that $S_h^{(1)}(a) > 0$ for all $h=1, \dots, c$ and all a unless $h=c$ and $a = a^+$, in which case $S_c^{(1)}(a^+) = F(a^+) = 0$. In particular, $F(a) > 0$ for all $a \succ a^*$.

5. DISCUSSION

It might seem that the monotone sign property of Lemma 4.2 is of fundamental importance, given that our argument begins there and establishes local convexity of $F(a)$ near a^+ . But the

MSP does not hold globally, and even the components $\{F_{1,j'}^{(i,i')}(a^+): j'=2, \dots, c\}$ of Lemma 4.3 do not have the MSP. As Lemma 4.3 and our approach to proving Theorem 4.1 suggest, however, it is the *positivity of cumulative sums* which is of fundamental importance. Indeed, above we demonstrated positivity of the cumulative sums $S_h^{(1)}(a) > 0$ for each h . In fact, we have observed empirically that the cumulative sums $S_h^{(i)}(a)$ are positive for every $i \in I$ (in fact, with $F_j^{(i)}(a) > 0$ for $j=1, \dots, b$), for each $h=1, \dots, c$, and every $a \succ w \succ a^*$. This observation itself provides a straightforward demonstration of Theorem 4.1, for then the argument at (4.7) shows that with $w_j > 1$ for $j \leq b$ and $w_j = 1$ for $j > b$ (1 $\leq j \leq c$),

$$\begin{aligned} \frac{\partial F(a)}{\partial a_i} &= \sum_{j=1}^{\lambda} F_j^{(i)}(a) \log w_j \\ &= S_1^{(i)}(a) \log w_1 + \sum_{j=2}^{\lambda} F_j^{(i)}(a) \log w_j \geq S_1^{(i)}(a) \log w_2 + \sum_{j=2}^{\lambda} F_j^{(i)}(a) \log w_j \\ &= S_2^{(i)}(a) \log w_2 + \sum_{j=3}^{\lambda} F_j^{(i)}(a) \log w_j \geq S_2^{(i)}(a) \log w_3 + \sum_{j=3}^{\lambda} F_j^{(i)}(a) \log w_j \\ &\quad \vdots \\ &= S_{\lambda-1}^{(i)}(a) \log w_{\lambda-1} + F_{\lambda}^{(i)}(a) \log w_{\lambda} \geq S_{\lambda-1}^{(i)}(a) \log w_{\lambda} + F_{\lambda}^{(i)}(a) \log w_{\lambda} \\ &= S_{\lambda}^{(i)}(a) \log w_{\lambda} \geq 0 \end{aligned}$$

with at least one strict inequality if $w_1 > w_c$. A similar argument applies to the second derivatives of $F(a)$, so that $F(a)$ is convex and increasing in each of its arguments, and since

$R(a^*)=0$ we conclude $R(a)>0$ for all $a \succ_W a^*$. Thus having all cumulative sums positive fully “explains why” $R(a)$ behaves as it does and is a promising strategy for a completely general proof.

Acknowledgments

We thank the editors and referees for their selfless efforts on our behalf.

References

- Bechhofer, RE.; Kiefer, J.; Sobel, M. Sequential Identification and Ranking Procedures. Chicago: University of Chicago Press; 1968.
- Gupta, SS. Mimeograph Series 150, Institute of Statistics. Chapel Hill: University of North Carolina; 1956. On a Decision Rule for a Problem in Ranking Means.
- Gupta SS. On Some Multiple Decision (Selection and Ranking) Rules. *Technometrics*. 1965; 7:225–245.
- Hardy, GH.; Littlewood, JE.; Pólya, G. Inequalities. Cambridge: Cambridge University Press; 1934.
- Leu, C-S.; Cheung, Y-K.; Levin, B. Subset Selection in Comparative Selection trials. In: Bhattacharjee, M.; Dhar, SK.; Subramanian, S., editors. Recent Advances in Biostatistics: False Discovery, Survival Analysis, and Other Topics. Vol. 4. London: World Scientific; 2011. p. 271–288.
- Leu CS, Levin B. On the Probability of Correct Selection in the Levin-Robbins Sequential Elimination Procedure. *Statistica Sinica*. 1999a; 9:879–91.
- Leu CS, Levin B. Proof of a Lower Bound Formula for the Expected Reward in the Levin-Robbins Sequential Elimination Procedure. *Sequential Analysis*. 1999b; 18:81–105.
- Leu, CS.; Levin, B. Technical Report #B-91. Department of Biostatistics, Columbia University; 2004 Jul 1. 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 $b = 2, c = 4$. available at <<http://biostats.bepress.com/columbiabiostat/>>
- Leu, CS.; Levin, B. Technical Report #B-98. Department of Biostatistics, Columbia University; 2006 Dec 15. Proof of the Lower Bound Formula for the Probability of Correct Binomial Subset Selection with the Levin-Robbins-Leu Sequential Elimination and Recruitment Procedure in the Case $b = 2, c = 4$. available at <<http://biostats.bepress.com/columbiabiostat/>>
- Leu CS, Levin B. On a Conjecture of Bechhofer, Kiefer, and Sobel for the Levin-Robbins-Leu Binomial Subset Selection Procedures. *Sequential Analysis*. 2008a; 27:106–125.
- Leu CS, Levin B. A Generalization of the Levin-Robbins Procedure for Binomial Subset Selection and Recruitment Problems. *Statistica Sinica*. 2008b; 18:203–218.
- Levin B, Leu CS. A Comparison of Two Procedures to Select the Best Binomial Population with Sequential Elimination of Inferior Populations. *Journal of Statistical Planning and Inference*. 2007; 137:245–263.
- Levin, B.; Leu, CS. Technical Report #B-148. Department of Biostatistics, Columbia University; 2013 Mar 18. On Two Lemmas Used to Establish a Key Inequality that Implies the Lower Bound Formula for the Probability of Correct Selection in the Levin-Robbins-Leu Family of Sequential Binomial Subset Selection Procedures. available at <<http://www.columbia.edu/~bl6/appendices/>>
- Levin B, Robbins H. Selecting the Highest Probability in Binomial or Multinomial Trials. *Proceedings of the National Academy of Sciences USA*. 1981; 78:4663–6.
- Marshall, AW.; Olkin, I.; Arnold, BC. Inequalities: Theory of Majorization and Its Applications. 2. New York: Springer; 2011.
- Muirhead RF. Some Methods Applicable to Identities and Inequalities of Symmetric Algebraic Functions of n Letters. *Proceedings of the Edinburgh Mathematical Society*. 1903; 21:144–157.
- Wald, A. *Sequential Analysis*. New York: John Wiley & Sons; 1947.
- Zybert P, Levin B. Selecting the Highest of Three Binomial Probabilities. *Proceedings of the National Academy of Sciences USA*. 1987; 84:8180–8184.