# WSC15 paper

### July 2, 2015

## Reviewer 1.

1. *Frazier 2014 was cite lots of times, modify the cite commands accordingly to make sure the correct format.* We have corrected this.

2. *Second corrections.* We have corrected this.

3. *Use Section 2, 3, 4 instead just 2,3,4 when referring to those sections.* We have corrected this. To preserve space, we have used the section symbol, rather than the word "Section".

4. *Page 3, 3rd paragraph.* Now it is correct.

5. *Page 3, 4th paragraph.* Now it is correct.

6. *Page 3, 1st paragraph in Section 3.* Now it is correct.

7. *3rd paragraph in Section 3.*
   *– It's not clear what n_ tx is, especially the range for t and x.*
   *– Says 100 is the recommended value for n_ 0, but later in the algorithm and experiments used something else.*

   We have clarified the description of $n_{tx}$.

   Regarding $n_0$, we do indeed recommend the value of 100, but only when variances are unknown. We have clarified this in the text in the noted place in Section 3. Additionally, in the unknown variance example in the numerical experiments, we use a value lower than the recommended value to make correct selection more difficult, and to observe the full range of behavior from the algorithm. When the recommended value is used, the algorithm provides performance similar to that shown in Figure 1(a). We have clarified this in the text discussing the second example in Section 5.

8. *End of page 3.* We have clarified the definition of $Y_{n_{tx},x}$.

9. *Section 4, 3rd paragraph.* Now it is correct.

10. *Section 6.* Now it is correct.

# Reviewer 2.

1. *It seems that the choice of n0 is relatively arbitrary. In practice, can problem properties be used to obtain some understanding of how large/small n0 should be?*

   Our experience is that when n0 is chosen reasonably large ($\geq 100$), then the performance of BIZ is robust, in terms of providing a PCS that is consistently above $P^*$. When n0 is smaller than this, we observe, as we did in the numerical experiments, that PCS may fall below $P^*$. However, we have not yet done enough experiments to say anything beyond the statements in the current paper recommending a value of 100 for n0.

   In Section 5, we previously stated,

   "In this example, we have intentionally chosen $n_0$ to be small, and have chosen a large variance for the best system, to cause BIZ to fail to meet the IZ guarantee for $\delta > 0$. Increasing the parameter $n_0$ typically causes BIZ to meet the IZ guarantee for all $\delta$, and we recommend a larger value of $n_0$ in practice."

   We have added to this the sentence, "The choice of $n_0$, and its impact on PCS, merits further study."

2. *It does seem to me that we choose a far too stringent threshold, the resulting problem may be rendered infeasible since μ/k are exogenously specified and presumably unavailable a prior and δ is a user-specified parameter. In re-reading the paper, it seems that the alternatives may have nearly the same means, in which case, the likelihood of infeasibility may be modest.*

   Actually, in the problem that we study, there cannot be infeasibility. We believe that the reviewer is referring to the fact that the configuration may not be in the preference zone. In this case, the indifference zone guarantee does not guarantee a probability of correct selection, but the problem is not infeasible. This is a property of the indifference-zone formulation of ranking and selection, and is not specific to the current paper.

   In practice, most procedures that satisfy the indifference-zone guarantee also satisfy what is called a "probability of good selection" guarantee, which is defined by Nelson and Banerjee (2001). This guarantee does apply to configurations inside of the preference-zone — it says that with probability $P^*$, the procedure selects an alternative within $\delta$ of the best. We have added a brief discussion of this point, and the reference to Nelson and Banerjee (2001), to the paper. Extending our results to a probability of good selection guarantee is worthwhile, and we hope to pursue it in future work.

3. *While the BIZ procedure is asymptotically valid, naturally in practice, the schemes have to be terminated after a finite number of samples. Are there error bounds available, even in a restricted regime, that can govern the choice of when the procedure may be terminated.*

   The BIZ procedure terminates on its own, in a finite number of samples. The user does not need to make this decision. The asymptotic aspect of the analysis refers to $\delta$ becoming small. When the procedure terminates, in the cases considered in Frazier (2014) (which includes common known variance with normal samples), the procedure satisfies the IZ guarantee, which is an error bound, even which $\delta > 0$.

   We clarified this point in section 3 by stating that the algorithm ends in a finite number of steps.

4. *In the numerical experiments, the choice of μ and δ ensures feasibility. Suppose $\mu_{100}$ is set to $\delta+1$, then unless I have misunderstood how the scheme works, the threshold constraint would render the problem infeasible. In this case, how would the scheme respond?*

As discussed above, the problem would not become infeasible. Instead, a problem configuration might not be in the preference zone. For problem configurations outside of the preference zone, the indifference-zone guarantee does not provide any guarantee on PCS.

5. *It might be useful to see how the BIZ procedure is indeed less conservative in practice compared to its counterparts in the R&S literature. The authors claim that the proposed procedure requires less samples – presumably this is only an empirical observation. If so, please emphasize that this is the case; else it might be worth pointing to the formal statement that guarantees this claim.*

   This statement is based on extensive numerical experiments in Frazier (2014).

   We state in the introduction, when the BIZ procedure is first mentioned, that

   "In numerical experiments, the number of samples required by BIZ is significantly smaller than that of procedures like the $P_{\mathcal{B}}^*$ procedure of Bechhofer, Kiefer, and Sobel (1968) and the KN procedure of Kim and Nelson (2001), especially on problems with many alternatives."

   This sentence states that the observation is empirical, as it is based on numerical experiments.

   We have modified the conclusion to also point out the empirical nature of this statement:

   "[BIZ] has been observed empirically to take fewer samples than other IZ procedures, especially for problems with large numbers of alternatives..."

6. *Finally, some quick observations regarding the references.*

   We have corrected these issues. We reference Billingsley twice because we use results from two different editions of the book (1968 and 1999).