Supporting Information for "Bandit solutions provide unified ethical models for randomized clinical trials and comparative effectiveness research"

William H. Press

Department of Computer Science and School of Biological Sciences,
The University of Texas at Austin, Austin, TX 78712

In this Supporting Information we fill in some missing mathematical steps and supporting details for the results given in the main paper.

State Variables and Successor Points

The successor points $\mathbf{m}^{(i+)}$ mentioned after Eq. **4** in the main text are explicitly

$$\begin{aligned}
\mathbf{m}^{(1+)} &= (m_A + 1, \overline{m}_A, m_B, \overline{m}_B) \\
\mathbf{m}^{(2+)} &= (m_A, \overline{m}_A + 1, m_B, \overline{m}_B) \\
\mathbf{m}^{(3+)} &= (m_A, \overline{m}_A, m_B + 1, \overline{m}_B) \\
\mathbf{m}^{(4+)} &= (m_A, \overline{m}_A, m_B, \overline{m}_B + 1)
\end{aligned} \tag{1}$$

and similarly for the predecessor points $\mathbf{m}^{(i-)}$, substituting minus for plus. Figure S1 shows the state vector $\mathbf{m}$ and its successor states. Which successor state is chosen depends, first, on whether a patient is assigned to treatment A or B, and second, on whether the treatment is a success or a failure, which occurs with respective probabilities a and b for A and B.

From the various definitions it is easy to prove relations like

$$\begin{aligned}
aP(a|\mathbf{m}) &= \frac{m_A + 1}{m_A + \overline{m}_A + 2} P(a|\mathbf{m}^{(1+)}) \equiv \langle a \rangle_{\mathbf{m}} P(a|\mathbf{m}^{(1+)}) \\
(1-a)P(a|\mathbf{m}) &= \frac{\overline{m}_A + 1}{m_A + \overline{m}_A + 2} P(a|\mathbf{m}^{(2+)}) \equiv \langle 1-a \rangle_{\mathbf{m}} P(a|\mathbf{m}^{(2+)})
\end{aligned} \tag{2}$$

and correspondingly for b, with $\mathbf{m}^{(3+)}$ and $\mathbf{m}^{(4+)}$.

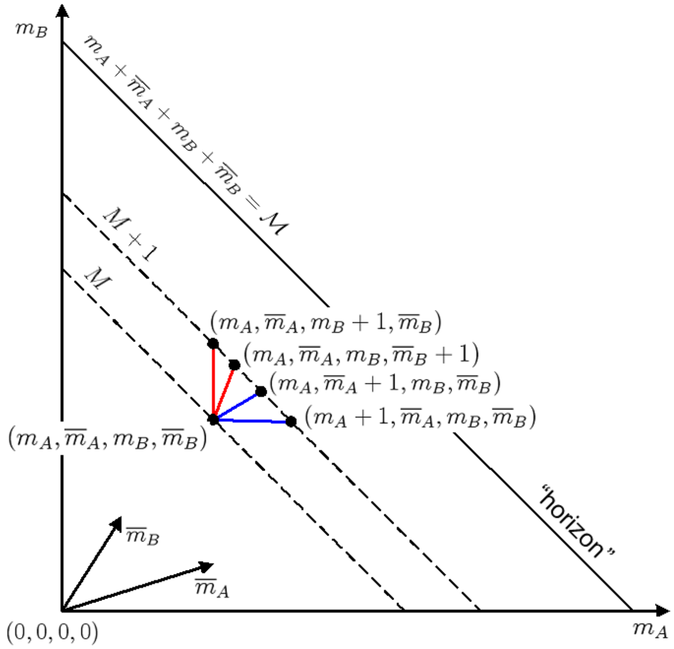

Figure 1: Schematic of the 4-dimensional state space. A state is a point on the nonnegative integer lattice. It has one of 4 successor points, depending on whether treatment A (blue) or B (red) is assigned and on whether the treatment succeeds or fails.

Recurrence for the Exact Solution

The probability that a path that goes through $\mathbf{m}_0$ passes through a point $\mathbf{m}$ is denoted $p(\mathbf{m}|a, b, \mathbf{r}, \mathbf{m}_0)$. A forward recurrence for $p(\mathbf{m}|a, b, \mathbf{r}, \mathbf{m}_0)$ simply tracks where the probability at point $\mathbf{m}$ came from:

$$\begin{aligned} p(\mathbf{m}|a, b, \mathbf{r}, \mathbf{m}_0) =& a\, r_{\mathbf{m}^{(1-)}}\, p(\mathbf{m}^{(1-)}|a, b, \mathbf{r}, \mathbf{m}_0) \\ &+ (1-a)\, r_{\mathbf{m}^{(2-)}}\, p(\mathbf{m}^{(2-)}|a, b, \mathbf{r}, \mathbf{m}_0) \\ &+ b\,(1 - r_{\mathbf{m}^{(3-)}})\, p(\mathbf{m}^{(3-)}|a, b, \mathbf{r}, \mathbf{m}_0) \\ &+ (1-b)\,(1 - r_{\mathbf{m}^{(4-)}})\, p(\mathbf{m}^{(4-)}|a, b, \mathbf{r}, \mathbf{m}_0) \end{aligned} \tag{3}$$

valid for $\mathbf{m} \succ \mathbf{m}_0$. Of course we also have $p(\mathbf{m}_0|a, b, \mathbf{r}, \mathbf{m}_0) \equiv 1$.

A backwards recurrence is obtained by explicitly enumerating how the probability at $\mathbf{m}_0$ devolves to its four successors:

$$\begin{aligned} p(\mathbf{m}|a, b, \mathbf{r}, \mathbf{m}_0) =& a\, r_{\mathbf{m}_0}\, p(\mathbf{m}|a, b, \mathbf{r}, \mathbf{m}^{(1+)}) \\ &+ (1-a)\, r_{\mathbf{m}_0}\, p(\mathbf{m}|a, b, \mathbf{r}, \mathbf{m}^{(2+)}) \\ &+ b\,(1 - r_{\mathbf{m}_0})\, p(\mathbf{m}|a, b, \mathbf{r}, \mathbf{m}^{(3+)}) \\ &+ (1-b)\,(1 - r_{\mathbf{m}_0})\, p(\mathbf{m}|a, b, \mathbf{r}, \mathbf{m}^{(4+)}) \end{aligned} \tag{4}$$

Start with the definition of the total cost

$$C(\mathbf{m}_0|\mathbf{r},M_h) = \left\langle \sum_{\mathbf{m}\succeq\mathbf{m}_0}^{M<M_h} c(\mathbf{m}|a,b,\mathbf{r})p(\mathbf{m}|a,b,\mathbf{r},\mathbf{m}_0) \right\rangle_{\mathbf{m}_0} \tag{5}$$

(main text Eq. **7**). Writing the first term in the sum explicitly gives

$$C(\mathbf{m}_0|\mathbf{r},M_h) = \left\langle c(\mathbf{m}_0|a,b,\mathbf{r}) + \sum_{\mathbf{m}\succ\mathbf{m}_0}^{M<M_h} c(\mathbf{m}|a,b,\mathbf{r})p(\mathbf{m}|a,b,\mathbf{r},\mathbf{m}_0) \right\rangle_{\mathbf{m}_0} \tag{6}$$

Now substitute Eq. **4** for $p(\mathbf{m}|a,b,\mathbf{r},\mathbf{m}_0)$ and use Eq. **2** to get

$$\begin{aligned} C(\mathbf{m}_0|\mathbf{r},M_h) = {} & \langle c(\mathbf{m}_0|a,b,\mathbf{r})\rangle_{\mathbf{m}_0} \\ & + \langle a\rangle_{\mathbf{m}_0} r_{\mathbf{m}_0} C(\mathbf{m}_0^{(1+)}) + \langle 1-a\rangle_{\mathbf{m}_0} r_{\mathbf{m}_0} C(\mathbf{m}_0^{(2+)}) \\ & + \langle b\rangle_{\mathbf{m}_0} (1-r_{\mathbf{m}_0}) C(\mathbf{m}_0^{(3+)}) + \langle 1-b\rangle_{\mathbf{m}_0} (1-r_{\mathbf{m}_0}) C(\mathbf{m}_0^{(4+)}) \end{aligned} \tag{7}$$

This is the principal result used in the main text. It is a backward recurrence for the cost $C(\mathbf{m}_0|\mathbf{r},M_h)$ in terms of the 4 costs $C(\mathbf{m}_0^{(i+)}|\mathbf{r},M_h)$ with $i = 1,2,3,4$, and can be started at the horizon M_h with

$$C(\mathbf{m}_0|\mathbf{r},M_h) \equiv 0 \quad \text{when} \quad M \geq M_h \tag{8}$$

If at each point we locally choose $r_{\mathbf{m}}$ to minimize C, then, as noted in the main text, the recurrence guarantees that each point will acquire the globally smallest cost C to the horizon; so we get an optimal strategy $\mathbf{r}$.

Fitting the Heuristic

Here we give some further justification for the functional form of main text Eq. **23**. We first consider the dependence of t_{crit} on the horizon M_h. For definiteness, suppose that $\mu_{A|\mathbf{m}} > \mu_{B|\mathbf{m}}$ so that t is positive, indicating that treatment A is likely superior to B. The tail probability of this indication being wrong scales as (neglecting polynomial terms) $\exp(-\frac{1}{2}t^2)$. If the indication is wrong, but we follow it anyway, then lost successes accrue for $\sim M_h$ patients. So the cost of the indication scales as

$$C_{\text{wrong}} \sim \exp(-\tfrac{1}{2}t^2)M_h \tag{9}$$

On the other hand, if we don't follow the indicated superiority of A, then we are essentially throwing away $\sim M$ data points in favor of a new data set. The cost of this scales as

$$C_{\text{right}} \sim M \tag{10}$$

Equating C_{wrong} and C_{right}, to find the boundary, gives

$$t_{\text{crit}} \sim \sqrt{\log(M_h/M)} \tag{11}$$

Trying this functional form on data from the exact solutions, one finds that the actual dependency on M_h is slightly weaker. Fitting for the best-fitting exponent gives

$$t_{\text{crit}} \sim [\log(M_h/M)]^{0.42} \tag{12}$$

as mentioned in the main text. There is nothing natural about this form, and it seems more likely that the real answer is Eq. **11**, but modified by subdominant effects. However, Eq. **12** is good enough for present purposes.

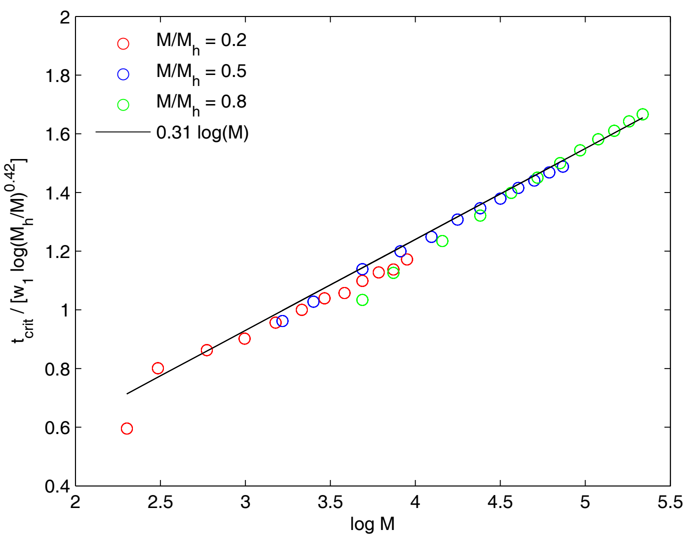

Figure 2: Residual dependence of t_{crit} on M after the dependence of Eq. **12** is removed. One sees very nearly linear dependence on $\log M$.

Now assuming that all dependence on M_h is captured in Eq. **12**, we can examine the dependence of t_{crit} on M, for a range of values $0 < M/M_h < 1$. Figure 2 shows typical data of this kind. One sees that the dependence is very nearly linear in $\log M$, and we adopt this functional dependence in the main text's Eq. **23**.

One could get a more accurate, but more complicated, functional form by noting in the Figure that different values of M/M_h have visibly slightly different slopes and intercepts, indicating a slight dependency on M_h not modeled by Eq. **12**. One could in principle fit an empirical model to these small differences, but this is unnecessary here.

Probability Distribution of Costs

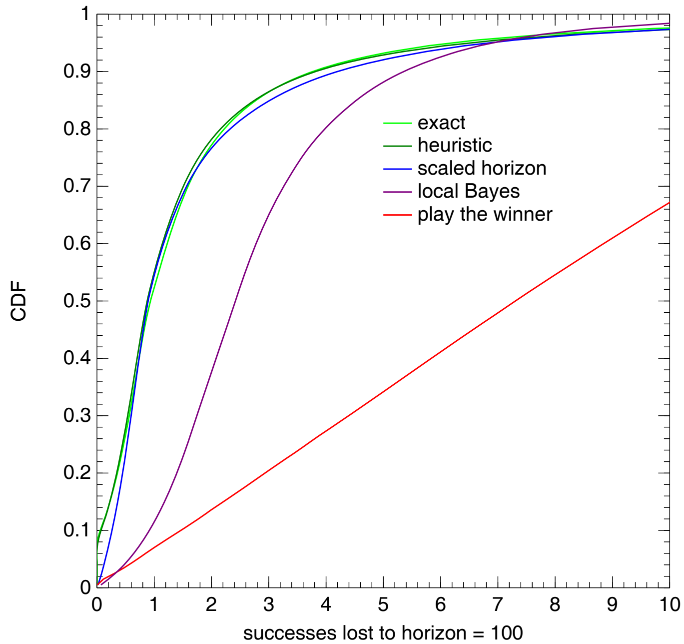

Figure 3: Cumulative distribution functions for the trial ESL costs for trials with $M_h = 100$ and starting points $(0, 0, 0, 0)$. These are distributions whose means are given in Table 1 in the main text.

Table 1 in the main text compared various strategies by evaluating, via Monte Carlo, their expected (i.e., mean) costs. However, other properties of the distributions can also be relevant. Figure 3 shows the full distributions for the case $M_h = 100$ and starting point $(0, 0, 0, 0)$, that is, no prior information. One sees that the exact and heuristic strategies are barely distinguishable. Both have a finite CDF at near-zero cost, because there are occasional trials in which an inferior treatment is assigned nearly zero times. The scaled-horizon strategy is seen to be almost as good as the exact and heuristic strategies, except in these unusually lucky cases.

Comparing the two horizonless strategies, one sees that the scaled-horizon strategy is significantly superior to the local Bayes strategy. However, it is somewhat heavier-tailed, so that it can produce poorer results in unfavorable cases, about 4% of the time (CDF > 0.96). It is not difficult to design small modifications to scaled-horizon that substantially improve its tail performance at the expense of slightly widening the gap between it and the optimal (fixed horizon) strategy.

Play the winner is roughly uniform up to about twice its (costly) mean, falling off rapidly thereafter.