INDIAN INSTITUTE OF MANAGEMENT KOZHIKODE

Working Paper

IIMK/WPS/459/ITS/2021/05

March 2021

BBECT: Bandit -based Ethical Clinical Trials

Mohammed Shahid Abdulla[1]
L Ramprasath[2]

[1] Associate Professor, Information Technology and Systems, Indian Institute of Management, Kozhikode, IIMK Campus PO, Kunnamangalam, Kozhikode, Kerala 673570, India; Email: shahid@iimk.ac.in, Phone Number (+91) 495 – 2809254

[2] Associate Professor, Finance, Accounting and Control , Indian Institute of Management, Kozhikode, IIMK Campus PO, Kunnamangalam, Kozhikode, Kerala 673570, India; Email: lrprasath@iimk.ac.in, Phone Number (+91) 495 – 2809248

BBECT: Bandit -based Ethical Clinical Trials

Mohammed Shahid Abdulla[†] & L Ramprasath[*]
† *Information Systems Area, IIM Kozhikode*
[*]*Finance, Accounting and Control Area, IIM Kozhikode*

Abstract

An aim of Ethico-Optimal clinical trials of drugs in Phase III is to randomly allocate a new drug (ND) to patients in the sample, but with a greater fraction being administered ND if doing so is statistically justified. Such an adaptation is not possible in static trials designed with a sample size N in which approximately half the patients would receive the current drug or standard of care (SOC), despite evidence within the trial that ND is efficacious. We adapt a canonical stochastic multi-armed bandit algorithm named UCB1 to a clinical trials setting and analyse the resulting Type-2 error β, as also minimum sample size N required by such a trial for a certain β level. The difference in our proposal is not just in the allocation rule that applies to patients or volunteers in the trial, but also in the inference rule to decide if null hypothesis can be rejected. We also present simulations to establish that the ethical properties of such a trial are higher, both to verify our analysis and demonstrate an empirical advantage when compared to 2 existing methods. In these simulations, we also propose and demonstrate a device to achieve low or comparable Type-1 error α vis-a-vis existing methods.

1. Introduction

We first introduce the stochastic multi-armed bandit (SMAB) problem from statistical learning by means of a simple algorithm designed to solve the problem. A stochastic multi-armed bandit problem setting has K arms or levers, each producing an outcome with mean reward $\{\mu_k\}_{k=1}^K$, arranged such that $\mu_k > \mu_{k+1}$. After pulling each of these K arms in the first K rounds, the single player is permitted to pull any one of these K arms in each succeeding round $K+1 \leq t \leq N$, with each pull yielding a reward $X_t \in [0,1]$. Note here that the random variable X_t belongs to probabilty distribution function $\mathcal{F}_{a_t}$ with support over $[0,1]$, where $a_t \in \{a^1, a^2, ..., a^K\}$ is the arm or action chosen to be pulled at t. Also note that $E(X_t) = \mu_{a_t}$, with X_t being used to update empirical means $\{\bar{X}_t^k\}_{k=1}^K$ that the player maintains for each arm a^k. If $a_t = a^k$, then empirical mean $\bar{X}_t^k$ is updated as follows: $s_t^k := s_{t-1}^k + 1$, followed by $\bar{X}_t^k := \frac{s_{t-1}^k \bar{X}_{t-1}^k + X_t}{s_t^k}$. The quantity s_t^k is the count of pulls of arm k till (and including) t, and the information $\{s_t^k\}_{k=1}^K$ is also retained by the player as she goes to round $t+1$.

A common and efficient algorithm for the SMAB algorithm is UCB1 (Upper Confidence Bound variant 1) [1] which infers a_{t+1} as follows:

$$a_{t+1} \quad = \quad \arg \max_{1 \leq k \leq K} \{\bar{X}_t^k + \sqrt{\frac{2 \log(t)}{s_t^k}}\}.$$

UCB1 is a *logarithmic* regret SMAB algorithm, in that it has been established:

$$E(s_N^k) \quad \leq \quad \frac{8 \log(N)}{\Delta^2} + (1 + \frac{\pi^2}{3}), \text{ for } k > 1, \text{ for all } N > \tfrac{8 \log(N)}{\Delta^2}, \tag{1}$$

where $\Delta = \mu_1 - \mu_2$. Such an assurance indicates that only $O(\log T)$ of T opportunities to pull arms were lost to the sub-optimal choice a^k, $k > 1$.

In a simple $2-$arm binary-response Phase 3 clinical trial for a new drug (ND), a sample size N of patients with a particular condition is decided using certain measurements made in the preceding Phase 2 trials. A key input into deciding N is the statistical significance required in the Phase 3 trial's conclusion, most notably the clinical trial's Type-1 error α. This error is the probability of recommending ND when said drug's performance is not statistically different from current standard of care (SOC) or *placebo*. In a randomized controlled trial, which we call static clincal trial alternatingly, roughly $\frac{N}{2}$ patients are administered ND, whilst the other half are administered SOC. This is done in such a way that patients do not know and cannot reliably infer which of the 2 drugs, a^1 or a^2, they have received. It is considered ethical within such a trial to administer more number of patients with the new drug if there is statistical evidence *till that point* in the trial of better outcomes. Such ethico-optimal trials have been investigated in [2] and in predecessor publications. If we assume $K = 2$ in the earlier description of UCB1, then being able to observe the outcome X_t of administering drug a_t to the $t-$th patient helps decide a_{t+1}. Further, the new drug likely has the greater mean outcome $\mu_1 > \mu_2$, as observed by the investigators in a Phase 2 trial, and hence from UCB1's analysis $s_N^2 = O(\log(N))$. The size of the trial can be set to have N subjects, such that

$$N \quad = \quad \min_{N' \in \mathcal{Z}} \text{ s.t. } N' \geq 2 \times \lceil \frac{8 \log(N')}{\Delta^2} \rceil. \tag{2}$$

Then, the number of patients administered SOC would be $O(\log(N))$, a quantity with promise of being lower than the $\frac{N}{2}$ in a static clinical trial. Our aim is to employ a variant of UCB1 - UCB1-MPA described below - as the basic unit of generic bandit-based clinical trial algorithms which we call BBECT.

We next describe in greater detail, from [3], the bandit-based recommendation algorithm UCB1-MPA (Most Played Arm):

$$a_{t+1} \quad = \quad \arg \max_{k \in \{1,2\}} \{\bar{X}_t^k + \sqrt{\frac{\alpha \log(t)}{s_t^k}}\}.$$

The method to *allocate* the $(t+1)-$th arm is similar to UCB1, except for the constant α in place of 2. However, at the end of N pulls, UCB1-MPA also *recommends* as the best therapy the arm $a* = \arg_{k=1,2} \max\{s_N^k\}$. This also termed as *inference* in a statistical trial, which is usally performed by constructing values such as Z-statistic and looking up tables. UCB1-MPA modifies UCB1's regret expression to derive an upper-bound on the probability that a^2 (SOC) would be recommended in place of a^1 (ND). Such a probability would thus be the *Type-2 error* β of the SMAB-based clinical trial: the probability that SOC will be recommended despite a statistically significant difference between ND and SOC. However, there appears to be no direct way to assess the *Type-1 error* except by simulating a high number of clinical trials where Δ is set to 0. UCB1-MPA currently uses the $\frac{N}{2}$ threshold to decide which of the 2 drugs to recommend, but with $\Delta = 0$ such an outcome would be noisy between simulations. A further modification to UCB1-MPA can be thought of such that a^1 is recommended only if it is used for a larger fraction of the N pulls, e.g. as the allocation of 60% of the N subjects in the trial. Such a modification would also help to capture the analogous conclusion in clinical trials where there is insufficient information to reject the null hypothesis.

Note, however, that N in (2) above is typically much larger than the N recommended by Z-statistic based methods used in static clinical trials. It is clarified here that $\Delta = \mu_1 - \mu_2$ is an approximate quantity known to the clinical trial investigators on account of Phase-2 findings, that precede Phase-3 for which BBECT is being proposed. An example here is that for $\Delta = 0.2$, the lowest possible N if using native UCB1 is 3233. Here, by native UCB1 we mean the UCB1 algorithm without the much lower N our proof technique allows to calculate. The N calculated from formula in [4], as required for static binary response clinical trials, is s.t. $N = 326$. In Section 2 below, we describe BBECT using UCB1-MPA with a proof that obtains an N much lower than in the proof of UCB1-MPA itself. We will also explain the formula for N required in static clinical trials later in Section 2. BBECT with UCB-1 is also comparable with binary-response static clinical trials, and has the advantage that more subjects obtain the ND. Further, in Section 3, we run a series of experiments with BBECT based on UCB1-MPA to compare with both static clinical trials and ethico-optimal clinical trials.

As the central module in BBECT, there also are other bandit algorithms apart from UCB1 that have various advantages:

- relaxation of $X_t \in [0, 1]$ can take place, such algorithms are called 'heavy-tailed bandit' algorithms, and
- regret quantity s_N^2 can be lower due to different scaling constants and offsets, but are still $O(\log(N))$.

We have, however, chosen UCB1's variant UCB1-MPA as the module for proposed BBECT due to UCB1 being a canonical and easy-to-analyse SMAB algorithm. Indeed, in recent work such as [5] *forward-looking* multi-armed bandit algorithms have been drafted for clinical trials. Such algorithms calculate indices at each step of the trial, also involving information such as number of enrolled trial participants left, to decide the allocation of patient one or the other arm of the trial. However, the work there deals with continuous and normally-distributed outcomes, assumes a prior distribution for parameters of both arms in the trial, and also has a worse tradeoff of trial's power vis-a-vis ethical outcome. When using BBECT with UCB1-MPA, the sample size N can be calculated in advance based on β required (in that sense it is not myopic as defined by [5]), and does not require any priors other than an estimate of Δ. Note also the theoretical formulation of regret in the Gittins Index method [6, (17)] where regret in N steps is greater than $\frac{1600}{\Delta} \cdot \log N$, an unfavourable scale coefficient compared to (1) above. The substitution for this is as follows: $\Delta + \frac{128}{\Delta} \log(2N^2) + 21 \cdot \Delta \lceil \frac{32}{\Delta^2} \log(4N^2) \rceil + 10N \cdot \frac{c''(\log N + \log_+(N\Delta^2))}{N\Delta}$, where we neglect the last term in our calculation. We thus get $\frac{(256+21 \cdot 64+20c'') \log N}{\Delta}$ and hence the assumption of $\frac{1600}{\Delta} \cdot \log N$ earlier. In [7], an improved UCB1 named 'Optimally-Confident UCB' is presented whose coefficients of regret are difficult to compare directly. However, it is observed there empirically that a Gittins-index -based strategy is the winner in regret terms among a large set of algorithms for small horizons e.g. $N \leq 1000$. An original investigation using bandit algorithms in binary response trials, that identifies multiple ethical criteria, was performed in [8].

Even within Management research literature, a shift from quantitative experiments to bandit-based adaptive experiments is being proposed in [9]. The work in [9] does not propose deriving a sample size for any particular bandit approach (though it employs Thompson sampling bandit). Further, as an illustration, a large N in the simulated trial in [9] results in an Efficient Allocation Proportion - i.e. the fraction of subjects randomized to the better therapy - of 83%. Our algorithm BBECT with UCB1-MPA achieves 80% for even the $1-$st percentile of outcomes from $10,000$ simulations. The publication, however, makes several valid points about Management literature requiring to shift from 'fixed and balanced' randomization to more 'exploration-exploitation' methods, at least in sequential trials.

2. Proof of BBECT regret in recommended action

An outline of UCB's proof of logarithmic regret (Theorem 1 in [1]) would serve as a useful illustration. The proof there requires that at least one of the following 3 events occur at an index $t < N$:

$$\begin{aligned} \bar{X}_t^1 &\leq \mu^1 - c_t^1 \\ \bar{X}_t^2 &\geq \mu^2 + c_t^2 \\ \mu_2 + 2c_t^2 &\geq \mu_1. \end{aligned} \tag{3}$$

The probability of these 3 events forms the upper bound for the probability of the event $\{\bar{X}_t^1 + c_t^1 \leq \bar{X}_t^2 + c_t^2\}$, which in turn indicates that at index $t+1$ the suboptimal arm 2 was pulled. Notice that the event in (3) does not occur if $s_t^2 > \frac{8\log(t)}{\Delta^2}$ due to the form of $c_t^2 = \sqrt{\frac{2\log(t)}{s_t^2}}$. It is thus sufficient if $s_t^2 > \frac{8\log(N)}{\Delta^2}$.

Also, by the Chernoff-Hoeffding concentration inequality, $P(\bar{X}_t^1 \leq \mu^1 - c_t^1) \leq e^{-2s_t^1(c_t^1)^2} = t^{-4}$. The same holds for $P(\bar{X}_t^2 \geq \mu^2 + c_t^2)$ above.

Theorem 1. *For each $\beta \in (0,1)$, $\exists L_0$ and N such that $N \geq 2L_0$, $L_0 = \frac{2(r_0+1)^2 \log(N)}{\Delta^2}$, where $r_0 < 1$, s.t. $\sum_{t=L_0+1}^{N} P\{\bar{X}_t^1 + c_t^1 \leq \bar{X}_t^2 + c_t^2\} < 1 - \beta$*

Proof. We begin with a modification to obtain a more favourable condition on s_t^2 above:

$$\begin{aligned} \bar{X}_t^1 &\leq \mu^1 + x_t c_t^1 - c_t^1 & (4) \\ \bar{X}_t^2 &\geq \mu^2 + r_t c_t^2 & (5) \\ \mu_2 + (r_t + 1)c_t^2 &\geq \mu_1 + x_t c_t^1. & (6) \end{aligned}$$

Just as the event (3) above, event (6) requires to be ruled out for s_t^2 exceeding a certain threshold. The following conditions on x_t and r_t result:

$$\begin{aligned} \mu_1 - \mu_2 &\geq -x_t c_t^1 + (r_t + 1)c_t^2 \text{ where, we solve} \\ \Delta &= -x_t\sqrt{\frac{2\log t}{S_t^1}} + (r_t+1)\sqrt{\frac{2\log t}{S_t^2}} \text{ as a sufficient condition.} \\ \text{In the above, } S_t^2 &= \frac{2(r_0+1)^2 \log N}{\Delta^2} \\ S_t^1 &= t - \frac{2(r_0+1)^2 \log N}{\Delta^2} \text{ and choose an } r_0 \text{ s.t.} \\ \frac{2(r_0+1)^2 \log N}{\Delta^2} &\ll \frac{8 \log N}{\Delta^2} \end{aligned}$$

Now notice from (4)-(5) that we can place similar concentration conditions for both $\bar{X}_t^1$ and $\bar{X}_t^2$ as in the original UCB1. Thus, we set the constraint that $1 - x_t = r_t$, and obtain the solution:

$$x_t = \frac{-1 + \frac{2}{r_0+1}\sqrt{\frac{\log(t)}{\log(N)}}}{\sqrt{\frac{2\log(t)}{t\Delta^2 - 2(1+r_0)^2\log(N)} + \frac{1}{r_0+1}\sqrt{\frac{\log(t)}{\log(N)}}}}$$

Further $L_0 = \frac{2(r_0+1)^2 \log N}{\Delta^2}$. Notice that Chernoff-Hoeffding bound can be applied as follows: $P(\bar{X}_t^2 \leq \mu^2 + r_t c_t^2) \leq e^{-2s_t^2(r_t c_t^2)^2} = e^{-2 \cdot 2r_t^2 \log t}$. Further, the resultant value is $(e^{-4\log t})^{r_t^2}$. Use x_t, r_t to obtain N such that $\sum_{t=L_0+1}^{N} 2t^{-4r_t^2} < 1 - \beta$, where β is the desired power of the statistical test. □

For example, we obtain $N = 668$ with $L_0 = 334$ (corresponding to $r_0 = 0.013$) when we input $\beta = 0.9$ and $\Delta = 0.2$. We have assumed that the process $\{\bar{X}_t^1\}$ is an empirical mean of s_t^1 events of type $\{0,1\}$ with Bernoulli parameter p_1. For the static clinical test, there exists a combination $p_1 = 0.6$, $p_2 = 0.4$, for which the $N = 326$ obtained is much lesser. Yet, we will demonstrate using simulation that treatment failures in BBECT are lower (alternatively, efficient allocation proportion for BBECT is superior). A grid search for r_0 for each possible value of Δ (with $\beta = 0.9$ employed) yields value of N suited to BBECT using UCB1-MPA. These values are compared below in Table 1 for BBECT versus static clinical trials, where the maximum possible value of N over varied p_1, p_2 pairs is recorded, s.t. $p_1 - p_2 = \Delta$. We use the formula for hypothesis testing applicable to settings of dichotomous outcomes and 2 independent samples, as given in [4]. In practice, BBECT using UCB1-MPA will also be more ethical due to the larger number of patients allocated to ND, i.e. the arm with efficacy p_1. Note an important difference above compared to UCB1's proof [1, (6)] where the following derivation is used:

Δ	N for BBECT ($\beta = 0.9$)	N in static trial ($\beta = 0.9$, $1 - \alpha = 0.99$)
0.1	3222	1302
0.15	1290	578
0.2	668	326
0.25	400	208
0.3	262	145
0.35	184	107
0.4	134	82
0.45	102	65
0.5	80	53

Table 1: Minimum sample size required for power $\beta = 0.9$

$$\begin{aligned}
T_i(N) &\leq L_0 + \sum_{t>L_0+1}^{N} \sum_{s_t^1=1, s_t^2=t-s_t^1}^{s_t^1=t-(L_0+1)} \{\bar{X}_t^1 + c_t^1 \leq \bar{X}_t^2 + c_t^2 | s_t^1, s_t^2\} \\
E(T_i(N)) &\leq L_0 + \sum_{t>L_0+1}^{N} \sum_{s_t^1=1, s_t^2=t-s_t^1}^{s_t^1=t-(L_0+1)} P\{\bar{X}_t^1 \leq \mu_1 - c_t^1 | s_t^1\} \\
&+ P\{\bar{X}_t^2 \geq \mu_2 + c_t^2 | s_t^2\} \\
E(T_i(N)) &\leq L_0 + \sum_{t>L_0+1}^{N} \sum_{s_t^1=1, s_t^2=t-s_t^1}^{s_t^1=t-(L_0+1)} 2t^{-4} \; < \; L_0 + 2\zeta(3),
\end{aligned}$$

Where ζ is the Reimannian zeta function. Note in the above that, since it is a conditional probability, $P\{\bar{X}_t^1 \leq \mu_1 - c_t^1 | s_t^1\} \leq 2t^{-4} \cdot P(s_t^1)$. Hence we may write the above without conditioning on s_t^1, s_t^2 values to obtain a more compact expression. This change allows us a low L_0 and a suitable $N \geq 2L_0$ such that adverse probability is bounded by a small $1 - \beta$:

$$E(s_N^2) \;\; \leq \;\; L_0 + \sum_{t>L_0+1}^{N} P(\{\bar{X}_t^1 + c_t^1 \leq \bar{X}_t^2 + c_t^2\})$$

$$E(s_N^2) \leq L_0 + \sum_{t>L_0+1}^{N} 2t^{-4}$$

It is useful to reiterate what Theorem 1 implies: suppose that the first L_0 subjects are allocated to B, followed by which 1 subject is allocated to A. Then, the probability of even 1 more subject from the remaining $N-(L_0+1)$ subjects being allocated to arm B is less than β if BBECT with UCB1-MPA is used.

A note about calculation of N for static trial, [4], is also required here for the sake of completeness. The minimum sample size N is calculated based on a basic quantity named 'effect size' $E(p_1, p_2)$:

$$\begin{aligned}
E(p_1, p_2) &= \frac{p_2 - p_1}{\sqrt{p(1-p)}} \text{ where,} \\
p &= \frac{p_1 + p_2}{2} \\
N &= \lceil 4\left(\frac{z_{1-\alpha} + z_\beta}{E(p_1, p_2)}\right)^2 \rceil
\end{aligned}$$

The maximum N over a fine grid of possible (p_1, p_2), for each Δ, has been calculated and placed in the third column of Table 1 above. The N calculated in [10] for the comparisons below (e.g. Tables 4, 5) appear to be higher but the authors do not point there to any formula to infer N.

3. BBECT compared to Static and Ethico-Optimal Clinical trials

We implemented a simulation with $100,000$ trials where Bernoulli parameters p_1, p_2 were chosen randomly from $(0, 1)$ and $p_1 - p_2 = \Delta$, with Δ set to 0.2. The percentile values for number of patients in each trial, from a total of 668, that were allotted to treatment represented by p_1 were captured in the simulations. These are given in Table 2. This indicates that in 99% of the simulated BBECT runs, 73% or more of the patients were allocated to ND. Similarly, In 99% of the simulated BBECT runs where the effective Δ is such that $\Delta \in \{0.2, 0.3\}$, 77% or more of the patients were allocated to ND. This experiment models situations where Δ is not known accurately, but for inferring N it is sufficient to know a d such that $\Delta > d$ Note also how the ethical outcome is achieved: consider the $10-$th percentile mark when $\Delta = 0.2$, implying 90% of simulations have higher allocations to ND. We choose this level since the power of the test as designed according to Theorem 1 above is also pegged at 90%. For this particular level, reading off the table, note that $668 - 529 = 139$ is less than $0.5 \times 326 = 163$, where $N = 326$ is the maximum static trial size N for $\Delta = 0.2$ from Table 1. It may similarly be useful to compare the difference between N and the $10-$th percentile level for each Δ, with $\frac{N}{2}$ of a static trial, to verify the efficacy of BBECT with UCB1-MPA as a technique. This is done in Table 3, where notice the advantage for BBECT using UCB1-MPA for all $\Delta > 0.1$.

Percentile	$p_1 - p_2 = 0.2$, $p_1, p_2 \in (0,1)$	$p_1 - p_2 = \Delta$, s.t. $\Delta \in [0.2, 0.3]$
1	493	518
5	517	542
10	529	553
50	564	587
90	593	612
95	601	618
99	615	629

Table 2: Percentile threshold allotted to ND under BBECT using UCB1-MPA

Δ	10−th percentile of allocations to SOC (using BBECT)	$\frac{N}{2}$ from Table 1
0.10	658	651
0.15	267	289
0.20	139	163
0.25	84	104
0.30	55	73
0.35	39	54
0.40	28	41
0.45	22	33
0.50	17	27

Table 3: Allocation to SOC under BBECT using UCB1-MPA

We next tried a simulation with $1,000,000$ trials to estimate the level of significance α, often called the $p-$value of the test, by turning $\Delta = 0.0$ and modifying the algorithm slightly. We set $N = 668$ using Table 1, but set the criterion that ND would be declared as the better therapy - i.e. null hypothesis rejected - only if patients allotted to it exceed 60%. It is important to note here that situations where $\Delta = 0$ were handled differently. The p_A used in the simulation was $p_B + 0.05$ (with probab. 0.5), or alternatively $p_B := p_A - 0.05$ (with probab. 0.5). In none of the experiments was the expected response of either SOC or ND required, neither was information about variability.

However, the BBECT algorithm is distribution-dependent in the sense that information about approximate $\Delta \overset{\Delta}{=} \mu_1 - \mu_2$ is still required. Our simulations compare BBECT using UCB1-MPA within the setting of [10] which proposes an optimal adaptive rule for binary response trials. Taking $\Delta = p_B - p_A$, the situations compared are those where significance of trial (when $\Delta = 0$) as well as power of trial ($\Delta > 0$) is observed over 1 million simulations. Notice in Table 4 below that the 'error rate' when $\Delta > 0$ is better than the optimal adaptive rule, signifying more power than the 90% for which the optimal adaptive rule was designed. It is also the case that significance calculated empirically from BBECT simulations when $\Delta = 0$ is such that Type-1 error stays below 5%. Note that the N in these experiments - where power observed is more than 95%

- happens to be even lower than N calculated analytically for 90% power in Table 1.

Further, the BBECT algorithm also has lower 'treatment failure', i.e. a lesser number of subjects who did not recover, irrespective of whether they were allocated to the A or B arms. Table 5 presents the number of treatment failures with standard deviation (SD). The SD metric is observed as being higher in some pairs when BBECT with UCB1-MPA is employed. Note one deviation in Table 5, the last entry, which compares the best proposed algorithm in [5] viz. the Forward-Looking Gittins Index method (FLGI), with block size $b = 1$. While FLGI has a very favourable treatment failure rate compared to BBECT, the high standard deviation and lower power (0.33 as obtained from [5] compared to 0.96) are also worthy of note.

Also compare these results with the extensive simulation in [11] where 'Type-1 error inflation' for bandit-based methods occurs, incl. UCB1. This inflation is similar to the outcome we would obtain if we employed the original MPA rule of declaring arm as winner if more than 50% of pulls correspond to it. Notice also that C_α there has been tuned with simulations to suit the bandit algorithm, just as our mark of 60% is obtained here using simulations. For example, the standard C_α would be 1.645, but is adjusted to 2.068, 1.867, 1.701 ([11, Table 1]) for the algorithms UCB1, KL-UCB1 and Thomson Sampling, respectively. Notice also the power values of our algorithm in the lower half of Table 4 (all above 95%) whilst power values in [11, Table 1], even for the bandit methods, are less than 90%. Of these methods, KL-UCB has the best balance of statistical test's power and efficient allocation proportion (77% and 82%, respectively) both of which are unfavourable compared to our figures.

We also present the comparison with work in [12] where Bayesian bandit clinical trial algorithms are introduced for the Bernoulli case, like ours. There are 3 algorithms introduced there, using Gittins Index (GI), Whittle Index (WI), and a Randomized Gittins index (RGI), all having the advantage of being 'non-myopic', i.e. sensitive to the horizon left for the trial. The work compares p_A of 0.3 and p_B of 0.5 after deriving $N = 148$ for the static fixed allocation clinical trial. The BBECT figure for expected number of successes (ENS) was 65.95 at this N, compared to the approximately 70 (ratio $\frac{70}{148} = 0.473$) that GI and WI were able to achieve. However, note that statistical power of the GI, WI methods was very low, at $0.3 - 0.4$ compared to values greater than 90% for BBECT. If using $N = 588$ in this setting for BBECT, where N is calculated from Theorem 1 for power setting $1 - \beta = 0.8$, we have an ENS of 274.21 which at 0.491 exceeds the ratio that GI and WI achieve. Type-1 error evaluated using the method of perturbing p_A or p_B by 0.05 yields 0.056 for $N = 148$, whilst it is a low 0.01 for $N = 558$ (within limits for design criterion of 5%). Also compare the lower simulation-based N obtained to achieve or exceed the required power of 80% yet remain near the Type-1 error of 5%. This is a lower $N = 120$, whilst the ENS ratio obtained is however worse at 0.44.

p_A	p_B	N	BBECT with UCB1-MPA	Optimal Adaptive rule
0.10	0.10	200	0.00	0.04
0.30	0.30	200	0.04	0.05
0.50	0.50	200	0.05	0.04
0.70	0.70	200	0.04	0.04
0.90	0.90	200	0.00	0.04
0.10	0.20	526	0.96	0.89
0.10	0.30	162	0.98	0.89
0.10	0.40	82	0.99	0.89
0.40	0.60	254	0.99	0.89
0.60	0.90	82	0.98	0.90
0.70	0.90	162	0.98	0.91
0.80	0.90	526	0.96	0.90
0.0	0.545	116	1.00	0.77 (KL-UCB)

Table 4: Table comparing error-rate in [10] against BBECT with UCB1-MPA

p_A	p_B	N	BBECT with UCB1-MPA	Optimal Adaptive rule
0.10	0.20	526	436.4 (9.6)	443 (8.5)
0.10	0.30	162	122.0 (6.2)	126.2 (5.4)
0.10	0.40	82	55.2 (4.8)	58.5 (4.2)
0.40	0.60	254	113.3 (8.6)	124.4 (7.8)
0.60	0.90	82	14.1 (3.0)	19.3 (3.7)
0.70	0.90	162	24.7 (4.2)	31.5 (4.8)
0.80	0.90	526	68.1 (7.4)	78.3 (8.1)
0.0	0.545	116	38.7 (4.7)	71.11 (11.6)

Table 5: Treatment failures in [10] against proposed BBECT

4. Future Directions

The N proposed by Theorem 1 is higher when compared to a static clinical trial, and this requires registering more volunteers which is a challenge if the condition is rare. However, the Hoeffding bound used above produces an upper limit and alternative bounds exist whereby if $p_A - p_B \geq \Delta$ and $p_A - p_B \leq d$ are both known, then the bound is tighter. A closed form expression for the Type-1 error, even under the assumption of 'region of indifference' viz. a minor difference between p_A and p_B, would also be welcome for practitioners.

5. Acknowledgements

We thank Shivaram Kalyanakrishnan of IIT Bombay for reading an early draft of this paper and giving valuable inputs.

References

[1] P. Auer, N. Cesa-Bianchi, P. Fischer, Finite-time analysis of the multiarmed bandit problem, Machine Learning (47) (2002) 235–256.

[2] A. Biswas, R. Bhattacharya, Optimal response-adaptive allocation designs in phase III clinical trials: Incorporating ethics in optimality, Statistics and Probability Letters (81(8)) (2011) 1155–1160.

[3] S. Bubeck, R. Munos, G. Stoltz, Pure exploration in finitely-armed and continuous-armed bandits, Theoretical Computer Science (412(19)) (2011) 1832–1852.

[4] L. Sullivan, Power and sample size determination, https://sphweb.bumc.bu.edu/otlt/MPH-Modules/BS/BS704_Power/BS704_Power_print.html.

[5] S. F. Williamson, S. S. Villar, A response-adaptive randomization procedure for multi-armed clinical trials with normally distributed outcomes, Biometrics: Journal of the International Biometric Society (76(1)) (2020) 197–209.

[6] T. Lattimore, Regret analysis of the finite-horizon gittins index strategy formulti-armed bandits, Annual Conference on Learning Theory (49) (2016) 1–32.

[7] T. Lattimore, Optimally confident UCB: Improved regret for finite-armed bandits, arXiv preprint (arXiv:1507.07880) (2015).

[8] W. H. Press, Bandit solutions provide unified ethical models for randomized clinical trials and comparative effectiveness research, Proceedings of the National Academy of Sciences (106(52)) (2009) 22387–22392.

[9] C. Kaibel, T. Biemann, Rethinking the gold standard with multi-armed bandits: Machine learning allocation algorithms for experiments, Organizational Research Methods (24(1)) (2021) 78–103.

[10] W. F. Rosenberger, N. Stallard, A. Ivanova, C. N. Harper, M. L. Ricks, Optimal adaptive designs for binary response trials, Biometrics (57) (2001) 909–913.

[11] A. L. Smith, S. S. Villar, Bayesian adaptive bandit-based designs using the gittins index for multi-armed trials with normally distributed endpoints, Journal of Applied Statistics (45(6)) (2018) 1052–1076.

[12] S. S. Villar, J. Bowden, J. Wason, Multi-armed bandit models for the optimal design of clinical trials: Benefits and challenges, Statistical Science (30(2)) (2015) 199–215.

Research Office
Indian Institute of Management Kozhikode
IIMK Campus P. O.,
Kozhikode, Kerala, India,
PIN - 673 570
Phone: +91-495-2809237/ 238
Email: research@iimk.ac.in
Web: https://iimk.ac.in/faculty/publicationmenu.php