A Modification of the Bechhofer, Kiefer, and Sobel Sequential Binomial Subset Selection Procedure

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

[1] Department of Biostatistics, Columbia University,
722 West 168th Street,
New York, NY 10032, USA
(e-mail: cl94@columbia.edu)

[2] Department of Biostatistics, Columbia University,
722 West 168th Street,
New York, NY 10032, USA
(e-mail: Bruce.Levin@columbia.edu)

Abstract. We introduce a modification of the Bechhofer, Kiefer, and Sobel sequential subset selection procedure using the Levin-Robbins-Leu adaptive subset selection procedure to allow sequential elimination of inferior populations and recruitment of superior populations. Numerical evidence indicates that the new procedure also guarantees the pre-specified level of probability of correct selection while reducing the expected total number of observations and failures compared with the original non-adaptive method.

Keywords. Adaptive BKS procedure, elimination and recruitment, probability of correct selection, sequential selection, subset selection problem.

1 Introduction

In 1968, Bechhofer, Kiefer, and Sobel (BKS) introduced sequential subset selection procedures for selecting a subset of b best populations from among c candidate populations. The populations envisaged could represent, e.g., clinical treatments, devices, industrial processes, computer operating systems, etc. The BKS subset selection procedure can be applied when the observable outcomes follow any univariate distribution in the exponential family, but because we focus in this paper exclusively on the binomial case, we basically assume that we have c coins with various success probabilities and that our general goal is to select a subset of size $b \geq 1$ containing “best” coins, i.e., coins with the highest probabilities of success. The BKS procedure follows the *indifference zone* approach of Bechhofer (1954) and it is non-adaptive. However, more recent interest in the application of selection procedures to early phase clinical trials with rapid response times has shifted focus away from non-adaptive procedures toward others that, while maintaining equiprobable sampling from competing treatments, allow sequential elimination of inferior populations from the trial based on the accumulating evidence. (We do not consider more fully response-adaptive allocation schemes here.) To the best of our knowledge, no one has proposed an adaptive version of the BKS procedure. We show how to do that here.

2 The procedure

Given $c \geq 2$ coins with labels in the set $C=\{1,...,c\}$. For coin i, let p_i be the probability of heads on a single toss. For a given integer b ($1 \leq b < c$), our primary goal is to select a subset of b coins with the highest such probabilities, which we shall call a subset of b best coins. Without loss of generality, we assume that the coins are ordered from best to worst with success probabilities $p_1 \geq \cdots \geq p_c$. To avoid trivialities, we also assume p_1<1 and p_c>0. We 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, $w_1 \geq \cdots \geq w_c$. We note that the subset size b is fixed in advance. Our approach thus differs from the random subset size selection procedures of Gupta (1956, 1965).

The procedure begins by tossing the coins vector-at-a-time. For n=1,2,..., let $\boldsymbol{X}^{(n)}$= $(X_1^{(n)}, X_2^{(n)},..., X_c^{(n)})$ be the vector that reports the cumulative number of heads observed for each coin after n tosses, and let $\boldsymbol{X}^{[n]}$=$(X_1^{[n]}, X_2^{[n]},..., X_c^{[n]})$ be the ordered $\boldsymbol{X}^{(n)}$ vector with $X_1^{[n]} \geq X_2^{[n]} \geq \cdots \geq X_c^{[n]}$. Let $S(b,c)$ denote the set of all ordered b-tuples of integers $(i_1,...,i_b)$ with $1 \leq i_j \leq c$ for j=1,...,b and $i_1 < \cdots < i_b$. Thus $S(b,c)$ has $\binom{c}{b}$ elements. For any b-tuple $\beta = (i_1,...,i_b) \in S(b,c)$ let $\beta\boldsymbol{X}^{[n]}$ denote the operation $\beta\boldsymbol{X}^{[n]}$= $X_{i_1}^{[n]} + \cdots + X_{i_b}^{[n]}$ and let $(\beta-1)\boldsymbol{X}^{[n]}$ denote the operation

$$(\beta - 1)\boldsymbol{X}^{[n]} = (X_{i_1}^{[n]} + \cdots + X_{i_b}^{[n]}) - (X_1^{[n]} + \cdots + X_b^{[n]}),$$

wherein 1 denotes the b-tuple (1,2,...,b). For pre-specified $\theta > 1$ and P^* with $\binom{c}{b}^{-1} \leq P^* < 1$, the classical BKS procedure terminates with the following stopping criterion:

$$\mathcal{B}(\theta, P^*, b, c) = \inf\{n \geq 1: \sum_{\beta \in S(b,c)\backslash 1} \theta^{(\beta-1)\boldsymbol{X}^{[n]}} \leq \frac{1-P^*}{P^*}\}.$$

If $\mathcal{B}(\theta, P^*, b, c) = n$, the procedure stops and the b coins with the largest tallies are selected. It is possible for several coins to be tied at stopping time; if so, select at random from the ties. BKS (1968) proved that if $w_b / w_{b+1} \geq \theta$, then the probability of correct selection, $P[CS]$, is greater than or equal to P^*. Thus θ and P^* are design constants.

To make this procedure adaptive, after the n^{th} vector of tosses (n=1,2,...), if $\mathcal{B}(\theta, P^*, b, c) > n$, then we check for an elimination or recruitment with the "$\mathcal{E}/\mathcal{R}$" criterion:

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

$N_r^{(b,C)}$ is the *time of first elimination or recruitment* in a c-coin game with coins C and r is chosen in advance of all tosses as the smallest positive integer such that

$$\sum_{i=1}^{b \wedge (c-b)} \binom{b}{i}\binom{c-b}{i} \theta^{-ri} \leq \frac{1-P^*}{P^*}.$$

If $N_r^{(b,C)} = n$, we 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 $X_1^{[n]} - X_{b+1}^{[n]} = r$. 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. 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. If fewer than b coins are recruited and/or fewer than $c\!-\!b$ coins are eliminated at time n, then, before the next toss, we check an adjusted BKS criterion for the remaining subset of coins $C' \subset C$ with $c' = |C'|$ and where b' is b minus the total number of coins recruited. The adjusted BKS criterion, which we denote by $\mathcal{B}'(\theta, P^*, b', c')$, is to stop if

$$\sum_{\beta \in S(b',c')\setminus 1} \theta^{(\beta-1)\boldsymbol{X}^{[n]}} \le \frac{1-P^*}{P^*} - \text{adjustment},$$

where the adjustment is given by

$$\frac{\displaystyle\sum_{i=1}^{b\wedge(c-b)} \binom{b}{i}\binom{c-b}{i}\theta^{-ri} - \sum_{i=1}^{b'\wedge(c'-b')} \binom{b'}{i}\binom{c'-b'}{i}\theta^{-ri}}{P^*\left(1+\displaystyle\sum_{i=1}^{b\wedge(c-b)} \binom{b}{i}\binom{c-b}{i}\theta^{-ri} - \sum_{i=1}^{b'\wedge(c'-b')} \binom{b'}{i}\binom{c'-b'}{i}\theta^{-ri}\right)}.$$

If $\mathcal{B}'(\theta, P^*, b', c')$ is satisfied, then we stop and select the b' coins in the lead together with the recruited $b\text{-}b'$ coin(s) as the b best. If $\mathcal{B}'(\theta, P^*, b', c')$ is not satisfied, then the procedure continues with the next vector of tosses, starting from the current tallies of the remaining subset of coins C', and iterating with stopping criterion $\mathcal{B}'(\theta, P^*, b', c')$ and $\mathcal{E}/\mathcal{R}$ criterion $N_r^{(b',C')}$. Continuing in this way, we stop whenever the $\mathcal{B}'$ criterion is satisfied at some epoch with b'' out of c'' coins left to recruit, or 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.

3 Results

We conjecture the following holds true in general:

For any set of coins C={1,...,c} with success probabilities given by $p_1, \cdots, p_c$, *let* $w_i = p_i/(1\text{-}p_i)$, *and suppose, without loss of generality,* $p_1 \ge \cdots \ge p_c$. *For given **P*** with* $\binom{c}{b}^{-1} \le P^* < 1$ *and pre-specified* $\theta > 1$, *the proposed adaptive procedure achieves a probability of correct selection at least **P*** whenever* $w_b / w_{b+1} \ge \theta$.

We have proved the conjecture for the special case $b = 1$ and any number of c coins, though the proof for general $b > 1$ is still an open problem. Numerical evidence indicates that the new, adaptive procedure reduces the expected total number of observations and failures compared with the original, non-adaptive method, while continuing to guarantee that $P[CS] \ge P^*$.

4 Illustration

Consider selecting $b=2$ best coins from a set of $c=4$ coins with $\theta=2$ and $P^*=0.8$, 0.85, 0.875, or 0.888. The table below presents simulation results with $\underline{p}=(0.4, 0.4, 0.25, 0.25)$ from 40,000 replications of the original BKS procedure, the adaptive BKS procedure, and the Levin-Robbins-Leu (2008) procedure which uses only the adaptive rule $\mathcal{E}/\mathcal{R}$, but not the adjusted BKS criterion $\mathcal{B}'$, stopping only when b coins have been recruited and c-b coins eliminated,. The replications use coordinated, i.e., identical, sequences of binary outcomes for each procedure. Note that in order to guarantee $P[CS] \geq 0.8$ for the adaptive BKS procedure, the integer r used in the adaptive rule $\mathcal{E}/\mathcal{R}$ for eliminating inferior coins and recruiting superior coins needs to be at least 5 for θ=2. Leu and Levin (2008) show that the $P[CS]$ for the Levin-Robbins-Leu (LRL) procedure can be bounded from below by $\{1+\sum_{i=1}^{b\wedge(c-b)}\binom{b}{i}\binom{c-b}{i}\theta^{-ri}\}^{-1} \geq P^*$ for $w_b / w_{b+1} \geq \theta$. In our example, the expression on the left is 0.888, somewhat larger than required, resulting in a greater expected number of vectors of observations **(**$E_{\underline{p}}[N]$), a greater expected total number of tosses ($E_{\underline{p}}[T]$), and a greater expected number of failures ($E_{\underline{p}}[F]$) compared to the original BKS procedure. However, as P^* increases while the required r remains at the same level (i.e. r=5), $E_{\underline{p}}[T]$ and $E_{\underline{p}}[F]$ start to favor the LRL procedure. Nevertheless, no matter how P^* changes along the line while r stays the same, the adaptive BKS procedure seems to always perform better than both LRL and the original BKS procedure in terms of $E_{\underline{p}}[T]$ and $E_{\underline{p}}$[F] while still achieving a comparable level of $P[CS]$.

Table 1. Simulation results from 40,000 coordinated replications for the case b=2, c=4, $\underline{p}$=(0.4, 0.4, 0.25, 0.25), θ=2, r=5 (standard errors of simulation in parentheses)

	Operating characteristic			
	$P_{\underline{p}}[cs]$	$E_{\underline{p}}[N]$	$E_{\underline{p}}[T]$	$E_{\underline{p}}[F]$
<u>P^*=0.8</u>				
Levin-Robbins-Leu	0.9091 (0.0014)	49.3 (0.14)	152.7 (0.36)	103.3 (0.25)
original BKS	0.8425 (0.0018)	33.2 (0.09)	133.0 (0.37)	89.8 (0.25)
adaptive BKS	0.8448 (0.0018)	34.6 (0.09)	121.7 (0.29)	82.3 (0.20)
<u>P^*=0.85</u>				
Levin-Robbins-Leu	0.9091 (0.0014)	49.3 (0.14)	152.7 (0.36)	103.3 (0.25)
original BKS	0.8754 (0.0017)	38.0 (0.10)	152.2 (0.40)	102.8 (0.28)
adaptive BKS	0.8850 (0.0014)	41.6 (0.11)	137.4 (0.32)	93.0 (0.22)
<u>P^*=0.875</u>				
Levin-Robbins-Leu	0.9091 (0.0014)	49.3 (0.14)	152.7 (0.36)	103.3 (0.25)
original BKS	0.8950 (0.0015)	40.9 (0.11)	163.7 (0.43)	110.6 (0.30)
adaptive BKS	0.9086 (0.0014)	49.2 (0.14)	152.4 (0.36)	103.2 (0.25)
<u>P^*=0.888</u>				
Levin-Robbins-Leu	0.9091 (0.0014)	49.3 (0.14)	152.7 (0.36)	103.3 (0.25)
original BKS	0.9130 (0.0014)	43.9 (0.12)	175.7 (0.47)	118.6 (0.32)
adaptive BKS	0.9091 (0.0014)	49.3 (0.14)	152.7 (0.36)	103.3 (0.25)

References

Bechhofer, R.E. (1954). A Single-sample Multiple Decision Procedure for Ranking Means of Normal Populations with Known Variances. *Ann. Math. Statist.* **25**, 16-39.

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

Gupta, S. S. (1956). On a Decision Rule for a Problem in Ranking Means, *Mimeograph Series 150, Institute of Statistics*, Chapel Hill: University of North Carolina.

Gupta, S. S. (1965). On Some Multiple Decision (Selection and Ranking) Rules, *Technometrics* 7: 225–245.

Leu, C.S. and Levin, B. (2008). On a Conjecture of Bechhofer, Kiefer, and Sobel for the Levin-Robbins-Leu subset selection procedures. *Sequential Analysis* **27**, 106-125.