Peeking with PEAK: Sequential, Nonparametric Composite Hypothesis Tests for Means of Multiple Data Streams

Brian Cho [1] **Kyra Gan** [1] **Nathan Kallus** [1]

Abstract

We propose a novel nonparametric sequential test for composite hypotheses for means of multiple data streams. Our proposed method, *peeking with expectation-based averaged capital* (PEAK), builds upon the testing-by-betting framework and provides a non-asymptotic α-level test across any stopping time. Our contributions are two-fold: (1) we propose a novel betting scheme and provide theoretical guarantees on type-I error control, power, and asymptotic growth rate/e-power in the setting of a single data stream; (2) we introduce PEAK, a generalization of this betting scheme to multiple streams, that (i) avoids using wasteful union bounds via averaging, (ii) is a test of power one under mild regularity conditions on the sampling scheme of the streams, and (iii) reduces computational overhead when applying the testing-as-betting approaches for pure-exploration bandit problems. We illustrate the practical benefits of PEAK using both synthetic and real-world HeartSteps datasets. Our experiments show that PEAK provides up to an 85% reduction in the number of samples before stopping compared to existing stopping rules for pure-exploration bandit problems, and matches the performance of state-of-the-art sequential tests while improving upon computational complexity.

1. Introduction

Sequential experiments have important applications in various fields to optimize decision-making, including better content recommendation (Bouneffouf et al., 2012), personalized learning (Clement et al., 2015; Cai et al., 2021), and optimized digital interventions (Figueroa et al., 2021; Forman et al., 2019; Liao et al., 2020; Piette et al., 2022; Trella et al., 2022; 2023; Yom-Tov et al., 2017). Providing informative and anytime-valid feedback during sequential experiments could potentially lead to cost reduction through early stopping (Liang and Bojinov, 2023) and improved experiment outcomes. For example, in digital interventions, prompt "expert" feedback has been shown to increase user engagement, resulting in better treatment outcomes (Sharpe et al., 2017; Yardley et al., 2016). In this work, we aim to provide anytime-valid inference in sequential experiments for a task of interest (e.g., identifying the best arm/treatment in the bandit/reinforcement learning setting), where data collection adapts based on previously observed outcomes. The problem necessitates special inferential tools since naively repeatedly applying conventional inference such as t-tests at every time point will always at some point detect an effect even when none exists.

In settings where sample sizes are limited and multiple arms are to be compared, the ideal testing tools for sequential and possibly adaptive experiments should fulfill the following desiderata:

(**a**) Providing a non-asymptotic α-level test across any stopping time under nonparametric assumptions;

(**b**) Efficiently rejecting hypotheses that are incorrect across all potential distributions that satisfy our nonparametric assumptions;

(**c**) Enabling joint, composite hypothesis tests on multiple streams of data in a computationally tractable manner.

The first tools for non-asymptotic, α-level tests were proposed by Wald (1945), but these approaches rely on likelihood ratios, which may not be applicable in composite and nonparametric settings. To circumvent such issues, modern tests satisfying desideratum (**a**) use the notion of e-processes (Grünwald et al., 2023). It generalizes likelihood ratios to the nonparametric, composite setting (Ramdas et al., 2023) and offers an alternative to p-values that is more suitable for testing under optional stopping and continuation (Vovk and Wang, 2021; Ramdas et al., 2022).

Various sequential testing methods focus on asymptotic guarantees (Bibaut et al., 2022; Robbins and Siegmund, 1970; Waudby-Smith et al., 2021; Woong Ham et al., 2023), where approximately valid (and powered) inference only starts after sufficiently large sample sizes. While these methods appeal to other desiderata such as asymptotic calibration of

[1]Department of ORIE, Cornell Tech, New York, NY, USA. Correspondence to: Brian Cho <bmc233@cornell.edu>.

Proceedings of the 41st International Conference on Machine Learning, Vienna, Austria. PMLR 235, 2024. Copyright 2024 by the author(s).

type-I errors and optimal power, they may not be appropriate when inference is desired for small sample sizes, reflected in our desideratum (**a**).

Recent works utilizing e-processes for non-asymptotic anytime-valid inference fall into two groups. The first involves Chernoff-based approach (Howard et al., 2021a), which relies on sub-ψ assumptions (such as sub-gaussian/sub-bernoulli) combined with peeling (Cappé et al., 2013) or mixture techniques (de la Peña et al., 2004). The second involves the testing-by-betting framework (Waudby-Smith and Ramdas, 2023). In the bounded observation setting, both empirical (Waudby-Smith and Ramdas, 2023) and theoretical (Shekhar and Ramdas, 2023) evidence indicate that the latter outperforms the former. However, these approaches suffer from computational inefficiency – computing confidence sequences and conducting composite tests require performing a grid search over the hypothesis space of means, which grows in dimension as the number of data streams increases. Thus, while the current testing schemes proposed in Waudby-Smith and Ramdas (2023) satisfy both desiderata (**a**), (**b**), they fall short of satisfying (**c**).

Contributions. In this paper, we provide a novel nonparametric sequential test for composite hypotheses on bounded means that aim to achieve desiderata (**a**)-(**c**). We build upon the testing-by-betting framework and establish theoretical guarantees for our procedure. Our contributions are two-fold: (1) we propose a novel betting scheme, and provide theoretical guarantees on type-I error control, power, and asymptotic growth rate/e-power that have not been shown for other adaptive testing-by-betting methods in the setting of a single data stream; (2) we introduce PEAK, a generalization of this betting scheme to multiple streams, that (i) avoids using wasteful union bounds via averaging, (ii) is a test of power one under mild regularity conditions on the sampling scheme of the streams, and (iii) reduces computational overhead when applying the testing-by-betting approaches for pure-exploration bandit problems.

Outline. In Section 2, we define e-values and introduce the concept of e-power, which serves as the natural analogue of power in settings involving optional stopping and continuation. Next, we present the testing-by-betting framework by Waudby-Smith and Ramdas (2023). In Section 3, we introduce our testing procedure in the single arm/stream setting, and provide theoretical results demonstrating that this test (i) controls type-I error, (ii) is of power 1, and (iii) forms convex/continuous confidence intervals. In addition, we quantify the asymptotic suboptimality of our test. In Section 4, we provide a new test based on *averaging evidence across arms*, and show that this testing approach maintains computational tractability while strictly dominating a union-bound approach. In Section 5, we numerically test our sequential testing scheme for both single-arm and multi-arm settings on both simulated data and the mHealth HeartSteps dataset (Liao et al., 2020), demonstrating the efficacy of our approach.

2. Problem Formulation and Related Work

We consider the setting where we have W arms/streams, indexed by $a \in \{1, ..., W\} \equiv [W]$, with each arm corresponding to an unknown, independent distribution P_a over $[0, 1]$. We denote $\mu = [\mu_1, ..., \mu_W]$ as the means of the distributions $\{P_a\}_{a\in[W]}$. Let $X_t \in [0, 1]$ denote the outcome that we observe at time t. The data is collected in an online, sequential fashion: at time $t \in \mathbb{N}$, we observe pair $(A_t, X_t) \in [W] \times [0, 1]$. The choice of arm A_t is sampled from a (potentially adaptive) sampling policy π_t, which may be unknown. The sampling function $\pi_t \in \Delta^W$ is assumed to be $\mathcal{F}_{t-1}$-measurable, where Δ^W is the probability simplex over $[W]$. For the entirety of this paper, we use $\mathcal{F}_0, \mathcal{F}_1, \ldots$ as the canonical filtration, i.e., $\mathcal{F}_{t-1} = \sigma((A_i, X_i)_{i=1}^{t-1})$, with $\mathcal{F}_0$ as the empty sigma field.

We assume the nonparametric setting, making no further assumptions about the data-generating process (DGP), allowing for continuous and discrete distributions over $[0, 1]$. For each hypothesis $m_a \in [0, 1]$ of arm a, let $\mathcal{P}(m_a)$ denote the set of all distributions on $[0, 1]$ with mean m_a. Each hypothesis vector $m = (m_1, ..., m_W) \in [0, 1]^W$ denotes the composite hypothesis $\cap_{a\in[W]}\{P_a \in \mathcal{P}(m_a)\}$, in which we fix the distribution means at m but otherwise allow for arbitrary distributions over $[0, 1]$ for each arm.[1] We use $R \subseteq [0, 1]^W$ to denote the set of composite hypotheses of means (corresponding to regions of the hypothesis space) that we are interested in testing at time $t \in \mathbb{N}$. We defer further discussion on R to Section 2.2.

A sequential test $T_t(R)$ for a composite hypothesis R is a binary-valued $\mathcal{F}_t$-measurable variable, with value 1 indicating rejection of R. Our goal is to develop a computationally tractable sequential test $T_t(R)$ that maintains a α-level type-I error rate, i.e., whenever $P_a \in \mathcal{P}(m_a)$ for all a for some $m \in R$, we have

$$\mathbb{P}(\exists t \in \mathbb{N} : T_t(R) = 1) \leq \alpha.$$

The test we developed will follow from the e-value and the testing-by-betting frameworks (Waudby-Smith and Ramdas, 2023). We formally introduce these concepts below.

2.1. e-Processes and Testing-by-Betting Framework

The notation of e-values, first introduced by Ville (1939), can be viewed as an alternative to p-values under optional

[1] The minimal assumption necessary on the arm distributions for all results in this document can be found in Section 2, paragraph 2 of Waudby-Smith and Ramdas, 2023. It only requires stationary of the expectations for each arm conditional on the previous history.

stopping and continuation (Grünwald et al., 2023). An e-value of hypotheses $\mathcal{H}$ is defined as a non-negative random variable, E, that satisfies the following:

$$\mathbb{E}_P[E] \leq 1, \forall P \in \mathcal{H}. \tag{1}$$

The resulting α-level test from an e-value is simply thresholding E against $1/\alpha$, with rejection if $E \geq 1/\alpha$. In our setting, the e-value generalizes to the e-process:

Definition 1 (e-process, Definition 1 of Grünwald et al. 2023)**.** *We say that E_t is a conditional e-variable (w.r.t. $(\mathcal{F}_t)_{t\in\mathbb{N}}$) for hypothesis $\mathcal{H}$ if it is a $\mathcal{F}_t$-measurable nonnegative random variable that satisfies $\mathbb{E}_P[E_t|\mathcal{F}_{t-1}] \leq 1$ a.s. for all $P \in \mathcal{H}$. The product of conditional e-variables $K_t = \prod_{i=1}^t E_i$ is an e-process, meaning that for any potentially infinite stopping time τ, K_τ is an e-value.*

If K_t is an e-process for hypothesis $\mathcal{H}$, the α-level sequential test is to reject $\mathcal{H}$ if K_t ever exceeds the threshold $1/\alpha$. The design of e-value-based tests directly leads to the interpretation of K_t as *evidence against the null*: we reject $\mathcal{H}$ when K_t is sufficiently large. The analogue of power for an e-process-based test, called the growth rate or e-power (Vovk and Wang, 2024), corresponds to this intuition.

Definition 2 (e-power)**.** *For an e-value E concerning the null hypothesis P, the growth rate of E with respect to an alternative hypothesis Q is given by $\mathbb{E}_Q[\log(E)]$.*

In the case of e-process K_t, the e-power is the sum of the expected log values for each conditional e-variable E_t. Maximizing the e-power intuitively means maximizing the evidence against the null when the null is misspecified; under certain conditions, this directly corresponds to minimizing the expected stopping time in the i.i.d. single arm setting (Grünwald et al., 2023; ter Schure et al., 2023). Moreover, for simple null hypothesis P and simple alternative hypothesis Q where Q is absolutely continuous with respect to P, the likelihood ratio dQ/dP maximizes the e-power (Lemma 2.1 of Vovk and Wang, 2024), aligning directly with the same notion of power in classical parametric sequential tests (Wald and Wolfowitz, 1948).

Testing-by-Betting Framework. In the single-arm setting for bounded means, with null hypothesis $R = m$, the test we develop can be seen as a special case of the more general testing-by-betting framework (Waudby-Smith and Ramdas, 2023), which constructs tests based on *sizing bets*. Specifically, the testing-by-betting framework constructs the e-process $M_t(m)$ as follows:

$$M_t(m) = \prod_{i=1}^{t} \left(1 + \lambda_t(m)(X_i - m)\right), \tag{2}$$

for some $\lambda_t(m) \in [-1/(1-m), 1/m]$, where $\lambda_t(m)$ is $\mathcal{F}_{t-1}$-measurable.[2] The sequence $\lambda_t(m)$ can be interpreted as bets against the belief of whether μ is above m, with the sign indicating the direction of the belief and the magnitude indicating the strength of the belief (i.e., how much of our current evidence $M_t(m)$ we are willing to put at stake). We discuss our novel choice of $\lambda_t(m)$ in Section 3, using the single-arm case as a guiding example.

The proposed testing-by-betting methods (i.e., choice of $\lambda_t(m)$) in Waudby-Smith and Ramdas (2023) suffer from computational inefficiency when generalized to the multi-armed case. In the single-stream setting, these tests require a search over the hypothesis space $[0, 1]$, using grid search to rule out hypotheses m sequentially. In the multi-stream setting, the size of the grid search grows exponentially with the dimension W. Our choice of sequential test, even with multiple streams, removes the need for grid search and allows various convex optimization approaches to be used.

2.2. Hypothesis Testing of Means for Multiple Streams

In the multi-stream setting, we are interested in testing hypotheses on the joint hypothesis space $[0, 1]^W$. A common hypothesis of interest is whether arm $a \in W$ is at least as large as all other arms, often referred to as the best arm identification (BAI) problem (Audibert et al., 2010), i.e.,

$$R_{\text{BAI}}(a) = \{m \in [0, 1]^W : m_a \geq m_i, \forall i \in [W] \setminus a\}. \tag{3}$$

Other potential hypotheses of interest include determining if each stream's mean is above/below a certain threshold (Kano et al., 2019). For such objectives, the hypotheses form convex regions of the hypothesis space $[0, 1]^W$. In Section 4.2, we show our test enables computationally tractable testing for such convex regions.

Another work that directly considers joint sequential tests across arms is Kaufmann and Koolen (2021), which uses a mixture-based approach to construct e-processes for a single arm and combines this evidence using the product of evidence across arms. While their objective aligns with ours, the setting of their work differ drastically: the parameters of the data streams are assumed to belong to 1-dimensional exponential families, rather than our nonparametric setting.

3. Tests for a Single Stream of Data

We first introduce our test in the setting where we have a single stream of data, i.e., $W = 1$, with $m \in \mathbb{R}$, and null hypothesis $R = m$. This serves as the base case for our approach. We define the *capital process* $K_t(m)$:

Definition 3 (Single-Arm Capital Process)**.** *The capital process $K_t(m)$ and the corresponding log capital process*

[2] Let $1/m = \infty$ when $m = 0$, $-\frac{1}{1-m} = -\infty$ when $m = 1$.

$l_t(m)$ are defined as follows:

$$K_t(m) = \prod_{i=1}^{t} \left(1 + \frac{(\hat{\mu}_{i-1} - m)(X_i - m)}{c}\right), \quad (4)$$

$$l_t(m_a) = \sum_{i=1}^{t} \log\left(1 + \frac{(\hat{\mu}_{i-1} - m)(X_i - m)}{c}\right), \quad (5)$$

where $\hat{\mu}_i = m_a$ if $i = 1$ and $\hat{\mu}_{i-1} = \frac{1}{i-1}\sum_{j=1}^{i-1} X_j$ for all $i > 1$. Moreover, $c \geq 1/4$ is a constant.

Let $e_i(m) = 1 + \frac{(\hat{\mu}_{i-1}-m)(X_i-m)}{c}$. Then, (1) implies that $e_i(m)$ is an e-variable. The capital process $K_t(m)$ can be written as $K_t(m) = \prod_{i=1}^{t} e_i(m)$. By Definition 1, the capital process $K_t(m)$ forms an e-process for hypothesis m. We note that $K_t(m)$, (6), is a special case of the nonnegative martingale construction $M_t(m)$, (2), with $\mathcal{F}_{t-1}$-measurable betting sequence $\{\lambda_t(m) = \frac{(\hat{\mu}_{t-1}-m)}{c}\}_{t\in\mathbb{N}}$. This betting scheme is intuitive: we place our bet according to the sign of $\hat{\mu}_{t-1} - m$ (a plug-in estimate of the sign of $X_t - m_a$) and base the bet's magnitude on the absolute distance between $\hat{\mu}_{t-1}$ (our running estimate of X_t) and hypothesis m.

We note that our choice of $\lambda_t(m) = \frac{\hat{\mu}_{t-1}-m}{c}$ differs from all existing betting schemes. The closest betting scheme to our choice of $\{\lambda_t(m)\}_{t\in\mathbb{N}}$ is the AGRAPA method (Appendix B, Waudby-Smith and Ramdas 2023), which sets $\lambda_i(m) = \max(\frac{-l}{1-m}, \min\{\frac{l}{m}, \frac{\hat{\mu}_{i-1}-m}{\sigma_{i-1}^2+(\hat{\mu}_{i-1}-m)^2}\})$ for some $l \leq 1$ and a $\mathcal{F}_{i-1}$-measurable estimate of the variance, $\hat{\sigma}_{i-1}^2$. We discuss the differences between our λ scheme and AGRAPA in Prop. 1 that enable computational benefits.

Because $K_t(m)$ is an e-process, it defines the following sequential test and confidence sequence.

Definition 4. *Let $T_t(m, \alpha) = 1$ denote the rejection of null (composite) hypothesis $\mathcal{P}(m)$. The sequential capital process test is as follows:*

$$T_t(m, \alpha) = \mathbf{1}\left[\max_{1\leq i\leq t} K_i(m) \geq 1/\alpha\right]. \quad (6)$$

The dual confidence sequence is respectively defined as

$$C_t(\alpha) = \{m \in [0, 1] : T_t(m, \alpha) < 1/\alpha\}. \quad (7)$$

Theorem 1 below guarantees the correctness of this test. The probability of rejecting m when $m = \mu$ is less than or equal to α for all $t \in \mathbb{N}$ (proof in Appendix A.1):

Theorem 1. *For all $c \geq 1/4$, $T_t(m, \alpha)$ defines a sequential test with type-I error α for null hypothesis m, i.e., if $\mu = m$ (equivalently, $P_1 \in \mathcal{P}(m)$), $\mathbb{P}(\exists t \in \mathbb{N} : T_t(m) = 1) \leq \alpha$.*

This result is obtained by a direct application of Ville's inequality. While Theorem 1 guarantees the correctness of our test when the null hypothesis $m = \mu$, it does not provide guarantees to incorrect hypotheses. To address this, Theorem 2 (proof in Appendix A.3) states that for any $m \neq \mu$, $T_t(m, \alpha)$ is a test of power one, i.e., the stopping time of our test for an incorrect mean is finite.

Theorem 2. *For null hypothesis $m \neq \mu$, any $\alpha \in (0, 1)$ and any $c > 1/4$, (i) the capital process $K_t(m) \to \infty$ a.s. for any $m \neq \mu$, and (ii) the sequential test $T_t(m, \alpha) = \mathbf{1}[i \leq t : K_i(m) \geq 1/\alpha]$ is a test of power one, i.e.,*

$$\inf_{P\in\mathcal{P}(\mu)} P(\exists t < \infty : T_t(m, \alpha) = 1) = 1.$$

We next verify that $C_t(\alpha)$, the confidence sequence defined in Equation (7), is an interval by establishing the convexity of $K_t(m)$ below. Then a direct implication of Theorem 2 is that $C_t(\alpha)$ shrinks to volume 0 as $t \to \infty$ for all $\alpha \in (0, 1)$.

Proposition 1. *For all $t \in \mathbb{N}$, and for any sequence $(X_i)_{i=1}^t \in [0, 1]^t$, $K_t(m)$ is strictly convex for all $m \in \mathbb{R}$. Thus, there exists a unique minima m^*, which is the only $m \in \mathbb{R}$ that satisfies the following equation:*

$$\sum_{i=1}^{t} \frac{2m - (\hat{\mu}_{i-1} - X_i)}{c + (\hat{\mu}_{i-1} - m)(X_i - m)} = 0.$$

Furthermore, the unconstrained minima m^* is in $[0, 1]$.

The convexity of our confidence sequence $C_t(\alpha)$ follows directly from Proposition 1 (proof in Appendix A.5): at time $t = 0$, $C_0(\alpha) = [0, 1]$; for $t \in \mathbb{N}$, each test results in a convex set; repeated intersections $C_t(\alpha)$ is also convex. Note that this property does not generally hold in testing-by-betting schemes by Waudby-Smith and Ramdas (2023).

Remark 1. *For example, the AGRAPA betting scheme, despite its similar form to our choice of betting scheme, is not guaranteed to form convex confidence sets (Waudby-Smith and Ramdas, 2023, Appendix E.4). This requires one to test each $m \in [0, 1]$ to construct confidence sets (or each $m \in R$ to determine whether one can reject a region of interest R). Our choice of λ remedies this issue by (1) forming strictly convex confidence sets with respect to m, avoiding the need to test a grid of values over $[0, 1]$ when constructing confidence sets, and thus (2) enabling a simple characterization of the minima m in a convex region R, avoiding the need to test a fine grid of $m \in R$. These two benefits reduce the computational burden of our proposed testing method PEAK, a generalization of this simple betting scheme $K_t(m)$, discussed in Section 4.*

While Theorems 1 and 2 guarantee that μ will be identified in our test, they do not provide additional insights. To better understand the behavior of our test, in Section 3.1, we characterize the asymptotic e-power/growth rate (Definition 2) of our process, $K_t(m)$, which defines the expected linear rate of increase of $l_t(m)$ as $t \to \infty$. We discuss the

selection of parameter c in Lemma 1. In Lemma 2, we establish that the worst-case instance of e-power is strictly positive for $m \neq \mu$. Finally, given that our test is nonparametric, we compare the asymptotic growth rate of our test in the parametric setting to the best-achievable growth rate in Lemma 3, establishing that the e-power of our test is near-optimal in the parametric setting.

3.1. Asymptotic e-Power/Growth Rate

The asymptotic e-power of $K_t(m)$ is defined as follows:

Definition 5. *Let P be a distribution with mean μ, i.e., $P \in \mathcal{P}(\mu)$. We define the asymptotic growth rate of $K_t(m)$ as $G(c, m, P) = \lim_{t\to\infty} \mathbb{E}_P[\log(1+\frac{(\hat{\mu}_{t-1}-m)(X-m)}{c})] = \mathbb{E}_P[\log(1+\frac{(\mu-m)(X_t-m)}{c})]$.*[3]

In Lemma 1 (proof in Appendix A.4), we establish the relationship between variable c and asymptotic growth rate $G(c, m, P)$. This provides a natural choice of $c \in [1/4, \infty)$.

Lemma 1 (Asymptotic Growth Rate as a Function of c)**.** *For any $m \in [0, 1]$ and any alternative hypothesis P, the growth rate $G(c, m, P)$ monotonically decreases with c.*

While setting c to its minimal value maximizes the asymptotic growth, we recommend setting $c = 1/4+\gamma$, where γ is a small constant such as $\gamma = 0.01$ for the sake of stability.[4]

Next, in Lemma 2 (proof in Appendix A.2), we establish that the worst-case growth rate of $G(c, m, P)$ is strictly positive for hypotheses $m \neq \mu$, and increases as a function of the absolute difference between the null hypothesis m and the true mean μ.

Lemma 2 (Worst-Case Positive Asymptotic Growth Rate)**.** *For any fixed $m \in [0, 1], \mu \in [0, 1]$, the worst-case instance $P^* := \inf_{P\in\mathcal{P}(\mu)} G(c, m, P)$ is $P^* \equiv Bern(\mu)$.*

Furthermore, for any $m \in [0, 1]$, and any distribution P over $[0, 1]$ with mean μ, $c \geq 1/4$, the worst-case growth rate is strictly nonnegative, i.e., $\inf_{P\in\mathcal{P}(\mu)} G(c, m, P) \geq 0$. In particular, (1) $G(c, m, P)$ is equal to 0 only if $\mu = m$, and (2) $G(c, m, P)$ monotonically increases w.r.t. $|\mu - m|$.

Finally, our test is nonparametric. When compared with the likelihood ratios of Wald (1945), naturally, one would expect our worst-case asymptotic growth rate to deviate significantly from the best achievable parametric growth rate. However, Lemma 3 (proof in Appendix A.4) shows that our test achieves near-optimal asymptotic growth rate in the simple-vs-simple Bernoulli setting, an instance of our worst possible growth rate. Let $\mu \in (0, 1)$. We have:

[3]The interchange of limits and integrals in the last equality is valid by Leibnitz's rule.

[4]Consider the scenario with $c = 1/4$, where the estimated mean from previous observations $\hat{\mu}_{i-1} = 1$, $m = 0.5$, and the next observed data point $X_i = 0$. Then, $K_t(m) = 0$. For any time after t, $K_t(m) = 0$, meaning that the hypotheses $m = 0.5$ cannot be rejected for *any* time in the future.

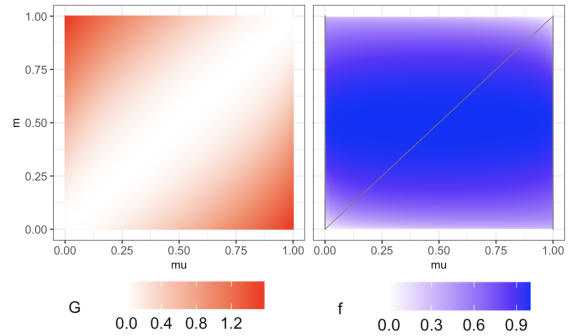

Figure 1. Growth rate visualization for different ground truth μ and hypothesis m combinations under Bernoulli distributions, $P \equiv$ Bern(μ), and $c = 0.26$. Left: Asymptotic growth rate of $K_t(m)$, $G(c, m, P)$, for null hypothesis m, with darker colors representing larger growth rates. Right: The *ratio* of $G(c, m, P)$ to the best achievable growth rate, $f(c, m, P)$, for null hypothesis m, with darker colors representing larger ratios.

Lemma 3 (Relative Growth Rate in Worst-Case Setting)**.** *When testing a Bernoulli distribution $X_i \sim Bern(\mu)$, the ratio of $G(c, m, P)$ to the best achievable growth rate in this simple-vs-simple sequential test is given by:*

$$f(c, m, \mu) = \frac{\log((1 + \frac{(\mu-m)(1-m)}{c})^{\mu}(1 - \frac{(\mu-m)m}{c})^{1-\mu})}{\log((\frac{\mu}{m})^{\mu}(\frac{1-\mu}{1-m})^{1-\mu})}.$$

To get a better sense of the ratio $f(c, m, \mu)$, we plot both $G(c, m, P)$ and $f(c, m, \mu)$ in Figure 1. We observe that as the distance between μ and m decreases, $f(m, \mu)$ increases rapidly in Figure 1 (right), meaning that $K_t(m)$ preserves the best-possible expected growth rate in difficult instances (i.e., the null hypothesis m is close to the alternative hypothesis μ). For instances where our growth rate demonstrates suboptimality (i.e., $m \approx 0$ or $m \approx 1$), we observe that $G(c, m, P)$ is still large (white/orange regions of Fig. 1 left). Lemma 3 and Figure 1 demonstrate that for incorrect hypotheses close to the true mean, the asymptotic growth rate of our process $K_t(m)$, in the worst case, is near-optimal with respect to the best-achievable growth rate.

4. Testing in the Multi-Stream Setting

In this section, we generalize the results of the single-stream/arm setting to W streams/arms, where $W \in \mathbb{N}$, and demonstrate how to test for a general null composite hypothesis R that is convex in the hypothesis space $m \in [0, 1]^W$. Let $N_t(a) = \sum_{i=1}^{t} \mathbf{1}[A_i = a]$ and $\hat{\mu}_t(a) = \sum_{i=1}^{t} \mathbf{1}[A_i = a]X_i/N_t(a)$. We first define our product capital process and the test for the case $R = m$ below:

Definition 6. *The capital process $K_t^a(m_a)$ for each stream/arm and the corresponding joint capital process*

$E_t(m)$ are defined as follows:

$$K_t^a(m_a) = \prod_{i=1}^{t} \Bigg(1 - \mathbf{1}[A_t = a] + \mathbf{1}[A_t = a] \cdot \left(1 + \frac{(\hat{\mu}_{i-1}(a) - m_a)(X_i - m_a)}{c}\right)\Bigg) \quad (8)$$

$$E_t(m) = \frac{1}{W} \sum_{a \in [W]} K_t^a(m_a). \quad (9)$$

The resulting α-level test for hypothesis m is as follows:

$$T_t(m, \alpha) = \mathbf{1}\left[\max_{1 \le i \le t} E_t(m) \ge 1/\alpha\right]. \quad (10)$$

This process generalizes our single-arm setting by averaging evidence across each arm for any hypothesis m, which gives rise to our procedure's name as *(P)eeking with (E)xpectation-based, (A)veraged (K)apital*.

Why should we average the evidence? The naive way to generalize our sequential tests is to use union bounds: we divide α by W and reject the hypothesis at time t if each component $K_t^a(m_a)$ exceeds the W/α threshold. In Lemma 4 (proof in Appendix A.8), we establish that the product test strictly dominates the union bound test.

Lemma 4. *Let $T'_t(m, \alpha)$ be the test under union bounding:*

$$T'_t(m, \alpha) = \prod_{a \in [W]} \mathbf{1}\left[\max_{1 \le i \le t} K_i^a(m_a) \ge W/\alpha\right].$$

For all $t \in \mathbb{N}$, any $m \in [0,1]^W$ rejected by union bounding test $T'_t(m, \alpha)$ is also rejected by the test $T_t(m, \alpha) = \mathbf{1}\left[\max_{1 \le i \le t} E_i(m) \ge 1/\alpha\right]$ for any realization of the data.

An immediate consequence of Lemma 4 is that for any realization of the data, the confidence sequences formed by inverting the test $T_t(m, \alpha)$ is a *strict subset* of the confidence sequences formed by union bounding across arms.

4.1. Theoretical Guarantees for Joint Capital Process

Let $\pi_t : H^{t-1} \to \Delta^W$ be a sampling scheme, where $H^{t-1} = \{A_1, X_1, ..., A_{t-1}, X_{t-1}\}$. This process $E_t(m)$ shares the same theoretical guarantees of the single arm setting for 1) type-I error control for all π_t, and 2) power under mild assumptions of π_t.

Theorem 3 (Correctness of Joint Capital Test)**.** *For $m = \mu$, $c \ge 1/4$, for any sampling scheme π_t, the test $T_t(m, \alpha)$ has type-I error probability $1 - \alpha$.*

Theorem 3 (proof in Appendix A.6) provides type-I error control for any possible sampling scheme π_t, and maintains the correctness of our test. To show that $T_t(m, \alpha)$ is a test of power one for any $m \neq \mu$, we require a relatively mild condition as stated in Theorem 4 (proof in Appendix A.7).

Theorem 4 (Test of Power 1)**.** *For any $m \neq \mu$, $c > 1/4$, let $\mathcal{I}(m) = \{a \in [W] : m_a \neq \mu_a\}$. Assume that the chosen sampling scheme π_t satisfies the following: there exists $a \in \mathcal{I}(m)$ s.t. $N_t(a) = \sum_{i=1}^{t} \mathbf{1}[A_t = a] \to \infty$ as $t \to \infty$. Then, $T_t(m, \alpha)$ is a test of power 1, i.e., $\inf_{P \in \mathcal{P}(\mu)} P(\exists t < \infty : T_t(m, \alpha) = 1) = 1$.*

Intuitively, the condition of Theorem 4 states that for an incorrect hypothesis $m \neq \mu$, we must pull at least one misspecified arm infinitely often (not necessarily the same arm across data realizations) as $t \to \infty$. Simple examples of sampling schemes that satisfy this property include $\pi_t(a) = 1/W$ or bandit sampling schemes such as ϵ-greedy or action elimination (Jamieson and Nowak, 2014).

4.2. Testing Composite Convex Hypotheses

The key benefit of our testing approach, as compared to using the betting schemes of Waudby-Smith and Ramdas (2023), is computational tractability for composite null tests that define convex regions in the hypothesis space, $[0,1]^W$. To reject a region R of the hypothesis space, we require that every hypothesis $m \in R$ in the region is rejected. In terms of our test, the criterion to reject a region R is as follows:

$$T_t(R, \alpha) = \mathbf{1}\left[\min_{1 \le i \le t, m \in R} E_i(m) \ge 1/\alpha\right]. \quad (11)$$

For such composite tests, it is crucial to find the minima of $E_t(m)$ within region R efficiently. We note that a straightforward extension of certain testing-by-betting schemes[5] by Waudby-Smith and Ramdas (2023) to the multi-arm setting would involve a grid search over the region $R \subset [0,1]^W$ to test hypothesis R. Because our composite test across joint means is simply the average of $K_t^a(m_a)$, it inherits the properties described in Proposition 1, such as strict convexity and a unique global minima $\tilde{m} \in [0,1]^W$. Consequently, the global minima of $E_t(m)$ is characterized by the following:

Proposition 2 (Global Minima of $E_t(m)$.)**.** *For time t, the vector $m^* = [m_1^*, ..., m_W^*] \in [0,1]^W$ that minimizes $E_t(m)$ is 1) unique, and 2) characterized component-wise by the following condition:*

$$\sum_{i=1}^{t} \sum_{a \in [W]} \mathbf{1}[A_t = a] \frac{2m_a^* - \hat{\mu}_{i-1}(a) - X_i}{c + (\hat{\mu}_{i-1}(a) - m_a^*)(X_i - m_a^*)} = 0.$$

To find the minima within region R, we project our global minima m^* as characterized in Proposition 2. Below, We provide two examples of this projection: threshold identification (THR) and best-arm identification (BAI), and provide visualizations in Figure B.2 in Appendix B.

[5] One such scheme is AGRAPA (Appendix B.4. in Waudby-Smith and Ramdas, 2023), which can form non-convex confidence regions.

Example 1: THR. For a threshold $\xi \in [0,1]$, each arm a in the threshold problem THR has two corresponding hypothesis regions: $R^0_{\text{THR}}(a) = \{m \in [0,1]^W : m_a < \xi\}$, and $R^1_{\text{THR}}(a) = [0,1]^W \setminus R^{\text{THR}}_0(a)$.

To test hypothesis $R^0_{\text{THR}}(a')$ at time t, we check if $E_t(m^0)$ exceeds the $1/\alpha$ threshold, where entries of m^0 are given by:

$$m^0_a = \begin{cases} m^*_a & \text{if } a \neq a' \\ \min(m^*_a, \xi) & \text{if } a = a' \end{cases}.$$

To test hypothesis $R^1_{\text{THR}}(a')$ at time t, we check if $E_t(m^1)$ exceeds the $1/\alpha$ threshold, where entries of m^1 are defined as:

$$m^1_a = \begin{cases} m^*_a & \text{if } a \neq a' \\ \max(m^*_a, \xi) & \text{if } a = a' \end{cases}.$$

The vectors m^0 and m^1 are the projections of the global minimizer m^*_a for regions $R^0_{\text{THR}}(a')$, $R^0_{\text{THR}}(a')$, respectively, and only involve thresholding the mean of arm a' based on ξ. Below, we consider a slightly more complicated example, BAI, which has constraints defined between pairs of arms.

Example 2: BAI. For best-arm identification, each arm a has a single hypothesis, as defined in Equation (3). Proposition 3 provides the conditions that define the minima of $E_t(m)$ for region $R_{\text{BAI}}(a)$ at time t.

Proposition 3 (Minima Condition for BAI Partition)**.** *Let $R_{BAI}(a')$ be the set defined in Equation (3). Let $\tilde{m} \in [0,1]^W$ be the global minimizer of $E_t(m)$. Then, $\min_{m \in R_{BAI}(a')} E_t(m)$ is obtained at solution m^*, given by:*

$$m^*_a = \begin{cases} \tilde{m}_a & \text{if } \tilde{m}_a < q \\ q & \text{if } \tilde{m}_a \geq q \end{cases},$$

where $q \in [0,1]$ is a constant that satisfies

$$\sum_{a \in W : \tilde{m}_a \geq q} \Bigg(\gamma(m^*, a) \sum_{i=1}^{t} \mathbf{1}[A_t = a] \frac{2q - X_i - \hat{\mu}_{i-1}(a)}{c + (\hat{\mu}_{i-1}(a) - q)(X_i - q)} \Bigg) = 0,$$

where $\gamma(m,a) = \prod_{i=1}^t (\mathbf{1}[A_i = a](1 + (X_i - m_a)(\hat{\mu}_{i-1}(a) - m_a)/c) + (1 - \mathbf{1}[A_i = a]))$.

Using this characterization, analysts can find the $m \in R_{\text{BAI}}(a')$ that achieves the minimum value within R_{BAI} by solving a single equation, as opposed searching across the entire region with grid values.

Beyond characterizing the constrained minima for specific hypothesis sets, finding the minimum value of $E_t(m)$ over any generic convex region $R \subset [0,1]^W$ defines a standard convex optimization problem. Thus, this allows analysts to (1) use the plethora of available convex optimization tools (Boyd and Vandenberghe, 2004) beyond grid-search style approaches, and (2) enables a direct characterization of the region-specific minima through KKT conditions, as done above for the BAI problem in Proposition 3.

5. Empirical Results

In this section, we provide empirical comparisons for our test that demonstrate (i) in the single-arm case, our test provides robust confidence sequences that verify the intuitions of Lemma 3, (ii) in the multi-stream setting with adaptive sampling, our test provides improved empirical performance to verify the composite hypothesis tests corresponding to threshold identification and best arm identification, and (iii) a stylized case-study using the mobile Health dataset HeartSteps (Liao et al., 2020). We use $\alpha = 0.05$ for our simulations and use $\alpha = 0.3$ in our case study.[6] For all experiments, we set $c = 0.26$[7].

5.1. Synthetic Experiments

Experiments in the Single-Stream Setting We compare the confidence sequences generated by our test to existing anytime-valid confidence intervals. We use the confidence sequence as a proxy for both the *power* and *growth rate* of our test. We compare our approach with four other nonparametric sequential tests that directly apply to bounded means: 1) Chernoff-based *sub-Bernoulli* (Sub-B) confidence sequences (Howard et al., 2021a),[8] 2) *empirical Bernstein* (Emp. Bern.) style confidence sequences (Waudby-Smith and Ramdas 2023, Thm. 2), 3) Hoeffding-based confidence sequences (PrPlH, Waudby-Smith and Ramdas 2023, Prop. 1), and 4) the hedged capital process confidence sequence (Hedged, Waudby-Smith and Ramdas 2023, Theorem 3, Eq. 26). For Hedged, we use a grid of 100 points for all simulations. We provide experiment details and additional experiments with nonstandard distributions in Appendix B.

Multi-Streams of Data, Adaptive Sampling We test multiple streams by 1) introducing adaptive sampling schemes that vary over time, and 2) requiring α-level guarantees across multiple parameters of interest. We set $W = 4$. In the Bernoulli setting (high variance case), the distribution of arm a is $P_a \equiv \text{Bern}(0.15 + 0.14a)$. In the Beta setting (low variance case), the distribution of arm a is $P_a \equiv \text{Beta}(1, \frac{0.85 - 0.14a}{0.15 + 0.14a})$. In both cases, $\mu_a = 0.14a + 0.15$. For both THR and BAI, Kano et al., 2019 and Jamieson and Nowak, 2014 (Eq. 2) propose a stopping criterion that uses sequential tests that union bound over each arm. We denote

[6] We note that our choice of $\alpha = 0.3$ is due to the lack of strong signal in this dataset (Liao et al., 2020).

[7] Code available at `https://github.com/brianc0413/PEAK`.

[8] as implemented using the package `confseq` (Howard et al., 2021b).

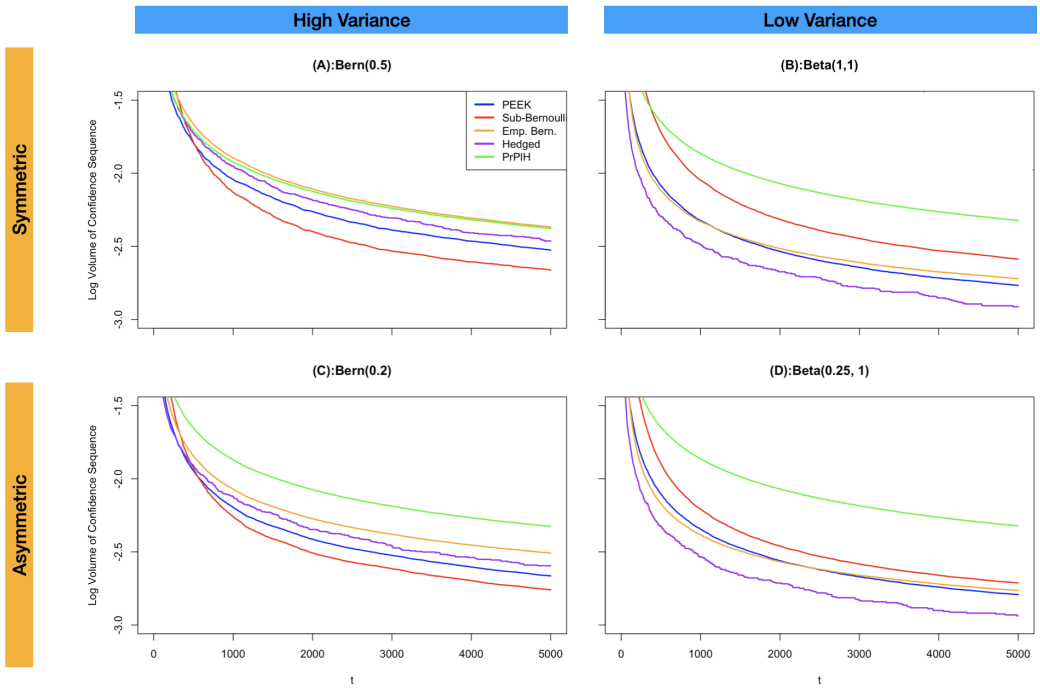

Figure 2. Log-widths of the confidence sequences across 5000 time steps. Curves represent the average width across 30 simulations.

Type	Stopping Rule	Stopping Time
Beta	PEAK	725.81 ± 218.49
	Base	4534.31 ± 482.94
	Sub-B	1618.32 ± 229.90
	Hedged	$\mathbf{479.16 \pm 205.97}$
Bern	PEAK	$\mathbf{1678.08 \pm 666.39}$
	Base	4795.51 ± 1101.09
	Sub-B	1686.81 ± 456.08
	Hedged	2241.18 ± 1092.49

Table 1. Stopping Times for THR. Stopping times represent the first time that we have classified all 4 arms as above or below $\xi = 0.5$. We report averaged stopping times and one standard error over 100 simulated sample paths. Across all simulations, each stopping time results in the correct conclusion (i.e., rejects the incorrect hypothesis).

Type	Stopping Rule	Stopping Time
Beta	PEAK	708.52 ± 266.34
	Base	4686.1 ± 565.24
	Sub-B	3175.3 ± 905.70
	Hedged	$\mathbf{500.12 \pm 214.08}$
Bern	PEAK	$\mathbf{1318.14 \pm 489.29}$
	Base	4631.66 ± 896.51
	Sub-B	4322.36 ± 3638.25
	Hedged	1734.64 ± 858.66

Table 2. Stopping Times for BAI. We report averaged stopping times and one standard error over 100 simulated sample paths. For all simulations, regardless of method, the correct best arm was identified.

their proposed stopping criterion as *Base*, and additionally test the two best-performing tests/confidence sequences (Sub-B, Hedged) in the single-arm case by using a similar union-bound approach. We refer to Appendix B for the exact adaptive sampling algorithm used for each bandit identification problem, and further experiment results, including simulations with nonstandard data distributions.

Discussion of Simulation Results In the single-arm case, Figure 1 plots the log-width of the confidence sequences across time. The log-width of our confidence interval suggests that PEAK provides a *robust testing procedure*, even in the single-arm case. Across any combination of variance settings and symmetry, our method performs the second-best across all tested approaches. In the high variance Bernoulli setting, Sub-B performs the best, as the sub-Bernoulli condition is tight with respect to the tested distribution. Compared to all other approaches, our test is closest in performance, corroborating the worst-case optimality results of Lemma 3. For the low-variance case, Hedged outperforms all other methods, but is closely followed by Emp. Bern. and PEAK. We additionally note that PEAK outperforms all benchmarks under mixture distributions as illustrated in Figure B.1.

In the multi-armed case, our test performs comparably to the state-of-the-art testing-by-betting method, Hedged, and outperforms the proposed stopping criterion and Sub-B. For both THR and BAI, the stopping times for PEAK are either

Method	Fidelity	Runtime (s)
PEAK	1×10^{-10}	5.21 ± 0.23
Hedged	1×10^{-2}	3.72 ± 0.19
	5×10^{-3}	6.982 ± 0.17
	2.5×10^{-3}	14.50 ± 0.65

Table 3. Runtimes Times for BAI Hypothesis Test, over a Fixed Horizon $T = 2000$, with tests at every other time point. We report averaged runtimes and one standard error over 100 simulated sample paths.

the lowest or second lowest on average empirically. For THR (Table B.5), the performance of our approach closely mirrors its performance in the single-arm case, where PEAK performs the second-best for both high and low variance settings, and Hedged and Sub-B perform the best for Beta (low variance) and Bernoulli (high variance) arms, respectively. For BAI, the testing-by-betting methods (Hedged and PEAK) drastically outperform Base and Sub-B as a stopping criterion. For the high variance, Bernoulli setting, PEAK outperforms Hedged, while Hedged outperforms PEAK in the low variance Beta setting.

Runtime To stress the key benefits of our approach, we test the best-arm hypothesis in the Bernoulli setting across a fixed $T = 2000$ time horizon (one test every two time steps), with no stopping, and report the average runtime for the fixed-horizon test over 100 simulations in Table 5.1. The runtimes for PEAK and Hedged are only comparable when the grid size is 100 (i.e., 10^{-2} fidelity) for Hedged. For larger grid sizes, such as 500 evenly spaced points, the runtime of Hedged is at least double that of PEAK, despite having far worse fidelity. The union-bound approach of using confidence intervals with the Hedged betting scheme scales linearly with the fidelity of the grid and becomes exponential if one were to average across arms, as we do with PEAK. This demonstrates that even with a relatively small number of streams, our approach drastically improves upon the computational requirements of existing betting schemes, such as Hedged.

5.2. Case Study with Mobile Health Data

We use the HeartSteps dataset (Liao et al., 2020) as a case study for our approach. The HeartSteps dataset was collected using a data-adaptive sampling scheme during a mobile health intervention with the HeartSteps app and contains user-level information on 37 users: (1) $A_{i,t}$ a binary intervention received at time t for user i, and (2) $X_{i,t}$, the number of user i's steps within a short interval after the intervention. In our case study, we investigate the value of τ_i, the first time in the trial in which we can determine whether $\mu_i(1) \geq \mu_i(0)$ (or vice versa) with 70% confidence. We make the following assumptions for the DGP: (1) each user i has an associated mean vector μ_i, where for all t, $\mathbb{E}[X_{i,t}|\mathcal{F}_{t-1}, A_{i,t} = a] = \mu_i(a)$, and (2) the number of steps that can be taken are bounded above for all users. These two assumptions are sufficient for the type-I error guarantees of PEAK. We apply the BAI termination condition for $W = 2$ for each i in the dataset using PEAK, Sub-B, and Hedged, the three most empirically promising testing schemes from Section 5.1, to obtain τ_i (if it occurs before the horizon of the data) for each user i.

In our experiments, we observe that Sub-B fails to conclude that the intervention is either helpful or harmful for any user i throughout the entire horizon of the trial. In contrast, both testing-by-betting methods are able to conclude that a treatment is helpful/harmful before the end of the trial, to varying degrees. Hedged concludes that the digital intervention provides no benefit and may be harmful (i.e., $\mu_i(1) \leq \mu_i(0)$) for User 22 using just half of the samples collected over the trajectory ($\tau_{22} = 4932$ over a horizon of 9072). For all other patients, Hedged fails to conclude that the intervention is either helpful or harmful. PEAK outperforms both approaches: it provides stopping times τ_i for 8 different patients before the end of the recorded data, identifying 8 different patients who all benefit from the treatment (i.e., $\mu_i(1) \geq \mu_i(0)$). If the analyst were to stop trials for the 8 users at their respective stopping times, they would have saved 18,563 total samples, reducing the total number of samples by roughly 10%. This simple case study demonstrates PEAK's potential to reduce experimental costs and provide resources for additional patient enrollment.

6. Conclusions and Future Directions

In our work, we propose a novel nonparametric sequential test for composite hypotheses over means of multiple data streams. In the single-arm case, our approach provides a robust option among existing sequential tests and anytime valid confidence sequences. In the multi-arm case, our test pools evidence across arms to avoid union bounds, and empirically outperforms other sequential tests both (i) as a stopping criterion for common pure-exploration bandit problems, such as threshold or best-arm identification, and (ii) in terms of runtime compared with leading approaches for sequential testing. Importantly, our approach is compatible with various existing convex optimization software, avoiding the need for grid-search style methods while maintaining the performance of other testing-by-betting methods.

Despite improved computational tractability, PEAK requires the use of solvers to test hypotheses; existing parametric e-processes have minimal computational overhead relative to any testing-by-betting approach, including PEAK. Our theoretical/empirical results suggest that PEAK is most appropriate for problems where (1) strong parametric assumptions are unrealistic, and (2) one wishes to preserve power against both high-variance and low-variance data streams.

Acknowledgements

Brian Cho was supported by the NDSEG fellowship.

Impact Statement

This paper presents work whose goal is to advance the field of Machine Learning. There are many potential societal consequences of our work, none which we feel must be specifically highlighted here.

References

Audibert, J.-Y., Bubeck, S., and Munos, R. (2010). Best arm identification in multi-armed bandits. In *Annual Conference Computational Learning Theory*.

Bibaut, A., Kallus, N., and Lindon, M. (2022). Near-optimal non-parametric sequential tests and confidence sequences with possibly dependent observations. *arXiv preprint arXiv:2212.14411*.

Bouneffouf, D., Bouzeghoub, A., and Gançarski, A. L. (2012). A contextual-bandit algorithm for mobile context-aware recommender system. In *Neural Information Processing: 19th International Conference, ICONIP 2012, Doha, Qatar, November 12-15, 2012, Proceedings, Part III 19*, pages 324–331. Springer.

Boyd, S. and Vandenberghe, L. (2004). *Convex optimization*. Cambridge university press.

Cai, W., Grossman, J., Lin, Z. J., Sheng, H., Wei, J. T.-Z., Williams, J. J., and Goel, S. (2021). Bandit algorithms to personalize educational chatbots. *Machine Learning*, 110(9):2389–2418.

Cappé, O., Garivier, A., Maillard, O.-A., Munos, R., and Stoltz, G. (2013). Kullback–leibler upper confidence bounds for optimal sequential allocation. *The Annals of Statistics*, 41(3).

Clement, B., Roy, D., Oudeyer, P.-Y., and Lopes, M. (2015). Multi-armed bandits for intelligent tutoring systems. *Journal of Educational Data Mining*.

de la Peña, V. H., Klass, M. J., and Leung Lai, T. (2004). Self-normalized processes: exponential inequalities, moment bounds and iterated logarithm laws. *The Annals of Probability*, 32(3).

Figueroa, C. A., Aguilera, A., Chakraborty, B., Modiri, A., Aggarwal, J., Deliu, N., Sarkar, U., Jay Williams, J., and Lyles, C. R. (2021). Adaptive learning algorithms to optimize mobile applications for behavioral health: guidelines for design decisions. *Journal of the American Medical Informatics Association*, 28(6):1225–1234.

Forman, E. M., Kerrigan, S. G., Butryn, M. L., Juarascio, A. S., Manasse, S. M., Ontañón, S., Dallal, D. H., Crochiere, R. J., and Moskow, D. (2019). Can the artificial intelligence technique of reinforcement learning use continuously-monitored digital data to optimize treatment for weight loss? *Journal of Behavioral Medicine*, 42:276–290.

Grünwald, P., de Heide, R., and Koolen, W. (2023). Safe testing.

Howard, S. R., Ramdas, A., McAuliffe, J., and Sekhon, J. (2021a). Time-uniform, nonparametric, nonasymptotic confidence sequences. *The Annals of Statistics*, 49(2).

Howard, S. R., Waudby-Smith, I., and Ramdas, A. (2021b). ConfSeq: software for confidence sequences and uniform boundaries.

Jamieson, K. and Nowak, R. (2014). Best-arm identification algorithms for multi-armed bandits in the fixed confidence setting. In *2014 48th Annual Conference on Information Sciences and Systems (CISS)*, pages 1–6.

Kano, H., Honda, J., Sakamaki, K., Matsuura, K., Nakamura, A., and Sugiyama, M. (2019). Good arm identification via bandit feedback. *Machine Learning*, 108:721–745.

Kaufmann, E. and Kalyanakrishnan, S. (2013). Information complexity in bandit subset selection. *Journal of Machine Learning Research*, 30:228–251.

Kaufmann, E. and Koolen, W. M. (2021). Mixture martingales revisited with applications to sequential tests and confidence intervals. *Journal of Machine Learning Research*, 22(246):1–44.

Liang, B. and Bojinov, I. (2023). An experimental design for anytime-valid causal inference on multi-armed bandits. *arXiv preprint arXiv:2311.05794*.

Liao, P., Greenewald, K., Klasnja, P., and Murphy, S. (2020). Personalized heartsteps: A reinforcement learning algorithm for optimizing physical activity. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 4(1):1–22.

Piette, J. D., Newman, S., Krein, S. L., Marinec, N., Chen, J., Williams, D. A., Edmond, S. N., Driscoll, M., LaChappelle, K. M., Maly, M., et al. (2022). Artificial intelligence (ai) to improve chronic pain care: Evidence of ai learning. *Intelligence-Based Medicine*, 6:100064.

Pérez-Ortiz, M. F., Lardy, T., de Heide, R., and Grünwald, P. (2023). E-statistics, group invariance and anytime valid testing.

Ramdas, A., Grünwald, P., Vovk, V., and Shafer, G. (2023). Game-theoretic statistics and safe anytime-valid inference. *Statistical Science*, 38(4):576–601.

Ramdas, A., Ruf, J., Larsson, M., and Koolen, W. (2022). Admissible anytime-valid sequential inference must rely on nonnegative martingales.

Robbins, H. and Siegmund, D. (1970). Boundary crossing probabilities for the wiener process and sample sums. *The Annals of Mathematical Statistics*, pages 1410–1429.

Sharpe, E. E., Karasouli, E., and Meyer, C. (2017). Examining factors of engagement with digital interventions for weight management: rapid review. *JMIR research protocols*, 6(10):e6059.

Shekhar, S. and Ramdas, A. (2023). On the near-optimality of betting confidence sets for bounded means.

Shin, J., Ramdas, A., and Rinaldo, A. (2021). On the bias, risk, and consistency of sample means in multi-armed bandits. *SIAM Journal on Mathematics of Data Science*, 3(4):1278–1300.

ter Schure, J., Perez-Ortiz, M. F., Ly, A., and Grunwald, P. (2023). The anytime-valid logrank test: Error control under continuous monitoring with unlimited horizon.

Trella, A. L., Zhang, K. W., Nahum-Shani, I., Shetty, V., Doshi-Velez, F., and Murphy, S. A. (2022). Designing reinforcement learning algorithms for digital interventions: pre-implementation guidelines. *Algorithms*, 15(8):255.

Trella, A. L., Zhang, K. W., Nahum-Shani, I., Shetty, V., Doshi-Velez, F., and Murphy, S. A. (2023). Reward design for an online reinforcement learning algorithm supporting oral self-care. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 15724–15730.

Ville, J. (1939). *Étude critique de la notion de collectif*.

Vovk, V. and Wang, R. (2021). E-values: Calibration, combination and applications. *The Annals of Statistics*, 49(3).

Vovk, V. and Wang, R. (2024). Efficiency of nonparametric e-tests.

Wald, A. (1945). Sequential tests of statistical hypotheses. *Annals of Mathematical Statistics*, 16:256–298.

Wald, A. and Wolfowitz, J. (1948). Optimum Character of the Sequential Probability Ratio Test. *The Annals of Mathematical Statistics*, 19(3):326 – 339.

Waudby-Smith, I., Arbour, D., Sinha, R., Kennedy, E. H., and Ramdas, A. (2021). Time-uniform central limit theory and asymptotic confidence sequences. *arXiv e-prints*, pages arXiv–2103.

Waudby-Smith, I. and Ramdas, A. (2023). Estimating means of bounded random variables by betting. *Journal of the Royal Statistical Society*.

Woong Ham, D., Lindon, M., Tingley, M., and Bojinov, I. (2023). Design-based confidence sequences: A general approach to risk mitigation in online experimentation. *Harvard Business School Technology & Operations Mgt. Unit Working Paper*, (23-070).

Yardley, L., Spring, B. J., Riper, H., Morrison, L. G., Crane, D. H., Curtis, K., Merchant, G. C., Naughton, F., and Blandford, A. (2016). Understanding and promoting effective engagement with digital behavior change interventions. *American Journal of Preventive Medicine*, 51(5):833–842.

Yom-Tov, E., Feraru, G., Kozdoba, M., Mannor, S., Tennenholtz, M., and Hochberg, I. (2017). Encouraging physical activity in patients with diabetes: Intervention using a reinforcement learning system. *Journal of Medical Internet Research*, 19(10):e338.

Appendix

For all α-level type I error correctness proofs, we will use the folllowing inequality:

Lemma 5 (Ville's Maximal Inequality)**.** *For any non-negative martinagale L_t and any $x > 1$, define a potentially infinite stopping time $N \coloneqq \inf\{t \geq 1 : L_t \geq x\}$. Then,*

$$\mathbb{P}(\exists t : L_t \geq x) \leq \mathbb{E}[L_0]/x.$$

Proof of Ville's Maximal Inequality. Define the expected overshoot when L_t surpasses x as $o = \mathbb{E}[L_N/x|N < \infty] \geq 1$. Using the optional stopping theorem for supermartingales,

$$\begin{aligned}
\mathbb{E}[L_0] &\geq \mathbb{E}[L_N] \\
&= \mathbb{E}[L_N|N < \infty]\mathbb{P}(N < \infty) + \mathbb{E}[L_\infty|N = \infty]\mathbb{P}(N = \infty) \\
&\geq \mathbb{E}[L_N|N < \infty]\mathbb{P}(N < \infty) \\
&= ox\mathbb{P}(N \leq \infty)
\end{aligned}$$

Thus, $\mathbb{P}(N \leq \infty) = \mathbb{P}(\exists t : L_t \geq x) \leq \frac{\mathbb{E}[L_0]}{ox} \leq \mathbb{E}[L_0]/x$. Setting $x = 1/\alpha$, we obtain the desired results for our test martingale. □

A. Proofs for Section 3

A.1. Proof of Theorem 1

Proof of Theorem 1. Note that we are in the setting where $\mu = m$, i.e. our null hypothesis is correct. We first show that $K_t(m)$ is nonnegative for all $t \in \mathbb{N}$, for $c \geq 1/4$. To prove that $K_t(m) = \prod_{i=1}^{t}(1 + \frac{(\hat{\mu}_{i-1} - m)}{c}(X_i - m))$ is nonnegative, it suffices to show that $\frac{(\hat{\mu}_{i-1} - m)}{c} \in (-1/(1-m), 1/m)$ for all $i \in \mathbb{N}$. For the upper bound,

$$\begin{aligned}
1/m - \frac{(\hat{\mu}_{i-1} - m)}{c} &= \frac{1}{m} - \frac{\hat{\mu}_{i-1} - m}{c} = \frac{1/4 + \hat{\sigma}_{i-1}^2 - m\hat{\mu}_{i-1} + m_a^2}{c} \\
&= \left((m - \frac{1}{2}\hat{\mu}_{i-1})^2 + c - \frac{1}{4}(\hat{\mu}_{i-1}^2)\right)/(cm) \\
&\geq 0.
\end{aligned}$$

For the lower bound,

$$\begin{aligned}
\frac{(\hat{\mu}_{i-1} - m)}{c} - (-\frac{1}{1-m}) &= \frac{(\hat{\mu}_{i-1} - m)(1-m) + c}{c(1-m_a)} \\
&= \left(\hat{\mu}_{i-1}(1-m) + (m - 1/2)^2 + c - 1/4\right)/(c(1-m)) \\
&\geq 0.
\end{aligned}$$

Not only is $K_t(m)$ nonnegative, but it is also a martingale:

$$\mathbb{E}[K_t(m)|\mathcal{F}_{t-1}] = K_{t-1}(m) * (1 + \frac{\hat{\mu}_{i-1} - m}{c}\mathbb{E}[(X_i - m)|\mathcal{F}_{t-1}]) = K_{t-1}(m).$$

Thus, by Ville's Maximal Inequality, Lemma (5), $\mathbb{P}(\exists t : K_t(m) \geq 1/\alpha) = \alpha$. □

A.2. Proof of Lemma 2.

Proof of Lemma 2. This proof proceeds in two parts:

1. First, we prove that $\mathbb{E}[\log(1+\frac{(\mu-m)(X-m)}{c})] \geq 0$, with equality iff $m=\mu$ for $X \sim \text{Bern}(\mu)$.
2. Second, we show that for any distribution with support $[0,1]$ and mean μ, the expectation $\mathbb{E}[\log(1+\frac{(\mu-m)(X_i-m)}{c})]$ is minimized at the Bernoulli distribution $\text{Bern}(\mu)$.

Step 1: Under a Bernoulli distribution $X \sim \text{Bern}(\mu)$, the expectation term becomes the following:

$$\mathbb{E}[\log(1+\frac{(\mu-m)(X-m)}{c})] = \mu_a \log(1+\frac{(\mu-m_a)(1-m)}{c}) + (1-\mu_a)\log(1+\frac{(\mu-m)(-m)}{c}).$$

Taking the first-order derivative with respect to m,

$$\begin{aligned}\frac{\delta}{\delta m}\mathbb{E}[\log(1+\frac{(\mu-m)(X-m)}{c})] &= \mu_a \frac{2m-1-\mu}{c+(\mu-m)(1-m)} + (1-\mu)\frac{2m-\mu}{c+(\mu-m)(-m)} \\ &= \frac{\mu(2m-1-\mu)(c+(\mu-m)(-m)) + (1-\mu)(2m-\mu)(c+(\mu-m)(1-m))}{(c+(\mu-m)(-m))(c+(\mu-m)(1-m))}\end{aligned}$$

Note that the denominator is strictly positive, and thus we only need to solve the numerator expression. The numerator expression reduces to the following expression, whose sign only depends on $m-\mu$:

$$\begin{aligned}&\mu(2m-1-\mu)(c+(\mu-m)(-m)) + (1-\mu)(2m-\mu)(c+(\mu-m)(1-m)) \\ &= 2\,(m-\mu)\,(c+(\mu-m)(-m)) + (1-\mu)(2m-\mu)(\mu-m) \\ &= 2(m-\mu)\left(c+(\mu-m)(-m) - \frac{1}{2}(1-\mu)(2m-\mu)\right) \\ &= 2(m-\mu)\underbrace{(c+m^2-m+1/2(\mu-\mu^2))}_{(i):>0}\end{aligned}$$

where statement (i) holds due to $m^2-m \geq -1/4$, $m,\mu \in [0,1]$ and $c \geq 1/4$. This derivative expression is solved when $m=\mu$, which results in $\mathbb{E}[\log(1+\frac{(\mu-m)(X-m)}{c})]=0$. The derivative is strictly negative when $m<\mu$, i.e. our function strictly decreases in the region $m \in [0,\mu)$. Symmetrically, the derivative is strictly positive when $m>\mu$, i.e. our function strictly increases in the region $m_a \in (\mu,1]$ - thus the asymptotic growth rate $G(c,m,P)$ monotonically increases with $|\mu-m|$. This also indicates that in the interior, $m=\mu$ is the minimum. To check endpoints of our feasible region, we only need to consider the cases $m_a \in \{0,1\}$, where both cases give us $\mathbb{E}[\log(1+\frac{(\mu-m)(X-m)}{1/4+\sigma^2})] > 0$. Thus, the global minimum is achieved at $m=\mu$, with $\mathbb{E}[\log(1+\frac{(\mu-m)(X-m)}{c})]=0$, and all other values of m_a result in $\mathbb{E}[\log(1+\frac{(\mu-m)(X-m)}{c})] > 0$.

Step 2: Now, we show that for any distribution $P \in \mathcal{P}(\mu)$, the expectation is lower bounded by the Bernoulli distribution $\text{Bern}(\mu)$. Let $X \sim P$, and let U be an independent $\text{Unif}[0,1]$ random variable. Note that the random variable $R=\mathbf{1}[U \leq X]$ has the following properties:

$$\mathbb{E}_U[\mathbf{1}[U \leq X]|X] = X, \quad \mathbb{E}[\mathbf{1}[U \leq X]] = \mu.$$

Thus, R is a Bernoulli random variable, with mean μ. Our function $f(X)=\log(1+\frac{(\mu-m)(X_i-m)}{c})$ is a concave function for all $\mu, m \in [0,1]$, and by using the conditional version of Jensen's inequality:

$$\mathbb{E}[f(R)|X] \leq f(\mathbb{E}[R|X]) = f(X).$$

By taking the expectation over X, we obtain the desired inequality:

$$\mathbb{E}[f(R)] \leq \mathbb{E}[f(X)] \implies \mathbb{E}_{X\sim\text{Bern}(\mu)}[\log(1+\frac{(\mu-m)(X_i-m)}{c})] \leq \mathbb{E}_P[\log(1+\frac{(\mu-m)(X_i-m)}{c})].$$

□

A.3. Proof of Theorem 2.

Proof of Theorem 2. We first prove the statement that $K_t(m) \to \infty$ almost surely as $t \to \infty$. First, note that we can include indicator functions to separate the positive and negative terms within the following expectation:

$$\mathbb{E}[\log((1 + \frac{\mu - m)(X_i - m)}{c}] =$$
$$\mathbb{E}\left[\mathbf{1}[X_i \geq m] \log((1 + \frac{\mu - m)(X_i - m)}{c}\right] + \mathbb{E}\left[\mathbf{1}[X_i < m_a] \log((1 + \frac{\mu - m)(X_i - m)}{c})\right] > 0.$$

Consider the following function, which equals the summation above at $\epsilon = 0$:

$$f(\epsilon, m) = \mathbb{E}\left[\mathbf{1}[X_i \geq m] \log(1 + \frac{(\mu - \epsilon - m)(X_i - m_a)}{c})\right] + \mathbb{E}\left[\mathbf{1}[X_i \geq m_a] \log(1 + \frac{(\mu + \epsilon - m)(X_i - m_a)}{c})\right]$$

Note that $f(\epsilon, m) = \mathbb{E}[\log((1 + \frac{\mu - m_)(X_i - m)}{c}] > 0$ at $\epsilon = 0$ (proven above). Because $f(\epsilon, m)$ is continuous with respect to ϵ, we know there exists an $\epsilon(m)$ such that $f(\epsilon(m), m) = \mathbb{E}[\log((1 + \frac{(\mu - m)(X_i - m)}{c}]/2 > 0$. By the Kolmogorov Strong Law of Large Numbers, we know $t(m, \omega) \in \mathbb{N}$, $t(m, \omega) < \infty$ such that $\hat{\mu}_t \in [\mu - \epsilon(m), \mu + \epsilon(m)]$ for all $t \geq t(m, \omega)$ almost surely, where $\omega \in \Omega$ s.t. $\mathbb{P}(\Omega) = 1$. We denote $t(m, \omega)$ as $t(m)$ to keep notational clutter minimal. The log capital process $l_t(m)$, for $t > t(m)$, takes the following form:

$$l_t(m) = \underbrace{\sum_{i=1}^{t(m)} \log(1 + \frac{(\hat{\mu}_{i-1} - m_a)(X_i - m_a)}{c})}_{(a)} + \underbrace{\sum_{i=t(m)}^{t} \log(1 + \frac{(\hat{\mu}_{i-1} - m_a)(X_i - m_a)}{c})}_{(b)}. \tag{12}$$

Summation (a) is guaranteed to be finite almost surely by $t(m) < \infty$, and we will now show that summation (b) diverges to ∞ for all $m_a \neq \mu$. We can rewrite summation (b) as the following:

$$\sum_{i=t(m)}^{t} \log(1 + \frac{(\hat{\mu}_{i-1} - m_a)(X_i - m_a)}{c}) =$$
$$\sum_{i=t(m)}^{\infty} \mathbf{1}[X_i \geq m_a] \log(1 + \frac{(\hat{\mu}_{i-1} - m_a)(X_i - m_a)}{c}) + \sum_{i=t(m)}^{t} \mathbf{1}[X_i < m_a] \log(1 + \frac{(\hat{\mu}_{i-1} - m_a)(X_i - m_a)}{c})$$

Recall that for all $i \geq t(m)$, $\hat{\mu}_i \in [\mu - \epsilon(m), \mu + \epsilon(m)]$. Using this, we can compare our summation (b) to the strictly smaller process for any realization $\{X_i\}_{i=1}^{\infty}$:

$$(b) \geq \sum_{i=t(m)}^{t} \mathbf{1}[X_i \geq m_a] \log(1 + \frac{(\mu - \epsilon(m_a) - m_a)(X_i - m_a)}{c})$$
$$+ \sum_{i=t(m)}^{t} \mathbf{1}[X_i < m_a] \log(1 + \frac{(\mu + \epsilon(m_a) - m_a)(X_i - m_a)}{c}).$$

By another application of SLLN, note that the RHS converges a.s. to a positive constant as $t \to \infty$:

$$\frac{\sum_{i=t(m)}^{t} \mathbf{1}[X_i \geq m_a] \log(1 + \frac{(\mu - \epsilon(m_a) - m_a)(X_i - m_a)}{c})}{t - t(m)} + \frac{\sum_{i=t(m)}^{t} \mathbf{1}[X_i \geq m_a] \log(1 + \frac{(\mu - \epsilon(m_a) - m_a)(X_i - m_a)}{c})}{t - t(m)}$$
$$\to f(\epsilon(m), m) > 0$$

which implies that the non-normalized summation diverges to ∞:

$$\mathbf{1}[X_i \geq m_a] \log(1 + \frac{(\mu - \epsilon(m_a) - m_a)(X_i - m_a)}{c}) + \sum_{i=t(m)}^{t} \mathbf{1}[X_i < m_a] \log(1 + \frac{(\mu + \epsilon(m_a) - m_a)(X_i - m_a)}{c}) \to \infty.$$

Because (b) is at least as large as the series which diverges to ∞ a.s., summation $(b) \to \infty$ a.s. as well, resulting in the following limit for $K_t(m)$:

$$l_t(m) = \underbrace{(a)}_{\text{finite.}} + \underbrace{(b)}_{\to\infty} \to \infty, \qquad K_t(m) = \exp(l_t(m)) \to \infty.$$

We now prove that the stopping time (i.e. time until rejection) of the test $T_t(m, \alpha)$ is finite for any fixed $m \neq \mu$ and any $\alpha \in (0, 1)$. Using $\epsilon(m), t(m)$ as above, if $T_t(m, \alpha) = 1$ for some $t \leq t(m)$, then we are guaranteed that the stopping time $\tau = \{\min_{t\in\mathbb{N}} t : T_t(m, \alpha) = 1\}$ is finite. If not, then we rewrite the test as follows using our log-capital expression $l_t(m)$ in Equation (12):

$$T_t(m, \alpha) = \mathbf{1}[i \leq t : \sum_{i=t(m)}^{t} \log(1 + \frac{(\hat{\mu}_{i-1} - m_a)(X_i - m_a)}{c}) \geq \ln(1/\alpha) - \sum_{i=1}^{t(m)} \log(1 + \frac{(\hat{\mu}_{i-1} - m_a)(X_i - m_a)}{c})].$$

Now, by replacing the term $\sum_{i=t(m)}^{t} \log(1 + \frac{(\hat{\mu}_{i-1}-m_a)(X_i-m_a)}{c})$ with a strictly smaller term $\sum_{i=t(m)}^{t} \mathbf{1}[X_i \geq m_a] \log(1 + \frac{(\mu-\epsilon(m_a)-m_a)(X_i-m_a)}{c}) + \sum_{i=t(m)}^{t} \mathbf{1}[X_i < m_a] \log(1 + \frac{(\mu+\epsilon(m_a)-m_a)(X_i-m_a)}{c})$, we construct a new stopping rule, $T'_t(m, a)$, which substitutes the LHS of the inequality for a strictly smaller term (shown above) as well as normalizing by $t - t(m)$.

$$\begin{aligned} T'_t(m, \alpha) = \mathbf{1}[i \leq t : \; & \frac{1}{t - t(m)} \sum_{i=t(m)}^{t} \mathbf{1}[X_i \geq m_a] \log(1 + \frac{(\mu - \epsilon(m_a) - m_a)(X_i - m_a)}{c}) + \\ & \frac{1}{t - t(m)} \sum_{i=t(m)}^{t} \mathbf{1}[X_i < m_a] \log(1 + \frac{(\mu + \epsilon(m_a) - m_a)(X_i - m_a)}{c}) \\ & \geq \frac{1}{t - t(m)} \left(\ln(1/\alpha) - \sum_{i=1}^{t(m)} \log(1 + \frac{(\hat{\mu}_{i-1} - m_a)(X_i - m_a)}{c}) \right)]. \end{aligned}$$

By the SLLN and by the fact that the LHS converges to $f(\epsilon(m), m)/2 > 0$, there exists a finite time $t'(m, \omega) \in \mathbb{N}$ almost surely for each sample path $\omega \in \Omega$ such that for all $t \geq t'(m, \omega) = t'(m)$, the following holds:

$$\begin{aligned} & \frac{\sum_{i=t(m)}^{t} \mathbf{1}[X_i \geq m_a] \log(1 + \frac{(\mu-\epsilon(m_a)-m_a)(X_i-m_a)}{c})}{t - t(m)} + \frac{\sum_{i=t(m)}^{t} \mathbf{1}[X_i \geq m_a] \log(1 + \frac{(\mu-\epsilon(m_a)-m_a)(X_i-m_a)}{c})}{t - t(m)} \\ & \geq f(\epsilon(m), m)/4 > 0. \end{aligned}$$

Let $t''(m)$ be the (random) time, $t''(m) \geq t'(m)$ such that the following holds:

$$\frac{1}{t''(m) - t(m)} \left(\ln(1/\alpha) - \sum_{i=1}^{t(m)} \log(1 + \frac{(\hat{\mu}_{i-1} - m_a)(X_i - m_a)}{c}) \right)] \leq f(\epsilon(m), m)/4.$$

Note that $t''(m)$ must also be finite, as $(\ln(1/\alpha) - \sum_{i=1}^{t(m)} \log(1 + \frac{(\hat{\mu}_{i-1}-m_a)(X_i-m_a)}{c})$ is finite due to the finiteness of $t(m)$. Thus, $\mathbb{P}(t < \infty : T'_t(m, \alpha) = 1) = 1$. Because the (unnormalized) LHS of the inequality is strictly smaller for $T'_t(m, \alpha)$ than for $T_t(m, \alpha)$, if $T'_t(m, \alpha) = 1$, then $T_t(m, \alpha) = 1$ as well. Thus,

$$\mathbb{P}(t < \infty : T_t(m, \alpha) = 1) \geq \mathbb{P}(t < \infty : T'_t(m, \alpha) = 1) = 1 \implies \mathbb{P}(t < \infty : T_t(m, \alpha) = 1) = 1.$$

□

A.4. Proof of Lemmas 1 and and 3

Proof of Lemma 1 . Let $c, c' \in [1/4, \infty)$, with $c' < c$. Let P be any distribution in $\mathcal{P}(m)$. Taking the difference of the two asymptotic growth rates, we obtain:

$$\begin{aligned}
&\mathbb{E}_P[\log(1 + \frac{(\mu - m)(X - m)}{c'})] - \mathbb{E}_P[\log(1 + \frac{(\mu - m)(X - m)}{c})] \\
&= \mathbb{E}_P[\log(c' + (\mu - m)(X - m)) - \log(c')] - \mathbb{E}_P[\log(c + (\mu - m)(X - m)) - \log(c)] \\
&= \mathbb{E}_P[\log(\frac{c' + (\mu - m)(X - m)}{c + (\mu - m)(X - m)})] + \log(\frac{c}{c'}) \\
&= \mathbb{E}_P[\log(c' + (\mu - m)(X - m))] - \mathbb{E}_P[\log(c' + (\mu - m)(X - m))] + \log(c) - \log(c').
\end{aligned}$$

Then, we use the identity: $\log(x + y) = \log(x) + \log(y) + \log(\frac{1}{x} + \frac{1}{y})$ to obtain the following expressions for each expectation term:

$$\begin{aligned}
\mathbb{E}_P[\log(c' + (\mu - m)(X - m))] &= \log(c') + \mathbb{E}_P[\log(\mu - m)(X - m) + \log(\frac{1}{c'} + \frac{1}{(\mu - m)(X - m)})] \\
\mathbb{E}_P[\log(c + (\mu - m)(X - m))] &= \log(c) + \mathbb{E}_P[\log(\mu - m)(X - m) + \log(\frac{1}{c} + \frac{1}{(\mu - m)(X - m)})]
\end{aligned}$$

Returning to the main expression,

$$\begin{aligned}
&\mathbb{E}_P[\log(1 + \frac{(\mu - m)(X - m)}{c'})] - \mathbb{E}_P[\log(1 + \frac{(\mu - m)(X - m)}{c})] \\
&= \mathbb{E}_P[\log(1/c' + (\mu - m)^{-1}(X - m)^{-1}) - \log(1/c + (\mu - m)^{-1}(X - m)^{-1})] \\
&> 0.
\end{aligned}$$

where the last inequality is because the LHS of the expression within the expectation is always strictly greater than the right hand side of the expression ($1/c' > 1/c$) for all X in the support of P. □

Proof of Lemma 3. First, note that the maximal rate of growth for any martingale-based test for the Bernoulli setting is given by the KL-divergence (Pérez-Ortiz et al., 2023):

$$G^{\text{opt.}}(m, \mu) = \mathbb{E}_\mu[\frac{\mu}{m}] = \mu \log(\frac{\mu}{m}) + (1 - \mu) \log(\frac{1 - \mu}{1 - m}).$$

In the Bernoulli case, the expectation of $G(c, m, P)$ is given by:

$$G(c, m, P) = \mu \log(1 + \frac{(\mu - m)(1 - m)}{c}) + (1 - \mu) \log(1 + \frac{(\mu - m)(-m)}{c}).$$

Taking the ratio, we obtain the desired result:

$$\begin{aligned}
\frac{G(c, m, P)}{G^{\text{opt.}}(m, \mu)} &= \frac{\log\left((1 + \frac{(\mu - m)(1 - m)}{c})^\mu (1 + \frac{(\mu - m)(-m)}{c})^{1-\mu}\right)}{\log\left((\frac{\mu}{m})^\mu (\frac{1-\mu}{1-m})^{1-\mu}\right)} \\
&= \log\left((1 + \frac{(\mu - m)(1 - m)}{c})^\mu (1 + \frac{(\mu - m)(-m)}{c})^{1-\mu} - (\frac{\mu}{m})^\mu (\frac{1 - \mu}{1 - m})^{1-\mu}\right)
\end{aligned}$$

□

A.5. Proof of Proposition 1

Proof of Proposition 1. We prove this theorem for the single arm case $K_t(a) = \sum_{i=1}^t \left(1 + \frac{(\hat{\mu}_{i-1} - m)(X_i - m)}{c}\right)$ - note that this theorem can be generalized to the multi-armed case by taking partial derivatives, which reduces into the single arm case for $K_t^a(a) = \sum_{i=1}^t \mathbf{1}[A_t = a] \left(1 + \frac{(\hat{\mu}_{i-1}(a) - m_a)(X_i - m_a)}{c}\right)$. We prove this statement through induction, under the

assumption that $K_t(m)$ is strictly convex for any $m \in \mathbb{R}$, i.e. its derivative is strictly monotone increasing for any realization of $(X_i)_{i=1}^t$.

As the base case for $t = 1$, for any $X_1 \in [0, 1]$, (1) $K_t(m)$ is clearly convex for any $m \in \mathbb{R}$, (2) $\frac{\delta}{\delta m} K_t(m)$ is strictly monotone increasing for $m \in \mathbb{R}$, and (3) there exists a unique minimizing value m_1^* between $[0, 1]$ that satisfies $\sum_{i=1}^t \frac{2m - \hat{\mu}_0 - X_1}{c + (\hat{\mu}_0 - m)(X_i - m)} = 0$ and $K_t(m_t^*) \geq 0$.

$$K_t(m) = \frac{c + (\hat{\mu}_0 - m)(X_1 - m)}{c}, \quad \frac{\delta}{\delta m} K_t(m) = (2m - \hat{\mu}_0 - X_1)/c, \quad m_t^* = \frac{\hat{\mu}_0 + x_1}{2} \in [0, 1].$$

We now turn to the induction step. By our inductive hypothesis, we assume that (1) $K_{t-1}(m)$ is strictly convex for $m \in \mathbb{R}$, (2) $\frac{\delta}{\delta m} K_{t-1}(m)$ is strictly monotonically increasing for $m \in \mathbb{R}$, and (3) the minimizing value m_{t-1}^* is unique, between $[0, 1]$, and $K_{t-1}(m_{t-1}^*) \geq 0$. We first rewrite the derivative of $K_t(m)$ as the following:

$$\begin{aligned}
\frac{\delta}{\delta m} K_t(m) &= \sum_{i=1}^{t} (2m - \hat{\mu}_{i-1} - X_i) \left[\prod_{j \neq i}^{t} c + (\hat{\mu}_{i-1} - m)(X_i - m) \right] \\
&= (c + (\hat{\mu}_{i-1} - m)(X_i - m)) \sum_{i=1}^{t-1} (2m - \hat{\mu}_{i-1} - X_i) \left[\prod_{j \neq i}^{t-1} c + (\hat{\mu}_{i-1} - m)(X_i - m) \right] \\
&\quad + (2m - \hat{\mu}_{t-1} - X_i) \left[\prod_{i=1}^{t-1} c + (\hat{\mu}_{i-1} - m)(X_i - m) \right]
\end{aligned}$$

To show this is monotone, we use the sign of the second derivative.

$$\begin{aligned}
\frac{\delta^2}{\delta m^2} K_t(m) &= 2 \left[\prod_{i=1}^{t-1} c + (\hat{\mu}_{i-1} - m)(X_i - m) \right] \\
&\quad + 2[(2m - \hat{\mu}_{t-1} - X_t)] \sum_{i=1}^{t-1} (2m - \hat{\mu}_{i-1} - X_i) \left[\prod_{j \neq i}^{t-1} c + (\hat{\mu}_{i-1} - m)(X_i - m) \right] \\
&\quad + (c + (\hat{\mu}_{i-1} - m)(X_i - m)) \frac{\delta}{\delta m} \sum_{i=1}^{t-1} (2m - \hat{\mu}_{i-1} - X_i) \left[\prod_{j \neq i}^{t-1} c + (\hat{\mu}_{i-1} - m)(X_i - m) \right] \\
&= 2c^{t-1} K_{t-1}(m) + 2c^{t-1}[(2m - \hat{\mu}_{t-1} - X_t)] \frac{\delta}{\delta m} K_{t-1}(m) + c^{t-1}(c + (\hat{\mu}_{i-1} - m)(X_i - m)) \frac{\delta^2}{\delta m^2} K_{t-1}(m).
\end{aligned}$$

Using the fact that $c^{t-1}(c + (\hat{\mu}_{i-1} - m)(X_i - m)) \frac{\delta^2}{\delta m^2} K_{t-1}(m) > 0$ for all $m \in [0, 1]$, we obtain the following inequality for the second derivative expression:

$$\frac{\delta^2}{\delta m^2} K_t(m) > 2c^{t-1}(K_{t-1}(m) + (2m - \hat{\mu}_{t-1} - X_t) \frac{\delta}{\delta m} K_{t-1}(m))$$

By our inductive hypothesis, (i) for any $\tilde{m} \in \mathbb{R}$, $K_{t-1}(\tilde{m}) \geq K_{t-1}(m_{t-1}^*) \geq 0$, and (ii) by convexity $K_{t-1}(m) \geq K_{t-1}(\tilde{m}) + (m - \tilde{m}) \frac{\delta}{\delta m} K_{t-1}(m)$. Using these assumptions, we obtain another lower bound on the second derivative:

$$\frac{\delta^2}{\delta m^2} K_t(m) > 2c^{t-1}(K_{t-1}(\tilde{m}) + (m - \tilde{m} + 2m - \hat{\mu}_{t-1} - X_t) \frac{\delta}{\delta m} K_{t-1}(m)), \quad \forall \tilde{m} \in \mathbb{R}.$$

Because $\tilde{m}$ is any real number, we can pick $\tilde{m} = (3m - \hat{\mu}_{t-1} - X_t)$ to obtain that $\frac{\delta^2}{\delta m^2} K_t(m) > 2c^{t-1} K_{t-1}(\tilde{m}) \geq 0$. Thus, for all $m \in \mathbb{R}$, $\frac{\delta^2}{\delta m^2} K_t(m) > 0$, and so (1) $\frac{\delta}{\delta m} K_t(m)$ is strictly monotonically increasing and (2) $K_t(m)$ is strictly convex for all $m \in \mathbb{R}$. By the monotonocity of the derivative, there exists only one $m_t^* \in \mathbb{R}$ where $\frac{\delta}{\delta m} K_t(m) = 0$, which

must be a minimum by convexity. To show that $m_t^* \in [0,1]$, we consider the derivative function again:

$$\begin{aligned}
\frac{\delta}{\delta m} K_t(m) &= \sum_{i=1}^{t} (2m - \hat{\mu}_{i-1} - X_i) \left[\prod_{j \neq i}^{t} c + (\hat{\mu}_{i-1} - m)(X_i - m) \right] \\
&= (c + (\hat{\mu}_{i-1} - m)(X_i - m)) \sum_{i=1}^{t-1} (2m - \hat{\mu}_{i-1} - X_i) \left[\prod_{j \neq i}^{t-1} c + (\hat{\mu}_{i-1} - m)(X_i - m) \right] \\
&\quad + (2m - \hat{\mu}_{t-1} - X_i) \left[\prod_{i=1}^{t-1} c + (\hat{\mu}_{i-1} - m)(X_i - m) \right] \\
&= c^{t-1}(c + (\hat{\mu}_{i-1} - m)(X_i - m)) \frac{\delta}{\delta m} K_{t-1}(m) + c^{t-1}(2m - \hat{\mu}_{t-1} - X_i) K_{t-1}(m)
\end{aligned}$$

We claim that $\min(m_{t-1}^*, \frac{\hat{\mu}_{t-1}+X_t}{2}) \leq m_t^* \leq \max(m_{t-1}^*, \frac{\hat{\mu}_{t-1}+X_t}{2})$. Assume this to be false: then, one of the following must be true.

- Case 1: $m < \min(m_{t-1}^*, \frac{\hat{\mu}_{t-1}+X_t}{2})$.

  In this case, $(2m - \hat{\mu}_{t-1} - X_t) < 0$. Because m_{t-1}^* is the unique point where $\frac{\delta}{\delta m} K_{t-1}(m) = 0$, and $\frac{\delta}{\delta m} K_{t-1}(m)$ is monotonic, $\frac{\delta}{\delta m} K_{t-1}(m) < 0$. Note that $c^{t-1}(c + (\hat{\mu}_{i-1} - m)(X_i - m)) > 0$, and $c^{t-1} K_{t-1}(m) > 0$, and thus $\frac{\delta}{\delta m} K_t(m) < 0$.

- Case 2: $m > \max(m_{t-1}^*, \frac{\hat{\mu}_{t-1}+X_t}{2})$.

  In this case, $(2m - \hat{\mu}_{t-1} - X_t) > 0$. Because m_{t-1}^* is the unique point where $\frac{\delta}{\delta m} K_{t-1}(m) = 0$, and $\frac{\delta}{\delta m} K_{t-1}(m)$ is monotonic, $\frac{\delta}{\delta m} K_{t-1}(m) > 0$. Note that $c^{t-1}(c + (\hat{\mu}_{i-1} - m)(X_i - m)) > 0$, and $c^{t-1} K_{t-1}(m) > 0$, and thus $\frac{\delta}{\delta m} K_t(m) > 0$.

Neither case can be true, and therefore $m_t^* \in [\min(m_{t-1}^*, \frac{\hat{\mu}_{t-1}+X_t}{2}), \max(m_{t-1}^*, \frac{\hat{\mu}_{t-1}+X_t}{2})]$, which is a subset of $[0,1]$ by the assumption that $m_{t-1}^* \in [0,1]$ and $\hat{\mu}_{t-1} + X_t \in [0,2]$. For any $m \in [0,1]$, $K_t(m) > 0$ by definition, and so (3) m_t^*, the unique minimizer of $K_t(m)$, lies within $[0,1]$ and $K_t(m_t^*) \geq 0$. This confirms our inductive hypothesis, and so for all $t \in \mathbb{N}$, $K_t(m)$ is a convex function with a unique minimizer m_t^* satisfying $\frac{\delta}{\delta m} K_t(m_t^*) = 0$.

□

Proofs for Section 4

A.6. Proof of Theorem 3

Proof of Theorem 3. We first note that because $E_t(m) = \frac{1}{W} \sum_{a \in [W]} K_t^a(m_a)$, and $K_t^a(m_a) \geq 0$ for all $a \in [A]$, $t \in \mathbb{N}$, $E_t(m)$ is nonnegative. We now show that this process is a martingale for $m = \mu$, regardless of sampling policy:

$$\begin{aligned}
\mathbb{E}_t[E_t(m) | \mathcal{F}_{t-1}] &= \sum_{a \in [W]} \pi_t(a) \left[(1 - \pi_t(a)) + \pi_t(a) * (1 - 0) \right] * E_{t-1}(m) \\
&= E_{t-1}(m).
\end{aligned}$$

Thus, by Ville's inequality 5, we obtain that the crossing probability $\mathbb{P}(\exists t \in \mathbb{N} : T_t(m) = \mathbf{1}[E_t(m) \geq 1/\alpha] = 1) \leq 1/\alpha$, guaranteeing our desired Type I error coverage. □

A.7. Proof of Theorem 4

Proof of Theorem 4. We prove this statement by building upon the proof of Theorem 2 and using Lemma 6 below:

Lemma 6 (Fact E.1 of Shin et al., 2021)**.** *Suppose that $Y_n \to Y$ a.s. as $n \to \infty$, and $N(t) \to \infty$ a.s. as $t \to \infty$. Then $Y_{N(t)} \to Y$ a.s. as $t \to \infty$.*

This establishes necessary convergence results for the assumptions we place on our policy, i.e. that for our sampling scheme π, $N_t(a) \to \infty$ a.s. for some $a \in \mathcal{I}(m)$. Let $\omega \in \Omega$ denote an event / sample path, where $\mathbb{P}(\Omega) = 1$. Because there exists

an arm $a^*(\omega) \in \mathcal{I}(\omega)$ such that $N_t(a^*(\omega))(\omega) \to \infty$ a.s., $K_t^{a^*(\omega)}(m_{a^*(\omega)}) \to \infty$ a.s. by Theorem 2, and $K_t^a(m_a) \geq 0$ for all $a \neq a^*(\omega)$, $E_t(m)(\omega) \to \infty$ almost surely. By the same argument as Theorem 2, this implies that the stopping time for test $T_t(m, \alpha)$ is finite for any fixed $m \neq \mu$, $\alpha \in (0, 1)$.

A.8. Proof of Lemma 4

This follows from the fact that for all $t \in \mathbb{N}$, $(A_i, X_i)_{i=1}^t \in ([W] \times [0, 1])^t$, $K_t^a(m_a)$ is nonnegative. Thus,

$$\frac{1}{W} \sum_{a \in [W]} K_t^a(a) \geq \frac{1}{W} K_t^a(m_a).$$

Note that $T_t'(m, \alpha) = \prod_{a \in [W]} \mathbf{1}[\max_{1 \leq i \leq t} K_i^a(m_a) \geq W/\alpha] = 1$, then there exists a time $j \in [1, ..., t]$ such that $K_t^a(m_a) \geq W/\alpha$ for some a. Then, for the same time $j \leq t$, $\sum_{a \in [W]} K_j^a(a) \geq W/\alpha$, which gives the desired result:

$$T_i(m, \alpha) = \mathbf{1}\left[\max_{1 \leq i \leq t} \sum_{a \in [W]} K_t^a(a) \geq \frac{W}{\alpha}\right] = 1.$$

.

A.9. Proof of Proposition 3

At time t, denote the global minimizer of $E_t(m)$ solution as $\tilde{m}$, which is given entry-wise by the following condition:

$$\frac{\delta}{\delta m_a} E_t(m) = 0 \iff \sum_{i=1}^t \mathbf{1}[A_t = a] \frac{2\tilde{m}_a - X_i - \hat{\mu}_{i-1}(a)}{c + (\hat{\mu}_{i-1}(a))} = 0 \quad \forall a \in [W].$$

Recall that our proposed solution is given by the form, with respect to partition $R_{\text{BAI}}(a')$, is given by the following form:

$$m_a^* = \begin{cases} \tilde{m}_a & \text{if } \tilde{m}_a < q \\ q & \text{if } \tilde{m}_a \geq q \end{cases}$$

where $q \in [0, 1]$ is a constant that satisfies

$$\sum_{a \in W : \tilde{m}_a \geq q} \gamma(m^*, a) \sum_{i=1}^t \mathbf{1}[A_t = a] \frac{2q - X_i - \hat{\mu}_{i-1}(a)}{c + (\hat{\mu}_{i-1}(a) - q)(X_i - q)} = 0,$$

where $\gamma(m, a) = \prod_{i=1}^t (\mathbf{1}[A_i = a](1 + (X_i - m_a)(\hat{\mu}_{i-1}(a) - m_a)/c) + (1 - \mathbf{1}[A_i = a]))$.

We verify that this is solution is indeed optimal through the KKT conditions (Chapt. 5, Boyd and Vandenberghe, 2004). To see this, we first convert our problem to standard convex optimization notation:

$$\min_m \frac{1}{W} \sum_{a \in [W]} K_t^a(m_a)$$
$$\text{s.t. } \mu_a - \mu_{a'} \leq 0 \quad \forall i \in [W] \setminus \{a'\}.$$

The lagrangian dual of this problem takes the form, for dual variable $\lambda \in \mathbb{R}^{W-1}$:

$$\mathcal{L}(m, \lambda) = \frac{1}{W} \sum_{a \in [W]} K_t^a(m_a) + \sum_{a \neq a'} \lambda_a (m_a - m_{a'}) \quad \text{s.t. } \lambda \geq 0.$$

Our proposed solution in terms of the Lagrangian is as follows:

$$\lambda_a = \begin{cases} 0 & \text{if } \tilde{m}_a < q \\ -\frac{1}{W} \frac{\delta}{\delta m_a} K_t^a(q) & \text{if } \tilde{m}_a \geq q \end{cases} \quad m_a^* = \begin{cases} \tilde{m}_a & \text{if } \tilde{m}_a < q \\ q & \text{if } \tilde{m}_a \geq q \end{cases}.$$

Note that $m_a^* \leq q = m_{a'}^*$, which satisfies the (1) primal feasibility requirements. For (2) dual feasibility, note that $\lambda_a = 0$ for $\tilde{m}_a < q$, and $-\frac{1}{W}\frac{\delta}{\delta m_a}K_t^a(q) > 0$ for $\tilde{m}_a \geq q$, which satisfies dual feasibility. The (3) complementary slackness condition is also satisfied, i.e.:

$$\lambda_a(m_a^* - m_{a'}^*) = 0 \quad \forall a \in [W] \setminus \{a'\}.$$

Lastly, to check (4) Lagrangian stationary, we begin with a such that $\tilde{m}_a < \tilde{m}_{a'}$:

$$\frac{\delta}{\delta m_a}L(m^*) = \frac{1}{W}\frac{\delta}{\delta m_a}K_t^a(m_a) = 0.$$

For a such that $\tilde{m}_a > \tilde{m}_{a'}$,

$$\frac{\delta}{\delta m_a}L(m^*) = \frac{1}{W}\frac{\delta}{\delta m_a}K_t^a(q) - \frac{\delta}{\delta m_a}K_t^a(q) = 0.$$

Finally, for $m_{a'}$,

$$\frac{\delta}{\delta m_{a'}}L(m^*) = \frac{1}{W}\frac{\delta}{\delta m_a}K_t^{a'}(q) - \sum_{a:\tilde{m}_a \geq q}\lambda_a = \frac{1}{W}\sum_{a\in[W]:\tilde{m}_a\geq q}\frac{\delta}{\delta m_a}K_t^a(q).$$

Note that this equation is solved by the condition:

$$\sum_{a\in[W]:\tilde{m}_a\geq q}\gamma(m,a)\sum_{i=1}^{t}\mathbf{1}[A_t = a']\frac{2m_a - X_i - \hat{\mu}_{i-1}(a)}{c + (X_i - m_a)(\hat{\mu}_{i-1}(a') - m_a)} = 0,$$

where $\gamma(m,a) = \prod_{i=1}^{t}\left(\mathbf{1}[A_i = a](c + (X_i - m_a)(\hat{\mu}_{i-1}(a) - m_a)) + (1 - \mathbf{1}[A_i = a])\right)$, and thus our solution solves this equation. Note that Slater's condition (pg. 226, Boyd and Vandenberghe, 2004) is trivially satisfied by $R_{\text{BAI}}(a')$ for any $a' \in [W]$ (i.e. feasible region has at least one interior point), and thus strong duality holds. Then, our KKT conditions are necessary and sufficient conditions for optimality, and therefore our solution solves our primal constrained minimization problem.

□

B. Experiment Details.

In this section, we provide additional details on the baseline methods of comparison, the data-adaptive sampling schemes used by in Section 5.2, additional details regarding Table 5.1, and additional simulations using non-standard underlying distributions for each arm.

B.1. Additional Details on Baseline Confidence Sequences / Anytime Valid Tests.

We provide additional details on confidence sequences used in the single arm experiments in Section 5.

Sub-Bernoulli. We use the package `confseq` (Howard et al., 2021b), and use the function `bernoulli_confidence_interval` with parameters optimal intrinsic time equal to $t = 2500$, and $\alpha = \alpha_{\text{opt}} = 0.05$.

PrPlH. The confidence sequence for PrPlH (Waudby-Smith and Ramdas, 2023) is given by:

$$C_t^{\text{PrPlH}} = \left(\frac{\sum_{i=1}^t \lambda_i X_i}{\sum_{i=1}^t \lambda_i} \pm \frac{\log(2/\alpha) + \sum_{i=1}^t \lambda_i^2/8}{\sum_{i=1}^t \lambda_i}\right), \quad \lambda_i = \min(\sqrt{\frac{8\log(2/\alpha)}{t\log(t+1)}}, 1).$$

Empirical Bernstein. The confidence sequence for Emp. Bern. (Waudby-Smith and Ramdas, 2023) is given by:

$$C_t^{\text{Emp. Bern.}} = \left(\frac{\sum_{i=1}^t \lambda_i X_i}{\sum_{i=1}^t \lambda_i} \pm \frac{\log(2/\alpha) + \sum_{i=1}^t v_i \psi_E(\lambda_i)}{\sum_{i=1}^t \lambda_i}\right).$$

where $\lambda_t^{\text{PrPl}} = \min(\sqrt{\frac{2\log(2/\alpha)}{\hat{\sigma}^2 t - 1 t\log(t+1)}}, 1/2)$, $\hat{\sigma}^2 = \frac{1/4 + \sum_{i=1}^t (X_i - \hat{\mu}_i)^2}{t+1}$, $\hat{\mu}_t = \frac{1/2 + \sum_{i=1}^t X_i}{t+1}$, $v_t = 4(X_i - \hat{\mu}_{i-1})^2$, and $\psi_E(\lambda) = (-\log(1-\lambda) - \lambda)/4$.

Hedged. We use the empirical-bernstein based variant of the hedged capital process discussed in Waudby-Smith and Ramdas (2023):

$$\begin{aligned}
\kappa_t^{\pm}(m) &= \max\left(\theta\kappa_t^{+}(m), (1-\theta)\kappa_t^{-}(m)\right), \\
\kappa_t^{+}(m) &= \prod_{i=1}^{t}(1+\lambda_i^{+}(m)(X_i - m)) \\
\kappa_t^{-}(m) &= \prod_{i=1}^{t}(1-\lambda_i^{-}(m)(X_i - m)), \\
C_t^{\text{Hedged}} &= \{m \in [0,1] : \kappa_t^{\pm} < 1/\alpha\}.
\end{aligned}$$

with $\theta = 1/2$, 100 grid points for each simulation, and $\lambda_t^{+} = \lambda_t^{-} = \lambda_t^{\text{PrPl}}$ as defined for the Emp. Bern. approach.

Baseline Methods for Multi-Arm Case. The termination condition for a single arm a (i.e. below/above threshold ϵ) is determined by the following anytime valid confidence interval in (Kano et al., 2019), which we refer to as C_t^{base}:

$$C_t^{\text{base}}(a) = \left\{\hat{\mu}_t(a) \pm \sqrt{\frac{\log(4WN_t^2(a)/\alpha)}{2N_t(a)}}\right\}.$$

The baseline confidence sequence used in Jamieson and Nowak (2014) for BAI is given by:

$$C_t^{\text{base}}(a) = \left\{\hat{\mu}_t(a) \pm \sqrt{\frac{\log(\frac{405.5Wt^{1.1}}{\alpha}\log(\frac{405.5Wt^{1}.1}{\delta}))}{2N_t(a)}}\right\}.$$

For Hedged and Sub-B., we generalize them to the multi-arm case by dividing $\alpha = 0.05$ by W to maintain our guarantees.

B.2. Additional Details For Sampling Schemes in Multi-Arm Case.

We provide the exact sampling schemes used for Section 5's experiments regarding threshold identification (THR) and best-arm identification (BAI).

Threshold Identification. We set threshold $\xi = 0.5$. We use the sampling algorithm HDoC proposed by Kano et al., 2019: at time t, HDoC sample arm a_t^{HDoC}, defined as $a_t^{\text{HDoC}} = \arg\max_{a\in[K]} \hat{\mu}_t(a) + \sqrt{\frac{\log(t)}{2N_t(a)}}$. We report the first time in which we have rejected i regions as τ_i, i.e. τ_i corresponds to the first time that we have labeled i arms in Table B.5, and the final stopping time (i.e. τ_4) in Table 5.1 in the main body of the text.

Best Arm Identification. Each arm a in the BAI problem has a single corresponding hypothesis region given in Equation (3). We use the Lower-Upper Confidence Bound (LUCB) sampling scheme (Kaufmann and Kalyanakrishnan, 2013): at time t, LUCB samples two arms: arm $a_t^U = \max_{a\in[W]} \hat{\mu}_t$ and $a_t^L = \max_{a\in[W]\setminus a_t^U} \hat{\mu}_t + \sqrt{\log\left(\frac{405.5At^{1.1}}{\alpha}\log(\frac{405.5At^{1.1}}{\alpha})\right)/2N_t(a)}$. Our stopping time τ is defined as the first time in which we have eliminated all regions but one, where the last non-rejected region $R_{\text{BAI}}(a)$ corresponds to the best arm a.

B.3. Runtime Testing.

While all tests were done in R, we test the runtime of each approach in Python. The grid size for Hedged is 100, 200, and 400 evenly spaced points on $[0,1]$ (i.e. fidelity $0.01, 0.005, 0.0025$ respectively), and tested in the manner of the description in Table 2 on page 50 of Waudby-Smith and Ramdas (2023). All runtimes were done locally on an Apple M2 Pro Chip, 16 Gb of RAM, with no parallelization.

B.4. Additional Simulations with Nonstandard Distributions

To emphasize that PEAK performs well across all arbitrary data distributions over $[0,1]$, we provide additional results that test nonstandard distributions beyond exponential families. In particular, we test the following distributions, and provide additional results that demonstrate the robustness of PEAK's performance.

- **Mixture Distributions for a Single Stream.** In the single-stream case (Section 5.1), we test the mixture distribution, where observations are generated from $\text{Unif}(0, 1)$, $\text{Beta}(1, 1)$, and $\text{Bern}(0.5)$ with probability 1/3 each. Figure B.1 plots the width of the confidence sequence at time t (i.e. the volume of the null hypothesis set we cannot reject at time t) formed by inverting PEAK, which empirically demonstrates that PEAK does not suffer (and is relatively better) than the other methods we test.

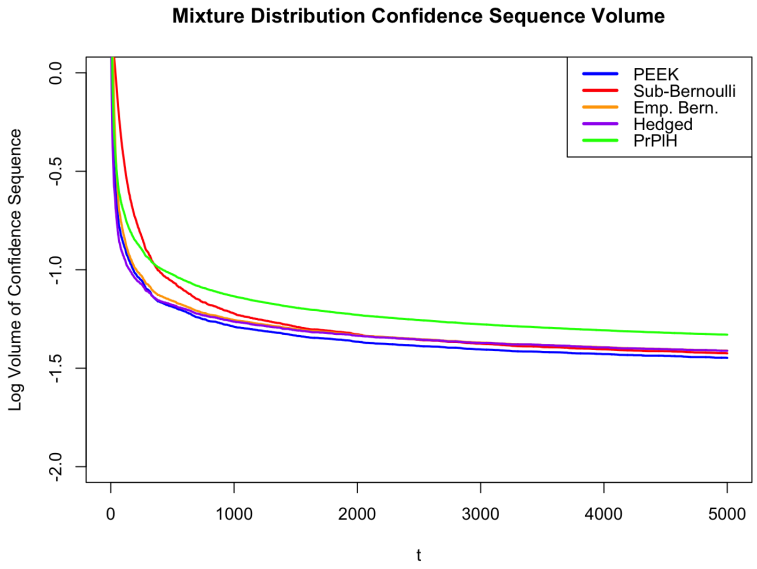

Figure B.1. Confidence Sequence Widths in the Single Arm Setting for Mixture Distribution Arm

- **Mixture Distribution for BAI.** For the multi-stream best-arm identification problem, we test distributions with adversarial heavy tails (outliers) in the best-arm identification setting by modifying our Beta distribution set-up. In our new simulations, we contaminate the distribution of arm 3 (second-best arm) with a point mass of 0.05 at 1, and contaminate the distribution of arm 4 (best arm) with a point mass of 0.05 at 0. To maintain the same μ across simulations, we choose Beta distributions with parameters $\alpha = 1$ for both arms 3 and 4, and $\beta = 0.43/0.52$, $\beta = 0.24/0.71$ for arms 3 and 4 respectively. The updated Table provides empirical results regarding the stopping times, which align closely to the uncontaminated low-variance (uncontaminated Beta) setting.

Type	Stopping Rule	Stopping Time
Beta	PEAK	708.52 ± 266.34
	Base	4686.1 ± 565.24
	Sub-B	3175.3 ± 905.70
	Hedged	$\mathbf{500.12} \pm \mathbf{214.08}$
Bern	PEAK	$\mathbf{1318.14} \pm \mathbf{489.29}$
	Base	4631.66 ± 896.51
	Sub-B	4322.36 ± 3638.25
	Hedged	1734.64 ± 858.66
Beta-Contaminated	PEAK	705.72 ± 284.76
	Base	4680.92 ± 676.86
	Sub-B	3376.22 ± 941.16
	Hedged	$\mathbf{542.42} \pm \mathbf{246.24}$

Table B.1. Stopping Times for BAI, including contaminated distribution example.

B.5. Additional Threshold Identification Results

We now provide the average stopping time for identifying each arm sequentially, where τ_i is the first time in which i arms are been labeled as above/below the threshold.

Arm Type	Stopping Rule	τ_1	τ_2	τ_3	τ_4
Beta	PEAK	**77.49 ± 34.62**	**461.78 ± 185.64**	**554.52 ± 178.33**	**725.81 ± 218.49**
	Base	253.71 ± 65.39	2391.64 ± 422.35	4476.97 ± 479.41	4534.31 ± 482.94
	Sub-B	199.98 ± 39.06	915.92 ± 219.18	1562.18 ± 230.75	1618.32 ± 229.90
	Hedged	**65.75 ± 28.73**	**316.74 ± 165.11**	**383.05 ± 161.48**	**479.16 ± 205.97**
Bernoulli	PEAK	**106.75 ± 70.29**	**905.64 ± 479.17**	**1329.32 ± 513.84**	**1678.08± 666.39**
	Base	253.83 ± 98.52	2479.43 ± 834.43	4731.87 ± 1107.71	4795.51 ± 1101.09
	Sub-B	208.85 ± 64.17	**944.26 ± 347.06**	1624.80 ± 464.31	**1686.81 ± 456.08**
	Hedged	**104.77 ± 70.57**	1212.75 ± 817.87	**1582.60 ± 867.27**	2241.18 ± 1092.49

Table B.2. Stopping Times τ for THR. Each τ_i represents the first time in which we have classified i arms as above or below $\xi = 0.5$. We report averaged stopping times and one standard error over 100 simulated sample paths. Bolded values represent the best and second-best average run-time respectively. Across all simulations, each stopping time results in the correct conclusion (i.e. rejects the incorrect hypothesis).

B.6. Visualization for Composite Hypothesis Testing for THR and BAI

We visualize Examples 1 and 2 in Section 4, providing simple intuition for the $W = 2$ case. Below, we use m_t^* to denote the global minimizer of $E_t(m)$, and m_t^i to denote the minimizer of $E_t(m)$ for region R_i. These simple plots demonstrate the intuitive solutions implied by solving the KKT system of equations.

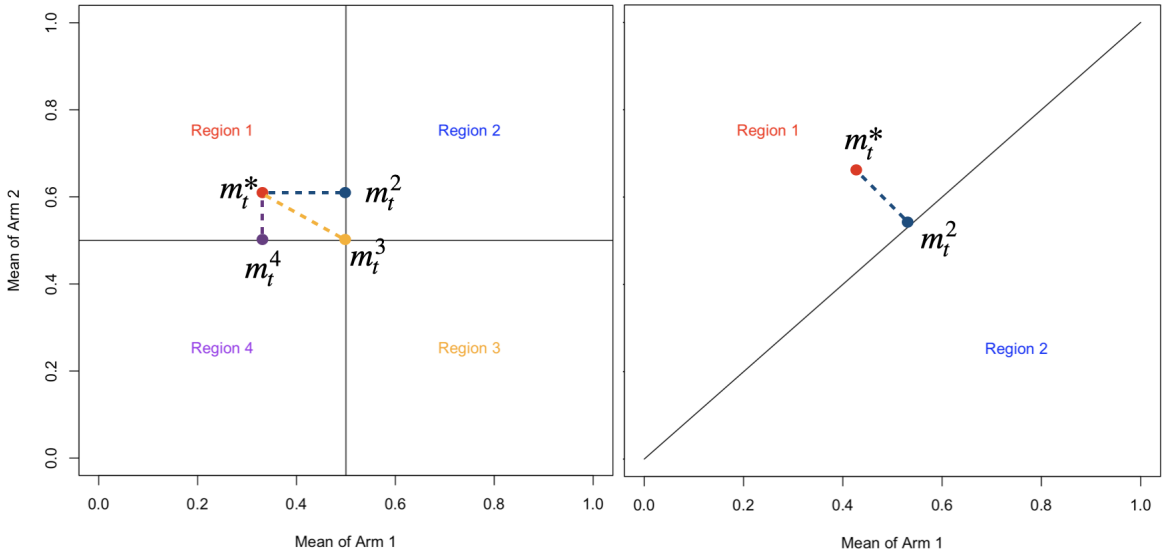

Figure B.2. Visualization of Minima within each region at time t, obtained by projecting the global minimizer of $E_t(m)$ onto the regions implied by THR (left) and BAI (right) for $W = 2$ case. In both plots, the current global minima at time t is contained in Region 1, and projected to obtain the minima in all other regions.