Acceleration of A/B/n Testing under time-varying signals

Jimmy Jin
Optimizely

Leo Pekelis
Opendoor

November 21, 2018

In 2018, Optimizely introduced Stats Accelerator - a multi-armed bandit designed to speed up A/B/n testing through dynamic traffic allocation while maintaining always valid inference, even in the presence of symmetric time variation [Optimizely, 2017].

The solution combines three ideas. First, an mSPRT (sequential hypothesis test) provides always-valid p-values and confidence intervals, allowing a user to continuously monitor results and detect true effects more efficiently [Johari et al., 2017]. Second, the test is modularized with a dynamic sampling allocation motivated by the lil'UCB algorithm of [Jamieson et al., 2014] designed for optimally fast detection of significant arms while maintaining FDR control [Jamieson and Jain, 2018]. Finally, a novel stratification-based augment to the mSPRT avoids bias from changing allocation when signals are symmetrically varying over time.

In this paper we highlight the last component of this solution. By dividing the timeline of the experiment into epochs of unchanging traffic allocation and averaging over those epochs, we maintain an unbiased estimate that integrates seamlessly into a sequential testing framework. We believe this method to be uniquely useful in practice because it

1. does not require estimation of the time variation (i.e. seasonality), and
2. does not require knowledge of the traffic allocation probabilities.

Our main contribution is a central limit theorem for the stratified estimator, suggesting that it should be able to be applied in most CLT-based methods such as the t-test and the mSPRT.

We provide full theoretical justification for its use in Optimizely's mSPRT in the presence of data dependent allocation and symmetric time variation, and show simulations demonstrating effectiveness of the full solution under a variety of dynamic allocation and temporal variation scenarios while retaining the performance improvements due to the dynamic traffic allocation.

1 Symmetric Time Variation and Simpsons Paradox

Consider a classic A/B testing scenario. Denote by $\{X_1, X_2, \ldots\}$ and $\{Y_1, Y_2, \ldots\}$ data arriving into the control and treatment arms of the experiment. A main goal of A/B testing is to determine whether or not the means of the data-generating distributions are equal.

A common underlying assumption is that means of the X_i's and Y_i's are constant over time, but in practice this is hardly the case, with $EX_i = \mu_C(t_i)$, $EY_j = \mu_T(t_j)$, and t_i indexing observations across time. One common form of time variation we see is **symmetric** time

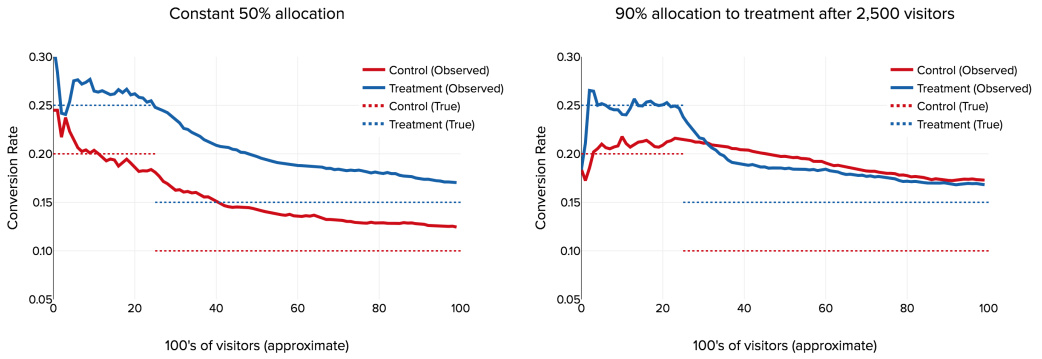

Figure 1: Simulations of two A/B tests of 10,000 visitors with and without Simpson's Paradox. In both cases, true conversion rates for the control and treatment begin at 0.20 and 0.25 respectively and drop to 0.10 and 0.15 respectively after 2,500 visitors. With constant allocation (left), the observed difference in conversion rates at 10,000 visitors is approximately equal to the true difference in conversion rates. With a change in traffic allocation at 2,500 visitors however, there is substantial bias in the observed difference in conversion rates at 10,000 visitors.

variation, where the time variation affects both arms of the experiment equally. For example, an additive form of this variation could be written $\mu_C(t) = \theta_C + f(t)$ and $\mu_T(t) = \theta_T + f(t)$.

Methods based on investigating differences in means are robust to this type of variation as long as the traffic distribution between the control and treatment is fixed, since the time variation term $f(t)$ is differenced out. However, when the traffic flow is adjusted in the presence of $f(t)$, the interaction between the two can cause bias, a phenomenon sometimes known as Simpson's Paradox.

The effect stems from the fact that most conversion rate estimates employ a running sum from the beginning of the experiment. Therefore if, for example, true conversion rates decrease equally for the control and treatment arms but relatively more of the low-converting traffic is sent to the control in the low conversion rate regime, the control will appear artificially depressed relative to the treatment arm. This scenario is illustrated in Figure 1.

Depending on the magnitude and direction of the bias, this can increase Type 1 or Type 2 error and is a concern for experimentation platforms who want to intelligently allocate traffic among different arms of experiments for their users. For example, Optimizely's dynamic sampling allocation can adjust traffic as often as hourly, and uses feedback from significance calculations to inform allocation updates. When updates line up with time variation underlying the experiment, results are exposed to potentially significant bias.

There has been increased interest in applying sequential algorithms to A/B testing in recent years.In addition to the mSPRT, first introduced in [Robbins, 1970] and explored for A/B Testing in [Johari et al., 2015] and [Johari et al., 2017], of particular note are the use of law-of-the-iterated-logarithm (LIL) bounds as in [Kaufmann et al., 2014]. With respect to time variation in the underlying parameters, most of the literature has fallen in the realm of regret minimization for *bandits with non-stationary rewards*, see for example [Besbes et al., 2014]. Our research generally lies in the same vein but is distinguished by a more statistical flavor

in aiming to explicitly control FDR.

2 Epoch Stats Engine

Our solution is based on the observation that bias due to the interaction of time variation and dynamic allocation cannot occur in periods when traffic allocation is fixed. To be precise, suppose there are $K(n)$ total epochs by the time the experiment has n points. Within each epoch k, denote by $n_{k,C}$ and $n_{k,T}$ the sample sizes of the control and treatment respectively, and by $\bar{X}_k$ and $\bar{Y}_k$ the sample means of the control and treatment respectively.

The epoch (stratified) estimator for the difference in means is

$$T_n = \sum_{k=1}^{K(n)} \frac{n_k}{n}(\bar{X}_k - \bar{Y}_k)$$

where $n_k = n_{k,C} + n_{k,T}$. Of particular concern here is performance of this estimator under dependence induced by a data-dependent allocation policy such as Stats Accelerator.

It turns out that plugging in a consistent variance estimate for T_n allows T_n to be used in almost all situations where the ordinary estimate $\bar{X}_n - \bar{Y}_n$ could be used. Letting $\theta := \theta_T - \theta_C$ denote the true difference in means, our first result is that for such a variance estimate $\hat{\text{Var}}(T_n)$ and data-dependent allocation rule, T_n satisfies a central limit theorem as $n \to \infty$:

> **Theorem 1.** Under mild assumptions on the behavior of the bandit governing traffic allocation changes, $\frac{T_n - \theta}{\hat{\text{Var}}(T_n)^{1/2}}$ is asymptotically normally distributed with $K(n)^{-1/2}$ rate of convergence.

The result follows from a fairly straightforward application of the martingale central limit theorem, see for example [Brown et al., 1971]. The "mild assumptions" on this bandit involve its influence on the conditional variance of the process $\{T_n : n \geq 1\}$ and are quite weak and can easily be shown to be satisfied by a fairly general class of bandits. We defer a detailed discussion of this proof and discussion of assumptions (and for the proposition below) to the appendix, but mention in passing that the variance for T_n can be well estimated by

$$\hat{\text{Var}}(T_n) = \sum_{k=1}^{K(n)} \left(\frac{n_k}{n}\right)^2 \left(\frac{\hat{\sigma}_C^2}{n_{k,C}} + \frac{\hat{\sigma}_T^2}{n_{k,T}}\right)$$

where $\hat{\sigma}_C^2, \hat{\sigma}_T^2$ are consistent estimates for the population variance of X_i and Y_i, the data-generating process for the control and treatment arms.

This asymptotic normality of Theorem 1 alone is enough to consider applicability in most CLT-based methods such as the t-test. However, the SPRT employed by Optimizely requires additional justification because of the need for approximations to the full likelihood of the observed data. In particular we show that

> **Corollary 1.** The mSPRT for $H_0 : \theta = 0$ has same asymptotic error control and run time as a 1-parameter mSPRT with mixture likelihood ratio, $\Lambda_n = \int \frac{\phi_{(\theta, \hat{s}_n)}(T_n)}{\phi_{(0, \hat{s}_n)}(T_n)} dF(\theta)$,

where $\hat{s}_n = \sqrt{\hat{\text{Var}}(T_n)}$, $\phi_{\theta,\sigma}$ is the density of a $N(\theta, \sigma^2)$ random variable, and F a mixing distribution with positive and continuous density, even under data dependent allocation and symmetric time variation.

The proof formalizes and generalizes Section 5 of [Johari et al., 2015]. As long as the allocation scheme is a stochastic process adapted to the filtration of data-generating process, sufficiency arguments along with the CLT for T_n and a similar result for $U_n = \sum_{k=1}^{K} \frac{n_k}{n}(\bar{X}_k + \bar{Y}_k)$ give an approximation to the full data likelihood,

$$L(\{X_1, \ldots, X_n\}, \{Y_1, \ldots, Y_n\}) = \phi_{\theta, s_n}(T_n) f(\{X_1, \ldots, X_n\}, \{Y_1, \ldots, Y_n\}) + o(1)$$

where f is not a function of θ. Asymptotic convergence of the mixture likelihood ratio immediately follows from asymptotic convergence of $\hat{s}_n$. Error control and run time performance require an examination of the rates of convergence of the above in the context of classical results for the mSPRT, for example [Lai et al., 1977].

3 Performance Under Bandit Policy

We validated the performance of Epoch Stats Engine under Simpson's Paradox conditions using simulated, time-varying data. Specifically, we generated 600,000 draws from 7 Bernoulli arms with one control and 6 variants, with 1 truly higher-converting arm and all others converting at the same rate as the control. The conversion rate for the control starts out at 0.10 and then undergoes cyclic time variation rising as high as 0.15.

On top of this data we observed the false discovery rate (FDR) and true discovery rate (TDR) over time for each of four allocation modes, averaged over 1000 simulations: under (bandit-driven) dynamic allocation with (ideas 1+2+3 in intro) and without (1+2 only) the epoch-based estimator enabled, and under a fixed (equal) traffic allocation policy with (1+3) and without the epoch-based estimator (1 only).

In the FDR comparison, we observe the non-epoch bandit policy shows up to 150% exceedance of the configured FDR level (0.10) while the bandit policy scenario with epochs shows proper control.

In the TDR comparison, we observe a large gap between the bandit allocation runs and the fixed allocation runs reflecting the fact that speedup due to bandit allocation is preserved under Epoch Stats Engine. Furthermore, we observe no significant difference in time to significance between the epoch and non-epoch scenarios under fixed allocation while we observe a small gap in time to significance between the epoch and non-epoch scenarios under the bandit policy. This gap can be ascribed to the fact that the non-epoch Stats Engine running under dynamic allocation experiences high sensitivity to time variation. Higher TPR is paid for with higher FDR.

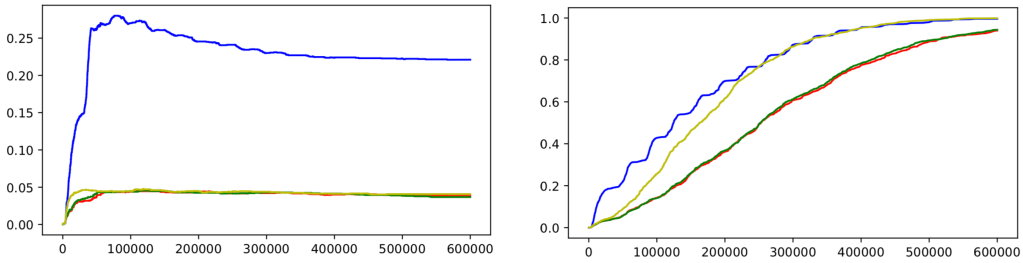

Figure 2: FDR (left) and TDR (right) over time. Both plots calculated over 1000 trials of 5 and 7 Bernoulli arms containing one true winner, under a configured significance (FDR) level of 0.10. Blue/yellow lines = bandit allocation without/with Epoch Stats Engine (respectively), red/green lines = fixed (equal) traffic allocation without/with Epoch Stats Engine (respectively).

References

[Besbes et al., 2014] Besbes, O., Gur, Y., and Zeevi, A. (2014). Stochastic multi-armed-bandit problem with non-stationary rewards. In Advances in neural information processing systems, pages 199–207.

[Bolthausen et al., 1982] Bolthausen, E. et al. (1982). Exact convergence rates in some martingale central limit theorems. The Annals of Probability, 10(3):672–688.

[Brown et al., 1971] Brown, B. M. et al. (1971). Martingale central limit theorems. The Annals of Mathematical Statistics, 42(1):59–66.

[Haeusler, 1988] Haeusler, E. (1988). On the rate of convergence in the central limit theorem for martingales with discrete and continuous time. The Annals of Probability, pages 275–299.

[Hall, 1988] Hall, P. (1988). On the effect of random norming on the rate of convergence in the central limit theorem. The Annals of Probability, pages 1265–1280.

[Jamieson and Jain, 2018] Jamieson, K. and Jain, L. (2018). A Bandit Approach to Multiple Testing with False Discovery Control. ArXiv e-prints.

[Jamieson et al., 2014] Jamieson, K., Malloy, M., Nowak, R., and Bubeck, S. (2014). lilucb: An optimal exploration algorithm for multi-armed bandits. In Conference on Learning Theory, pages 423–439.

[Johari et al., 2017] Johari, R., Koomen, P., Pekelis, L., and Walsh, D. (2017). Peeking at a/b tests: Why it matters, and what to do about it. In Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pages 1517–1525. ACM.

[Johari et al., 2015] Johari, R., Pekelis, L., and Walsh, D. J. (2015). Always Valid Inference: Bringing Sequential Analysis to A/B Testing. ArXiv e-prints.

[Kaufmann et al., 2014] Kaufmann, E., Cappé, O., and Garivier, A. (2014). On the complexity of a/b testing. In Conference on Learning Theory, pages 461–481.

[Lai et al., 1977] Lai, T., Siegmund, D., et al. (1977). A nonlinear renewal theory with applications to sequential analysis i. The Annals of Statistics, 5(5):946–954.

[Mourrat et al., 2013] Mourrat, J.-C. et al. (2013). On the rate of convergence in the martingale central limit theorem. Bernoulli, 19(2):633–645.

[Optimizely, 2017] Optimizely (2017). Accelerating experimentation through machine learning. `https://blog.optimizely.com/2017/10/18/stats-accelerator/`.

[Robbins, 1970] Robbins, H. (1970). Statistical methods related to the law of the iterated logarithm. The Annals of Mathematical Statistics, 41(5):1397–1409.

Appendix

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Define $I = \{0, 1, \dots\}$ as a discrete time index, X_n and Y_n independent, square-integrable random variables for $n \in I$ with $EX_n = \mu_C(n)$ and $EY_n = \mu_T(n)$, defined as above. An allocation scheme is a process $\delta_n : \Omega \to \{C, T\}$ such that the observed data in an experiment up to time n is $D_n = \{X_s\}_{s \in S_n(C)} \cup \{Y_s\}_{s \in S_n(T)}$ where $S_n(z) = \{s \leq n \ : \ \delta_s = z\}$. Note n is the number of total points in the experiment.

For all that follows, we consider the filtration $\{\mathcal{F}_n : n \geq 1\}$ defined by $\mathcal{F}_n = \sigma(D_s, s \leq n)$ and assume that $\{\delta_n : n \geq 1\}$ is adapted to this filtration.

Some more epoch-specific notation is required. Assume allocation decisions are constant on epochs of constant length n_K, i.e. δ_{n+j} is j-predictable for $j = 1, ..., n_K$ when $n \pmod{n_K} = 0$. Let $K(n)$ be the total number of epochs at the time during which the experiment has n total points, and let k index over the epochs from 0 to $K(n)$.

Proof of Theorem 1

In all that follows, we will show limiting results in terms of the number of epochs going to infinity. Therefore, we will need to reindex some quantities by epoch: Define $n(i)$ to be the total number of points in the experiment by the end of epoch i. We introduce the following epoch-indexed quantities:

1. $E_i = \frac{n_i}{n}(\bar{X}_i - \bar{Y}_i - \theta)$
2. $\{\mathcal{G}_i : 1 \leq i \leq K(n)\}$ be the sub-filtration of $\{\mathcal{F}_n : n \geq 1\}$ defined by $\mathcal{G}_i = \mathcal{F}_{n(i)}$
3. $\bar{T}_i = \sum_{j=1}^{i} E_i$

We start with a few simplifying assumptions.

$$\mathbb{E}X_i^4, \mathbb{E}Y_i^4 < \infty \tag{1}$$

$$\text{Var}(X_i) := \sigma_X^2 \text{ and } \text{Var}(Y_i) := \sigma_Y^2 \text{ are known} \tag{2}$$

$$P(\delta_n = C) \in (0, 1) \text{ for all } n \tag{3}$$

$$\{n_i : 1 \leq i \leq K(n)\} \text{ is predictable with respect to } \{\mathcal{G}_i : 1 \leq i \leq K(n)\} \tag{4}$$

$$K(n) \to \infty \text{ as } n \to \infty \tag{5}$$

(1) is needed only for the rate of convergence of the CLT. In practice this is not a restrictive condition since most random variable encountered in the wild are effectively bounded.

(2) is a simplifying assumption which we accommodate in practice by replacing the true variance with the sample variance. Due to strong concentration of the sample variance around the true variance and other considerations (see, e.g. [Hall, 1988]), this should not have a significant effect on the true rate of convergence.

(3) is a weak assumption that is satisfied by almost all bandits we know, and can often by forced by setting some $\epsilon > 0$ amount of forced exploration into the algorithm. (4) is a simplifying assumption which does not greatly affect the structure of the argument going forward and, moreover, is satisfied in practice due to regular traffic density patterns in most experiments. Finally, (5) allows us to connect the result as $n \to \infty$ in terms of the number of epochs $\to \infty$, and involves no loss of generality since the result with a bounded number of epochs follows from Slutsky combined with the ordinary CLT on a single epoch with size going to infinity.

To begin the proof we start by noting that under these conditions, the sequence $\{\bar{T}_i : 1 \leq i \leq K(n)\}$ is a $\mathcal{G}_i$-martingale, and so the increments $\{E_i : 1 \leq i \leq K(n)\}$ are orthogonal.

Define $\sigma_i^2 = \mathbb{E}(E_i^2|\mathcal{G}_{i-1})$. Use this to define the conditional variance $V_i^2 = \sum_{j=1}^{i} \sigma_i^2$ and its expectation $s_i^2 = \mathbb{E}V_i^2$ for each epoch $i = 1, \ldots, K(n)$. In what follows we assume that the allocation probability sequence is "well-behaved" in the sense of

$$V_{K(n)}^2 s_{K(n)}^{-2} \xrightarrow{\mathbb{P}} 1 \text{ as } n \to \infty \tag{6}$$

The significance of this condition is two. First, it expresses a certain stability to the changes in the allocation probabilities over time. To see this, note that by orthogonality

$$\text{Var}(\bar{T}_{K(n)}) = \sum_{i=1}^{K(n)} \text{Var}(E_i)$$

Clearly, the $\text{Var}(E_i)$'s are not iid, so the requirement for V_n^2 to satisfy a weak law of large numbers is non-trivial. Secondly, the norming used in practice is not exactly $\text{Var}(\bar{T}_{K(n)})^{1/2}$ but rather closer to the conditional quantity $V_{K(n)}^{1/2}$, since we observe exactly the allocation probabilities conditional on previous epochs. Therefore this requirement also codifies the requirement necessary to replace the unconditional norming $s_{K(n)}$ (in the statement below) by $V_{K(n)}$ in practice.

In general, we find this condition weak enough to be satisified by a large class of bandit algorithms, including Accelerate Learnings, and accepting this we are now in a position to state the CLT.

Lemma 1. If (6) holds, then

$$\frac{\bar{T}_{K(n)}}{s_{K(n)}} \xrightarrow{d} N(0,1) \quad \text{as } n \to \infty$$

Proof. We apply a martingale CLT to $\bar{T}_i := \sum_{j=1}^{i} E_j$ by showing a Lindeberg condition to guarantee that contributions to the variance from the martingale increments are not too large.

Following Lemma 2 in [Brown et al., 1971] we consider the equivalent conditional Lindeberg condition

$$V_{K(n)}^{-2} \sum_{i=1}^{K(n)} \mathbb{E}\left(E_i^2 \cdot \mathbb{1}\left\{|E_i| \geq \epsilon s_{K(n)}\right\} \mid \mathcal{G}_{i-1}\right) \xrightarrow{\mathbb{P}} 0 \quad \text{for all } \epsilon > 0 \tag{7}$$

Denote by $F_{n,i,\epsilon}$ the event $\{|E_i| \geq \epsilon s_{K(n)}\}$. Algebraic manipulation of (7) gives

$$V_{K(n)}^{-2} \sum_{i=1}^{K(n)} \mathbb{E}\left(E_i^2 \cdot \mathbb{1}\left\{|E_i| \geq \epsilon s_{K(n)}\right\} \mid \mathcal{G}_{i-1}\right) = \frac{\sum_{i=1}^{K(n)} \mathbb{E}\left[n_i^2(\bar{X}_i - \bar{Y}_i - \theta)^2 \cdot \mathbb{1}\left\{F_{n,i,\epsilon}\right\} \mid \mathcal{G}_{i-1}\right]}{\sum_{i=1}^{K(n)} \mathbb{E}\left[n_i^2(\bar{X}_i - \bar{Y}_i - \theta)^2 \mid \mathcal{G}_{i-1}\right]}$$

Analyzing the event $F_{n,i,\epsilon}$ we see that

$$\begin{aligned}
\mathbb{E}\left(\mathbb{1}\left\{F_{n,i,\epsilon}\right\} \mid \mathcal{G}_{i-1}\right) &= \mathbb{P}(E_i^2 \geq \epsilon^2 s_n^2 \mid \mathcal{G}_{i-1}) \\
&= \mathbb{P}\left(\frac{n_i^2(\bar{X}_i - \bar{Y}_i - \theta)^2}{\sum_{i=1}^{K(n)} \mathbb{E}(n_i^2(\bar{X}_i - \bar{Y}_i - \theta)^2)} \geq \epsilon^2 \;\middle|\; \mathcal{G}_{i-1}\right) \\
&\leq \frac{\mathbb{E}(n_i^2(\bar{X}_i - \bar{Y}_i - \theta)^2)}{\epsilon^2 \sum_{i=1}^{K(n)} \mathbb{E}(n_i^2(\bar{X}_i - \bar{Y}_i - \theta)^2)}
\end{aligned}$$

Now since the allocation probabilities $\in (0,1)$, there exists constants C_1, C_2 not depending on n_i or n satisfying

$$n_i C_1 \leq \mathbb{E}(n_i^2(\bar{X}_i - \bar{Y}_i - \theta)^2) \leq n_i C_2 \tag{8}$$

Therefore for some $C := C(\epsilon)$, we have

$$\mathbb{P}(F_{n,i,\epsilon} \mid \mathcal{G}_{i-1}) \leq C \cdot \frac{n_i}{\sum_{i=1}^{K(n)} n_i} \tag{9}$$

so that $\mathbb{P}(F_{n,i,\epsilon} \mid \mathcal{G}_{i-1}) \to 0$ uniformly as $n \to \infty$, which implies (7). □

For Optimizely's mSPRT however, more results are required for the rate of convergence due to the necessity of approximations to the full likelihood of the observed data.

In particular, we need a rate of convergence for Lemma 1 around $K(n)^{-1/2}$. As a first pass we will lean on the extensive literature on rates of convergence for martingale CLT (see e.g. [Mourrat et al., 2013], [Haeusler, 1988], and in particular [Bolthausen et al., 1982]) to show that this is satisfied with some additional conditions on higher moments of the within-epoch estimates E_i:

Lemma 2. Let Φ denote the standard normal CDF. Define

$$D_{K(n)} = \sup_{t \in \mathbb{R}} \left| \mathbb{P}(\bar{T}_{K(n)}/s_{K(n)} \leq t) - \Phi(t) \right|$$

Suppose that in addition to the existence of fourth moments (1), the weak law of (6) is strengthened in that, for some $p > 1$:

$$\sup_{1 \le i \le K(n)} \|\mathbb{E}(E_i^2 \mid \mathcal{G}_{i-1}) - \mathbb{E}(E_i^2)\|_p = O(K(n)^{-1/2}) \tag{10}$$

and furthermore that the conditional third moments of the martingale increments $\{E_i : 1 \le i \le K(n)\}$ become more and more nonrandom as $K(n) \to \infty$ in the sense of:

$$\sup_{1 \le i \le K(n)} \|\mathbb{E}(E_i^3 \mid \mathcal{G}_{i-1}) - \mathbb{E}(E_i^3)\|_\infty = O(1/\log K(n)) \tag{11}$$

Then there exists a constant C such that $D_{K(n)} \le C \cdot K(n)^{-1/2}$ for any $n \ge 1$.

Proof. Immediate from Theorem 4 of [Bolthausen et al., 1982]. □

Combining Lemmas 1 and 2 along with Slutsky to incorporate a consistent estimate $\hat{\mathrm{Var}}(T_n)^{1/2}$ in place of s_n completes the proof of Theorem (1).

Proof of Corollary 1

For this argument we will specialize to the case of binomial X_n, Y_n. With $n_{k,z} = |S_k(z)| - |S_{k-1}(z)|$, $z \in \{T, C\}$, $\bar{f}_{k,z} = \sum_{(k-1)n_k < s \le kn_k} f(s)\mathbb{E}\left[\frac{\mathbb{1}\{\delta_s = z\}}{n_{k,z}}\right]$, we make the technical assumption that

$$\lim_{n \to \infty} \frac{1}{K(n)} \bar{f}_{k,T} - \bar{f}_{k,C} = 0$$

which is satisfied by, for example, $f(s)$ decaying to 0 itself or periodic and δ_n a binomial random variable with parameter $d_n \in (0, 1)$ (itself a random process).

Denoting $L_n = L(D_n|\Theta_n)$, the likelihood of observed experiment data, where $\Theta_n = \{\theta, \tau, f(s), s \le n, n_K\}$, $\theta = \theta_T - \theta_C$, $\tau = \theta_T + \theta_C$, we prove

Proposition 1.

$$L_n = \phi_{\theta, \hat{s}_n}(T_n) L(D_n | T_n, \Theta_n \setminus \theta) + O(K(n)^{-1/2})$$

for $\hat{s}_n$ a root-$K(n)$ consistent variance estimate of $\sqrt{\mathrm{Var}(T_n)}$.

Proof. Let $\bar{X}_k$, $\bar{Y}_k$ denote sample averages of Xs and Ys observed in epoch k, and $(\delta)_n = (\delta_1, ..., \delta_n)$

$$\begin{aligned}
&L(D_n | \bar{X}_k, \bar{Y}_k, k \le K(n), \Theta_n) \\
&= \sum_{(\delta)_n} L(D_n | \bar{X}_k, \bar{Y}_k, k \le K(n), \Theta_n, (\delta)_n) L((\delta)_n | \bar{X}_k, \bar{Y}_k, k \le K(n), \Theta_n) \\
&= \sum_{(\delta)_n} L(D_n | \bar{X}_k, \bar{Y}_k, k \le K(n), \Theta_n \setminus \{\theta, \tau\}, (\delta)_n) L((\delta)_n | \bar{X}_k, \bar{Y}_k, k \le K(n), \Theta_n \setminus \{\theta, \tau\})
\end{aligned} \tag{12}$$

where the last line follows from the fact that $\bar{X}_k$ and $\bar{Y}_k$ are sufficient for θ and τ for any fixed allocation decision, and non-predictability of δ_n. Concretely, fixing $(\delta)_n$ along with $\bar{X}_k$ contains $\#\{X_s = 1, s \in S_nC \cap \{(k-1)n_K, \ldots, kn_K\}\}$, and so the randomness remaining is the position of values of X_s, parameterized by $f(s)$ only. Note, we make an assumption here that $S_n(C)$ and $S_n(T)$ are non-empty, which holds with probability approaching 1 by assumption (3) of Theorem 1.

Next,

$$L(\bar{X}_k, \bar{Y}_k, k \leq K(n)|\Theta_n) = L(\bar{X}_k, \bar{Y}_k, k \leq K(n)|T_n, U_n, \Theta_n \setminus \{\theta, \tau\})\ L(T_n, U_n|\Theta_n). \quad (13)$$

The 1st factor on the right hand side does not depend on θ or τ via a similar sufficiency argument as above. The vector (T_n, U_n) has limiting distribution (Z_T, Z_U), a bi-variate normal with

$$EZ_T = \theta\ ;\ EZ_U = \tau + \lim_{n\to\infty} \frac{1}{K(n)} \bar{f}_{k,T} + \bar{f}_{k,C}.$$

We know the limit in EZ_U exists due to CLT lemmas, and does not depend on θ since

$$\bar{f}_{k,T} + \bar{f}_{k,C} = \sum_{(k-1)n_K < s \leq kn_K} w(s)f(s)$$

and $w(s)$ is symmetric in X_s and Y_s. Note the "technical assumption" gives the first expectation. The result follows from approximating the co-variance matrix of (Z_T, Z_U) with root-$K(n)$ consistent estimates. □