Trustworthy Bayesian Inference in Batch-Adaptive Experimentation

Yicheng Li*, Yan Wang*, Jiayun Dong*
*Airbnb, San Francisco, CA, USA

I. Introduction

Large-scale online A/B experiment is central to data-driven decision making in industry [5]. Adaptive experimentation, particularly bandit-based methods, has been widely studied for improving experiment efficiency by reallocating traffic toward promising treatments [1]. However, most of this literature emphasizes regret minimization or best-arm identification, while practical experimentation requires valid post-experiment inference, including credible treatment effect estimates and formal hypothesis testing with principled error control across large experiment portfolios [2, 4, 3]. This paper addresses this gap by introducing a Bayesian framework for batch-adaptive experiments that is tailored to the operational realities of online platforms.

A. Challenges and Gaps

In a real-world adaptive A/B experiment, data are generated not only from newly arriving users but also from existing cohorts that continue to contribute outcomes as they age. This repeated and refreshed observation structure is common in practice, yet it has received little rigorous attention in the literature on batch adaptive experimentation. Current approaches largely overlook the persistence and cohort-aging dynamics that characterize actual online experimentation [12, 7].

Moreover, a central challenge in batch adaptive experimentation is the lack of tractable likelihood formulations. Most of the bandit literature is designed for regret minimization and allocation efficiency rather than post-experiment statistical inference [1, 10]. The relatively few works that study inference focus on establishing asymptotic normality or coverage guarantees under frequentist paradigms [6, 4, 2, 7], yet these results remain technically complex, lack asymptotic optimality characterizations, and are not easily applied in practice. A flexible likelihood framework that can accommodate a wide range of bandit algorithms—and that could serve as the foundation for valid statistical inference—has not been developed, let alone one tailored to the cohort-based data structures discussed above.

Finally, reliable impact estimation and formal hypothesis testing are essential for experiment-driven decision making in industry, where product launches and business strategies depend on trustworthy evidence. However, most bandit algorithms are designed to optimize allocation and regret, and they provide little or no guarantee about the validity of post-experiment effect estimates or error rates [8, 9]. As a result, practitioners face a gap between algorithms that efficiently allocate traffic during an experiment and statistical procedures that ensure credible inference once the data are collected.

B. Contributions

This paper develops a practical inference framework for batch adaptive experimentation that addresses these challenges. To capture the realities of online platforms, we introduce a data model that explicitly accounts for repeated and refreshed cohort observations, filling a gap overlooked in prior work. Building on this structure, we derive a novel likelihood approximation that converts adaptive histories into a tractable linear model, enabling flexible statistical inference across a wide range of bandit algorithms. In addition, we establish decision procedures that link evidence thresholds to reliable error control, providing practitioners with valid tools for estimating impacts and testing hypotheses in large-scale adaptive A/B experiments. Finally, we offer a complete guideline on batch-adaptive experiment design to ensure trustworthy inference in adaptive experimentation.

II. Methodology

A. Batch Adaptive Experiment Overview

Consider a two-arm adaptive experiment with a reference arm and a treatment arm. In a typical real-world industry setting, at each time t, a new cohort of subjects enters the experiment, with the arm traffic allocation determined based on data collected from prior periods. Concurrently, all existing cohorts continue to observe updated data under their original treatment assignments. Let $X^t_{a,i,j}$ represent the observed effect for subject j in cohort i at time t under treatment arm a ($a = 0$ for the reference arm and $a = 1$ for the treatment arm), modeled as:

$$X^t_{a,i,j} = \mu(a, i, t) + \epsilon_j,$$

where $\mu(a, i, t)$ denotes the expected effect, and ϵ_j is a mean-zero random variable with finite variance. For the purposes of this manuscript, we focus on the stationary expected effect setting, where $\mu(a, i, t) = \mu(a)$ across time and cohorts (see Appendix D for an age-specific non-stationary setting). The arm assignment is determined through a **one-time** sampling process based on the arm assignment rate $\pi_{a,i}$ at the time the cohort first enters the experiment. At any given time t, the complete set of average observed effects

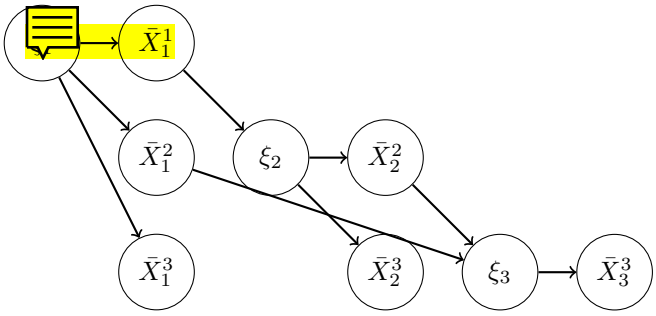


Fig. 1: A Directed Acyclic Graph (DAG) illustrating the data generation process in a typical adaptive experiment under Assumption 1. The superscript t indicates the time of observation and supscript i indicates the cohort (i.e., the time when the cohort first arrives in the experiment). Arm index a is suppressed for simplicity.

across all cohorts, each with sample size N_i, is denoted as $\bar{\mathbb{X}}^t := \{\bar{X}^t_{a,i}\}_{a\in[0,1],i\in[1,\dots,t]}$, where

$$\bar{X}^t_{a,i} = \frac{1}{N_i}\sum_{j=1}^{N_i} \xi_{a,i,j} X^t_{a,i,j}.$$

Our inference framework is designed to be agnostic to the specific bandit algorithm used for determining the arm assignment rates, provided the following assumption is satisfied:

Assumption 1 The cohort-specific arm assignment rate $\pi_{t+1} := \{\pi_{0,t+1}, \pi_{1,t+1}\}$ for the newly arriving cohort at time $t+1$ is fully determined by the average observed effects at time t, such that $\pi_{t+1} = \pi(\bar{\mathbb{X}}^t)$, where $\pi(\cdot)$ is any deterministic function defined by the traffic allocation policy. Additionally, π_{t+1} remains positive for both arms.

Assumption 1 is practical for real-world adaptive experiments, as it is widely applicable to common multi-arm bandit algorithms and ensures that the most recent data is leveraged to determine traffic allocation for newly arriving cohorts. Figure 1 illustrates the complete data generation process of a typical adaptive experiment under Assumption 1.

B. Bayesian Inference Framework

The main inference goals of the Bayesian batch adaptive experiment include **estimation** of the expect effects of both arm (and their difference) and **decision-making** through Bayesian hypothesis testing. We start with the likelihood approximation for the batch adaptive experiment data generation process.

1) Likelihood Approximation: By data generation procedure described in Section II-A, the joint probability of the full history of the average observed effects with repeated measurement **up to time** t can be decomposed as:

$$\mathbb{P}(\bar{\mathbb{X}}^1, \cdots, \bar{\mathbb{X}}^t) = \prod_{i=2}^{t} \mathbb{P}(\bar{\mathbb{X}}^t_i \mid \bar{\mathbb{X}}^{i-1}) \times \mathbb{P}(\bar{\mathbb{X}}^t_1),$$

where $\bar{\mathbb{X}}^t_i$ is the full history of the average observed effects for cohort i by time t since the cohort entered the experiment.

To approximate each conditional probability in the likelihood, we need the following assumptions:

Assumption 2 $\xi_i \perp X^t_{i,j} | \bar{\mathbb{X}}^{i-1} \quad \forall t \geq i$. This assumption states that the treatment assignment ξ_i for cohort i is independent of the individual observed effects for all time after the cohort enters the experiment $t \geq i$, conditional on the data observed prior to cohort i.

Assumption 3 $X^t_{i,j} \perp X^{t+1}_{i,j} \perp X^{t+2}_{i,j} \perp \dots | \xi_i$. This assumption posits that the observed effects are independent across time periods once the arm assignment is fixed and can be easily relaxed by incorporating a more realistic covariance structure if needed.

To approximate the conditional likelihood of cohort-level averages, we leverage the multivariate Central Limit Theorem under Assumptions 1–3, which yields an asymptotically normal representation for the evolving cohort means with covariance reflecting both assignment probabilities and outcome variability. The formal statement is given in Proposition 1 (Appendix A). For tractability, we further approximate the joint probability of effects within a cohort under the two arms by assuming independence, noting that this can be relaxed by incorporating richer cross-arm covariance structure. We also reparameterize $\mu := \{\mu(0), \mu(1)\}$ into $\tilde{\mu} := \{\mu(0), \delta(1)\}$, where $\delta(1) = \mu(1) - \mu(0)$, which is often of primary interest and easier to elicit priors for.

Building on these approximations, the adaptive history admits a linear representation

$$X^t = \mathbf{A}^t \tilde{\mu} + \epsilon, \quad \epsilon \sim N(0, \Sigma_\epsilon),$$

where X^t stacks cohort-level outcomes, $\mathbf{A}^t$ encodes the adaptive allocation, and Σ_ϵ captures the induced covariance. The full statement and derivation appear in Proposition 2 (Appendix B).

This formulation offers a general framework for adaptive experiment designs under the minimal Assumptions 1–2, which yields a flexible likelihood that unifies a broad class of batch adaptive policies and hold broadly in industry practice. Extensions are possible by relaxing Assumption 3, incorporating arm correlations, or adapting to non-stationary effects and multi-arm settings through modifications of the design matrix $\mathbf{A}$ and covariance structure Σ_ϵ. For example, correlations between arms can be modeled by replacing the diagonal block $(\Sigma^t_{0,i}/N_i, \Sigma^t_{1,i}/N_i)$ in Σ_ϵ with a covariance matrix including off-diagonal terms. An age-specific non-stationary extension is provided in the Appendix D.

2) Effect Estimation: To perform full Bayesian posterior inference, we define the prior distribution of $\tilde{\mu}$ as a multivariate Gaussian distribution: $\tilde{\mu} \sim \mathcal{N}(\mu_{\tilde{\mu}}, \Sigma_{\tilde{\mu}})$ with

$$\mu_{\tilde{\mu}} = (\mu_{\mu_0} \;\; 0)', \qquad \Sigma_{\tilde{\mu}} = \mathbf{diag}(\sigma^2_{\mu_0}, \sigma^2_\delta).$$

Here μ_{μ_0} represents the prior mean of the reference arm, and $\sigma^2_{\mu_0}$ and σ^2_δ denote the prior variances of the reference arm and the treatment effect size, respectively. To utilize the Normal-Normal conjugacy, we propose employing a plug-in estimate for μ_a in the data covariance matrix $\Sigma^t_{a,i}$. Our simulation study demonstrates that using the posterior mean from the previous

iteration provides robust and reliable results. Consequently, the posterior distribution of $\tilde{\mu}$ follows a Gaussian distribution (formula omitted).

3) Decision Making: We perform Bayesian hypothesis testing for adaptive experiment using the Bayes Factor (BF) to test $H_0 : \delta(1) = 0$ against $H_1 : \delta(1) \sim \mathcal{N}(0, \sigma_\delta^2)$. The Bayes Factor for comparing hypotheses H_1 and H_0 is defined as $\text{BF}_{10} := \mathbb{P}(X|H_1)/\mathbb{P}(X|H_0)$, where $\mathbb{P}(X|H_k)$ is the marginal likelihood under hypothesis H_k following

$$X|H_k \sim \mathcal{N}(\mathbf{A}\mu_{H_k}, \mathbf{A}\Sigma_{H_k}\mathbf{A}' + \Sigma_\epsilon).$$

Here the μ_{H_k} and Σ_{H_k} are the mean vector and covariance matrix from prior of the corresponding hypothesis H_k. We reject H_0 in favor of H_1 if $\text{BF}_{10} > \text{BF}_{\text{threshold}}$ for a pre-specified threshold for evidence strength. Bayes Factors provide a natural link to error control in adaptive experiments, as they bound the false discovery rate (FDR, the probability of falsely declaring an effect when none exists among discoveries) while also relating to the false non-discovery rate (FNDR, the probability of missing a true effect among non-discoveries). Our simulation studies reveal how FDR, FNDR, sample size, and treatment effect variance interact, providing the basis for the detailed guidelines on trustworthy batch-adaptive experiment design presented in Appendix C.

III. Simulation Study

A. Simulation Setup

In this simulation study, we consider the following scenario under the stationary expected effects:

- H_0: $\mu(0) \sim \mathcal{N}(\mu_{\mu_0}, 1)$ and $\delta(1) = 0$,
- H_1: $\mu(0) \sim \mathcal{N}(\mu_{\mu_0}, 1)$ and $\delta(1) \sim N(0, \sigma_\delta^2)$.

We generate the individual treatment effects $X_{a,i,j}^t \sim_{i.i.d.} \mathcal{N}(\mu_a, 1)$. In practice, practitioners can standardize the observed effects by the estimated population standard deviation first and treat δ as the effect size. We choose the standard Thompson sampling with a clipping at 0.05 as an example of the arm assignment algorithm for illustration. We experiment with different settings including number of batches, batch sample sizes, and standard deviations of the effect size δ. We repeat each experiment 100,000 times, with equal sampling from H_0 and H_1 and prior distributions known. In practice, prior distributions and prior odds should be properly estimated from historical experiments or suggested by domain expertise.

B. Estimation Biasness

Figure 2 illustrates the bias distribution of the treatment effect. The bias of the estimated treatment effect is a non-increasing function of the true value, centered at 0. Notably, the estimate remains unbiased when there is no treatment effect. However, when a real effect exists, the estimate exhibits shrinkage, a common practice often referred to as applying a "haircut" in impact analysis. Interestingly, the bias for the inferior arm is positive due to the influence of the prior. This is significant because, in adaptive experiments, the post-experiment estimate of the inferior arm is typically negatively biased due to insufficient sampling of the inferior arm [11].

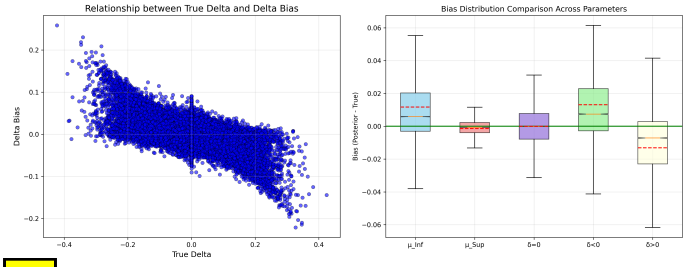


Fig. 2: Bias distribution under the stationary scenario with $\sigma_\delta^2 = 0.1$, consisting of 30 batches with batch sizes randomly sampled between 200 and 300. Left: Scatterplot showing the relationship between effect size bias and true values. Right: Boxplots illustrating the bias distribution by true effects, where μ_{inf} (resp. μ_{sup}) is arms with inferior (resp. superior) true effects. $\delta = 0$ is under H_0 and $\delta <> 0$ is under H_1.

C. Decision Making Errors

Figure 3 illustrates the FDR and FNDR under various experimental settings. The FDR is effectively controlled at the desired level by selecting an appropriate Bayes Factor threshold (e.g., $\text{BF}_{\text{threshold}} = 9$ for FDR ≤ 0.10). Meanwhile, for a fixed Bayes Factor threshold, the FNDR decreases as the total sample size or the standard deviation of the treatment effect increases. Additionally, we evaluated the FNDR against the batch size while keeping the total sample size constant. The results indicate that more frequent updates lead to improved FNDR and opportunity gain under a fixed sample size (see Figure 4). These findings provide practitioners with practical, statistically sound guidelines for designing batch adaptive experiments, as outlined in Appendix C.

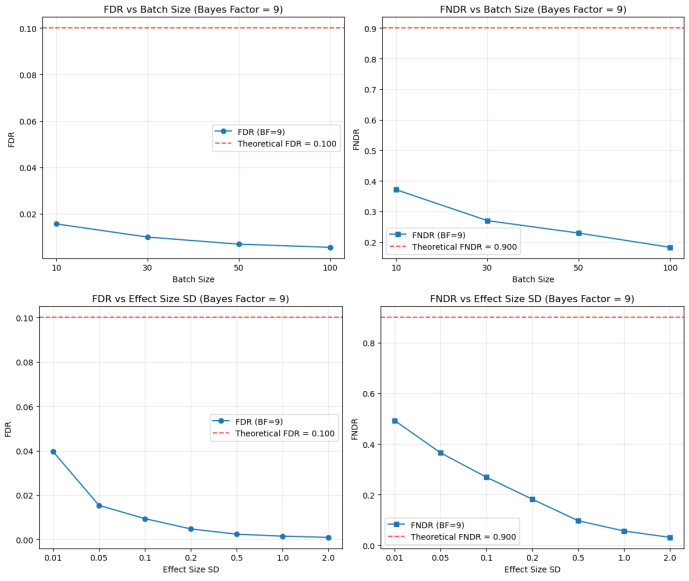


Fig. 3: False Discovery Rate (left) and False Non-Discovery Rate (right) under the stationary scenario with $\text{BF}_{\text{threshold}} = 9$. The in-batch sample size is randomly sampled between 200 and 300. Top: Variation in the number of batches. Bottom: Variation in σ_δ^2 with a fixed number of 30 batches.

References

[1] Shipra Agrawal and Navin Goyal. Analysis of thompson sampling for the multi-armed bandit problem. In *Conference on learning theory*, pages 39–1. JMLR Workshop and Conference Proceedings, 2012.

[2] Aurélien Bibaut and Nathan Kallus. Demystifying inference after adaptive experiments. *Annual Review of Statistics and its Application*, 12(1):407–423, 2025.

[3] Ethan Che, Daniel R Jiang, Hongseok Namkoong, and Jimmy Wang. Optimization-driven adaptive experimentation. *arXiv preprint arXiv:2408.04570*, 2024.

[4] Thomas Cook, Alan Mishler, and Aaditya Ramdas. Semiparametric efficient inference in adaptive experiments. In *Causal Learning and Reasoning*, pages 1033–1064. PMLR, 2024.

[5] Alex Deng, Jiannan Lu, and Shouyuan Chen. Continuous monitoring of a/b tests without pain: Optional stopping in bayesian testing. In *2016 IEEE International conference on data science and advanced analytics (DSAA)*, pages 243–252. IEEE, 2016.

[6] Vitor Hadad, David A Hirshberg, Ruohan Zhan, Stefan Wager, and Susan Athey. Confidence intervals for policy evaluation in adaptive experiments. *Proceedings of the national academy of sciences*, 118(15):e2014602118, 2021.

[7] Keisuke Hirano and Jack R Porter. Asymptotic representations for sequential decisions, adaptive experiments, and batched bandits. *arXiv preprint arXiv:2302.03117*, 2023.

[8] Kwang-Sung Jun, Kevin Jamieson, Robert Nowak, and Xiaojin Zhu. Top arm identification in multi-armed bandits with batch arm pulls. In *Artificial Intelligence and Statistics*, pages 139–148. PMLR, 2016.

[9] Ron Kohavi, Diane Tang, and Ya Xu. *Trustworthy online controlled experiments: A practical guide to a/b testing*. Cambridge University Press, 2020.

[10] Tor Lattimore and Csaba Szepesvári. *Bandit algorithms*. Cambridge University Press, 2020.

[11] Xinkun Nie, Xiaoying Tian, Jonathan Taylor, and James Zou. Why adaptively collected data have negative bias and how to correct for it. In *International Conference on Artificial Intelligence and Statistics*, pages 1261–1269. PMLR, 2018.

[12] Kelly Zhang, Lucas Janson, and Susan Murphy. Inference for batched bandits. *Advances in neural information processing systems*, 33:9818–9829, 2020.

Appendix

A. Asymptotic Normality of Cohort Averages

Proposition 1 Define

$$\Sigma^t_{a,i} = \pi_{a,i}(\sigma^t_{a,i})^2 \mathbb{I}_{t-i+1} + \pi_{a,i}(1-\pi_{a,i})(\mu_a)^2 \mathbf{1}_{t-i+1}\mathbf{1}'_{t-i+1}.$$

Then under Assumptions 1-3, the asymptotic distribution of $\bar{\mathbb{X}}^t_{a,i} \mid \bar{\mathbb{X}}^{i-1}$ follows

$$\sqrt{N_i}(\Sigma^t_{a,i})^{-\frac{1}{2}}(\bar{\mathbb{X}}^t_{a,i} - \pi_{a,i}\mu_a \mathbf{1}_{t-i+1}) \mid \pi_i(\bar{\mathbb{X}}^{i-1}) \to_D \mathcal{N}(0, \mathbf{I}).$$

B. Linear Model Representation of Adaptive Histories

Proposition 2 Denote $X^t := ((\bar{\mathbb{X}}^t_{0,1})', (\bar{\mathbb{X}}^t_{1,1})', \ldots, (\bar{\mathbb{X}}^t_{0,t})', (\bar{\mathbb{X}}^t_{1,t})')'$ as a 1-dim flattened vector of the full history and design matrix $\mathbf{A}^t := ((\mathbf{A}^t_{0,1})', (\mathbf{A}^t_{1,1})', \cdots, (\mathbf{A}^t_{0,t})', (\mathbf{A}^t_{1,t})')$ with $\mathbf{A}^t_{a,i}$ of size $(t-i+1)\times 2$ and

$$\mathbf{A}^t_{a,i}[p,q] = \begin{cases} \pi_{1,i} & \text{if } a=1 \text{ and all } p,q \\ \pi_{0,i} & \text{if } a=0 \text{ and } q=0 \text{ and all } p \\ 0 & \text{otherwise} \end{cases}.$$

In addition, define

$$\Sigma_\epsilon := \mathbf{diag}\left(\frac{\Sigma^t_{0,1}}{N_1}, \frac{\Sigma^t_{1,1}}{N_1}, \ldots, \frac{\Sigma^t_{0,t}}{N_t}, \frac{\Sigma^t_{1,t}}{N_t}\right).$$

Then under Assumptions 1-3, the likelihood of X^t is asymptotically equivalent to the following linear model:

$$X^t = \mathbf{A}^t\tilde{\mu} + \epsilon, \quad \epsilon \sim \mathcal{N}(0, \Sigma_\epsilon).$$

C. Experiment Design Guideline

In practice, we suggest practitioners to follow the steps before performing a Bayesian batch adaptive experiment.

1) Perform estimation on past experiments to determine parameters in the prior distributions including mean, variance of the control arm, and variance of the effect size, and PriorOdds_{10}.
2) Select the desired FDR rates (and its corresponding BF) and FNDR rates.
3) Estimate the traffic of the experiment of the most frequent practical updates with a minimal sample size requirement (e.g., if hourly updates are practice, then select hourly over daily updates and estimates the hourly traffic.)
4) Perform FNDR simulation to determine the number of batches required for the desired FNDR.

D. Non-stationary Scenario: age-specific expect effect

Consider the scenario that the expected effect of each arm is non-stationary in its mean across time t but rather changes with respect to the age of the cohort $|t-i|$, i.e., $\mu(a,i,t) = \mu(a,|t-i|)$. Additionally, the treatment effects between arms are assumed to be constant with respect to age, that is the treatment effect is stationary across time. This data model is more realistic as it allows the seasonality and maturity of the metric into consideration.

Similar to the stationary case, we perform the same reparametrization that is $\tilde{\mu} = (\mu(0,0)\cdots\mu(0,t-1)\delta(1))'$ and the corresponding design matrix $\mathbf{A}$ is defined as follows:

$$\mathbf{A}^t = \begin{bmatrix} (\mathbf{A}^t_{0,1})' & (\mathbf{A}^t_{1,1})' & \cdots & (\mathbf{A}^t_{0,t})' & (\mathbf{A}^t_{1,t})' \end{bmatrix}',$$

where $\mathbf{A}^t_{a,i}$ is $(t-i+1)\times(t+1)$ matrix with

$$\mathbf{A}^t_{a,i}[j,k] = \begin{cases} \pi_{1,i} & \text{if } a=1 \text{ and } (j=k \text{ or } k=t) \\ \pi_{0,i} & \text{if } a=0 \text{ and } j=k \\ 0 & \text{otherwise.} \end{cases}$$

The data covariance Σ_ϵ is of the same block-diagonal structure (under same arm independence and individual effect independence assumptions as in Proposition 2) with

$$\Sigma^t_{a,i} = \pi_{a,i}(\sigma^t_{a,i})^2 \mathbb{I}_{t-i+1} + \pi_{a,i}(1-\pi_{a,i})\mu_{a,i,t}\mu'_{a,i,t},$$

$$\mu_{a,i,t} := \begin{pmatrix} \mu(a,0) \\ \mu(a,1) \\ \vdots \\ \mu(a,|t-i|) \end{pmatrix}.$$

E. Opportunity and FNDR vs batch sample size

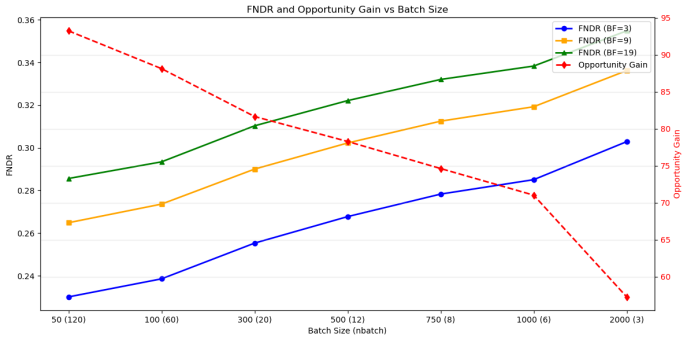


Fig. 4: FNDR (solid lines) and opportunity gain (dashed line) against different number of batch size under the stationary scenario with $\sigma^2_\delta = 0.1$ and a total sample size of 6,000.