Nonparametric Extensions of Randomized Response for Private Confidence Sets

Ian Waudby-Smith [1] **Zhiwei Steven Wu** [1] **Aaditya Ramdas** [1]

Abstract

This work derives methods for performing nonparametric, nonasymptotic statistical inference for population means under the constraint of local differential privacy (LDP). Given bounded observations $(X_1, \dots, X_n)$ with mean $\mu^\star$ that are privatized into $(Z_1, \dots, Z_n)$, we present confidence intervals (CI) and time-uniform confidence sequences (CS) for $\mu^\star$ when only given access to the privatized data. To achieve this, we introduce a nonparametric and sequentially interactive generalization of Warner's famous "randomized response" mechanism, satisfying LDP for arbitrary bounded random variables, and then provide CIs and CSs for their means given access to the resulting privatized observations. For example, our results yield private analogues of Hoeffding's inequality in both fixed-time and time-uniform regimes. We extend these Hoeffding-type CSs to capture time-varying (non-stationary) means, and conclude by illustrating how these methods can be used to conduct private online A/B tests.

1. Introduction

It is easier than ever for mobile apps and web browsers to collect massive amounts of sensitive data about individuals. *Differential privacy* (DP) provides a framework that leverages statistical noise to limit the risk of sensitive information disclosure (Dwork et al., 2006). The goal of private data analysis is to extract meaningful population-level information from the data (whether in the form of machine learning model training, statistical inference, etc.) while preserving the privacy of individuals via DP. In particular, this paper will focus on statistical inference (e.g. confidence intervals and p-values) for population means under DP constraints.

As motivating examples, suppose a city wishes to survey households to calculate the approval rating of their mayor, or an IT company aims to understand whether a redesigned homepage will lead to the average user spending more time on it. Both problems can be framed as estimating the mean of some (potentially large) population, but it may be infeasible to query every single household or all possible website users. Fortunately, a *sample mean* can still be used to estimate the population mean with some degree of precision. For example, a city may randomly choose households to query, or the technology company may show 10% of users the redesigned webpage at random. This is often referred to as "A/B testing", and we expand on this application under privacy constraints in Section 4. When making decisions, however, it is crucial to both calculate sample means *and* quantify the uncertainty in those estimates (e.g. using confidence intervals, reviewed in Section 1.2). However, calculating confidence intervals under local differential privacy constraints (defined in Section 1.1) poses a unique statistical challenge, because these intervals must incorporate both the uncertainty introduced from random sampling *and* from the privacy mechanism. This paper studies and provides a nonparametric solution to precisely this challenge.

1.1. Background I: Local Differential Privacy

There are two main models of privacy within the DP framework: *central* and *local* DP (LDP) (Dwork et al., 2006; Kasiviswanathan et al., 2011; Dwork & Roth, 2014). The former involves a centralized data aggregator that is trusted with constructing privatized output from raw data, while the latter performs privatization at the "local" or "personal" level (e.g. on an individual's smartphone before leaving the device) so that trust need not be placed in any data collector. Both models have their advantages and disadvantages: LDP is a more restrictive model of privacy and thus in general requires more noise to be added. On the other hand, the stronger privacy guarantees that do not require a trusted central aggregator make LDP an attractive framework in practice. This paper deals exclusively with LDP.

Making our setup more precise, suppose $X_1, X_2, \dots$ is a (potentially infinite) sequence of $[0, 1]$-valued random variables. We could instead have assumed boundedness on any known interval $[a, b]$ since we can always translate and scale the interval to $[0, 1]$ via the transformation $x \mapsto (x - a)/(b - a)$. We will refer to $(X_t)_{t=1}^\infty$ as the "raw" or "sensitive" data that are yet to be privatized. Fol-

[1]Carnegie Mellon University. Correspondence to: Ian Waudby-Smith <ianws@cmu.edu>.

Proceedings of the 40^{th} International Conference on Machine Learning, Honolulu, Hawaii, USA. PMLR 202, 2023.

lowing the notation of Duchi et al. (2013a) the privatized views $Z_1, Z_2, \ldots$ of $X_1, X_2, \ldots$, respectively are generated by a sequence of conditional distributions $Q_1, Q_2, \ldots$ which we refer to as the *privacy mechanism*. Throughout this paper, we will allow this privacy mechanism to be *sequentially interactive*, meaning that the distribution Q_i of Z_i may depend on the past privatized observations $Z_1^{i-1} := (Z_1, \ldots, Z_{i-1})$ (Duchi et al., 2013a). In other words, the privatized view Z_i of X_i has a conditional distribution $Q_i(\cdot \mid X_i = x, Z_1^{i-1} = z_1^{i-1})$. Following Duchi et al. (2013a; 2018) we say that Q_i satisfies ε-local differential privacy if for all $z_1, \ldots, z_{i-1} \in [0, 1]$ and $x, \widetilde{x} \in [0, 1]$, the following likelihood ratio is uniformly bounded:

$$\sup_{z \in [0,1]} \frac{q_i(z \mid X_i = x, Z_1^{i-1} = z_1^{i-1})}{q_i(z \mid X_i = \widetilde{x}, Z_1^{i-1} = z_1^{i-1})} \leqslant \exp\{\varepsilon\}, \qquad (1)$$

where q_i is the density (or Radon-Nikodym derivative) of Q_i with respect to some dominating measure. In the non-interactive case where the dependence on Z_1^{i-1} is dropped, (1) simplifies to the usual ε-LDP definition (Dwork & Roth, 2014). To put $\varepsilon > 0$ in a real-world context, Apple uses privacy levels in the range of $\varepsilon \in \{2, 4, 8\}$ on macOS and iOS devices for various ε-LDP data collection tasks, including health data type usage, emoji suggestions, and lookup hints (Apple Inc., 2022). See Figure 1 to intuit how ε affects the widths of confidence intervals that we develop. Next, we review time-uniform confidence sequences and how they differ from fixed-time confidence intervals.

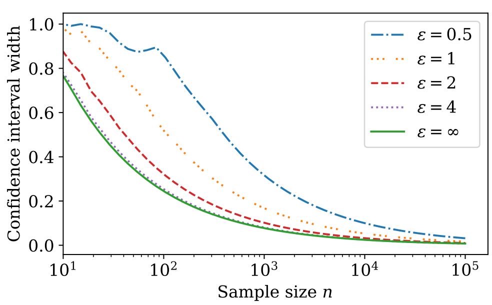

Figure 1. Widths of private 90%-CIs for the mean of a uniform distribution using our private Hoeffding CI given in (11) for various levels of ε ranging from $\varepsilon = 0.5$ (very private) to $\varepsilon = \infty$ (no privacy). Unsurprisingly, less privacy leads to sharper inference, but notice that inference is still practical, especially for $\varepsilon \geqslant 2$. For context, Apple uses $\varepsilon \in \{2, 4, 8\}$ for various data collection tasks on iPhones (Apple Inc., 2022). At these levels of privacy, our CIs perform nearly as well as — and are in some cases indistinguishable from — the non-private Hoeffding CI.

1.2. Background II: Confidence Sequences

One of the most fundamental tasks in statistical inference is the derivation of confidence intervals (CI) for a parameter of interest $\mu^\star \in \mathbb{R}$ (e.g. mean, variance, treatment effect, etc.). Given data $X_1, \ldots, X_n$, the interval $\dot{C}_n \equiv C(X_1, \ldots, X_n)$ is said to be a $(1-\alpha)$-CI for $\mu^\star$ if

$$\mathbb{P}(\mu^\star \notin \dot{C}_n) \leqslant \alpha, \qquad (2)$$

where $\alpha \in (0, 1)$ is a prespecified error tolerance. Notice that (2) is a "pointwise" or "fixed-time" statement, meaning that it only holds for a single fixed sample size n.

The "time-uniform" analogue of CIs are so-called *confidence sequences* (CS) — sequences of confidence intervals that are uniformly valid over a (potentially infinite) time horizon (Darling & Robbins, 1967; Robbins, 1970). We say that the sequence $(\bar{C}_t)_{t=1}^{\infty}$ is a $(1-\alpha)$-CS[1] for $\mu^\star$ if

$$\mathbb{P}(\exists t \geqslant 1 : \mu^\star \notin \bar{C}_t) \leqslant \alpha. \qquad (3)$$

The guarantee (3) has important implications for data analysis, giving practitioners the ability to (a) update inferences as new data become available, (b) continuously monitor studies without any statistical penalties for "peeking", and (c) make decisions based on valid inferences at arbitrary stopping times: for any stopping time τ, $\mathbb{P}(\mu^\star \notin \bar{C}_\tau) \leqslant \alpha$.

1.3. Contributions and Outline

Our primary contributions are threefold: (a) a new privacy mechanism, (b) CIs, and (c) time-uniform CSs.

(a) We introduce `NPRR` — a sequentially interactive, nonparametric generalization of Warner's randomized response (Warner, 1965) for bounded data (Section 2).

(b) We derive several CIs for the mean of bounded random variables that are privatized by `NPRR` (Section 3). We believe Section 3 introduces the first private nonparametric CIs for means of bounded random variables.

(c) We derive time-uniform CSs for the mean of bounded random variables that are privatized by `NPRR`, enabling private nonparametric sequential inference (Section 3.3). We also introduce a CS that is able to capture means that change over time under no stationarity conditions on the time-varying means (Section 3.4). We believe Sections 3.3 and 3.4 are the first private nonparametric CSs in the DP literature.

Furthermore, we show how all of the aforementioned techniques can be used to conduct private online A/B tests (Section 4). Finally, Section 5 summarizes our findings and discusses some additional results whose details can be found in the appendix. A Python package implementing our methods as well as code to reproduce the figures can be found on GitHub at github.com/WannabeSmith/nprr.

[1] As a mnemonic, we will use overhead bars $\bar{C}_t$ and dots $\dot{C}_n$ for time-uniform CSs and fixed-time CIs, respectively.

1.4. Related Work

The literature on differentially private statistical inference is rich, including nonparametric estimation rates (Wasserman & Zhou, 2010; Duchi et al., 2013a;b; 2018; Kamath et al., 2020; Butucea et al., 2020; Acharya et al., 2021b), parametric hypothesis testing and confidence intervals (Vu & Slavkovic, 2009; Wang et al., 2015; Gaboardi et al., 2016; Awan & Slavković, 2018; Karwa & Vadhan, 2018; Canonne et al., 2019; Joseph et al., 2019; Ferrando et al., 2022; Covington et al., 2021), median estimation (Drechsler et al., 2021), independence testing (Couch et al., 2019), online convex optimization (Jun & Orabona, 2019), and parametric sequential hypothesis testing (Wang et al., 2022). A more detailed summary can be found in Section D.

The aforementioned works do not study the problem of private nonparametric confidence sets for population means. Prior work does exist on confidence intervals for the *sample mean of the data* (Ding et al., 2017; Wang et al., 2019). The most closely related work is that of Ding et al. (2017, Section 2.1) who introduce the "1BitMean" mechanism which can be viewed as a special case of `NPRR` (Algorithm 2). They derive a private Hoeffding-type confidence interval for the *sample* mean of the data, but it is important to distinguish this from the more classical statistical task of *population* mean estimation. For example, if $X_1, \dots, X_n$ are random variables drawn from a distribution with mean $\mu^\star$, then the *population mean* is $\mu^\star$, while the *sample mean* is $\widehat{\mu}_n := \frac{1}{n}\sum_{i=1}^n X_i$. A private CI for $\mu^\star$ incorporates randomness from both the mechanism *and* the data, while a CI for $\widehat{\mu}_n$ incorporates randomness from the mechanism *only*. Neither is a special case of the other, and some of our techniques allow for the (sequential) estimation of sample means (see Appendix B.2 for details and explicit bounds) but this paper is primarily focused on the problem of private *population* mean estimation.

2. Extending Warner's Randomized Response

Before introducing our nonparametric extension of randomized response, let us briefly review Warner's classical randomized response mechanism as well as the Laplace mechanism, discuss their shortcomings, and introduce a new mechanism to remedy them.

Warner's randomized response. When the raw data $(X_t)_{t=1}^\infty$ are binary, one of the oldest and simplest privacy mechanisms is Warner's *randomized response* (RR) (Warner, 1965). Warner's RR was introduced decades before the very definition of DP, but was later shown to satisfy LDP by Dwork & Roth (2014). RR was introduced as a method to provide plausible deniability to subjects when answering sensitive survey questions (Warner, 1965), and proceeds as follows: when presented with a sensitive Yes/No question (e.g. "have you ever used illicit drugs?"), the subject flips a biased coin with $\mathbb{P}(\text{Heads}) = r \in (0, 1]$. If the coin comes up heads, then the subject answers truthfully; if tails, the subject answers "Yes" or "No" (encoded as 1 and 0, respectively) with equal probability 1/2. It is easy to see that this mechanism satisfies ε-LDP with $\varepsilon = \log(1 + \frac{2r}{1-r})$ by bounding the likelihood ratio of the privatized response distributions: for any true response $x \in \{0, 1\}$, let $q(z \mid X = x) = r\mathbb{1}(z = x) + (1 - r)/2$ denote the conditional probability mass function of its privatized view. Then for any $x, \widetilde{x} \in \{0, 1\}$,

$$\sup_{z\in\{0,1\}} \frac{q(z \mid X = x)}{q(z \mid X = \widetilde{x})} \leqslant 1 + \frac{2r}{1-r}, \tag{4}$$

and hence RR satisfies ε-LDP with $\varepsilon = \log(1 + \frac{2r}{1-r})$. In Appendix B.1, we show how one can derive a CI for the mean of Bernoulli random variables when they are privatized via RR, but as we will see in Section 3, this will be an immediate corollary of a more general result for bounded random variables (Theorem 4).

One downside of RR, however, is that it takes binary data as input. On the other hand, the famous Laplace mechanism satisfies ε-LDP for *bounded* data, including binary ones.

The Laplace mechanism. The Laplace mechanism appeared in the very same paper that introduced DP (Dwork et al., 2006). Algorithm 1 recalls the (sequentially interactive) Laplace mechanism (Duchi et al., 2013a). It is well-known that Z_t is (conditionally) ε_t-LDP (given Z_1^{t-1}) for each t (Dwork et al., 2006). Appendix B.4 derives novel CIs and CSs for population means under the Laplace mechanism, but we omit them here for brevity as our new mechanism (to be introduced shortly) will yield better bounds.

Algorithm 1 Sequentially interactive Laplace mechanism

```
for t = 1,2,... do
    Choose ε_t based on Z_1^{t-1}.
    Generate L_t ~ Laplace(1/ε_t)
    Z_t ← X_t + L_t
end for
```

Nonparametric randomized response (`NPRR`). Our mechanism "Nonparametric randomized response" (`NPRR`) serves as a sequentially interactive generalization of RR for arbitrary bounded data by combining stochastic rounding (Barnes et al., 1951; Forsythe, 1959; Hull & Swenson, 1966) with k-RR — a categorical but non-interactive generalization of Warner's RR introduced by Kairouz et al. (2014; 2016), and also considered by Li et al. (2020) under the name "Generalized Randomized Response". Note that Kairouz et al. (2014; 2016) use k to refer to the number of unique values that the input and output data can take on, which is $k = G + 1$ in the case of Algorithm 2. `NPRR` is explicitly described in Algorithm 2, and its LDP guarantees are summarized in Theorem 1.

Algorithm 2 Nonparametric randomized response (NPRR)

for $t = 1, 2, \dots$ **do**
  // Step 1: Discretize X_t into Y_t via stochastic rounding.
  Choose integer $G_t \geqslant 1$ based on Z_1^{t-1}
  $X_t^{\text{ceil}} \leftarrow \lceil G_t X_t \rceil / G_t$, $X_t^{\text{floor}} \leftarrow \lfloor G_t X_t \rfloor / G_t$
  if $X_t^{\text{ceil}} == X_t^{\text{floor}}$ **then**
    $Y_t \leftarrow X_t$
  else
    Generate $Y_t \sim \begin{cases} X_t^{\text{ceil}} & \text{w.p. } G_t \cdot (X_t - X_t^{\text{floor}}) \\ X_t^{\text{floor}} & \text{w.p. } G_t \cdot (X_t^{\text{ceil}} - X_t) \end{cases}$
  end if
  // Step 2: Privatize Y_t into Z_t via k-RR.
  Choose $r_t \in (0, 1]$ based on Z_1^{t-1}
  Generate $\mathcal{U}_t \sim \text{Unif}\left\{0, \frac{1}{G_t}, \frac{2}{G_t}, \dots, \frac{G_t}{G_t}\right\}$
  Generate $Z_t \sim \begin{cases} Y_t & \text{w.p. } r_t \\ \mathcal{U}_t & \text{w.p. } 1 - r_t \end{cases}$
end for

Notice that if $(X_t)_{t=1}^{\infty}$ are $\{0, 1\}$-valued, and if we set $G_1 = G_2 = \dots = 1$ and $r_1 = r_2 = \dots = r \in (0, 1]$, then no stochastic rounding occurs and NPRR recovers RR exactly, making NPRR a sequentially interactive and nonparametric generalization for bounded data. Also, if we let NPRR be non-interactive and set $G_1 = \dots = G_n = 1$, then NPRR recovers the "1BitMean" mechanism (Ding et al., 2017). However, Ding et al. (2017) do not explicitly point out the connection to stochastic rounding. Notice that Ding et al. (2017)'s α-point rounding mechanism is different from NPRR as NPRR shifts the mean of the inputs but alpha-point rounding leaves the mean unchanged. Let us now formalize NPRR's LDP guarantees.

Theorem 1 (NPRR satisfies LDP)**.** *Suppose $(Z_t)_{t=1}^{\infty}$ are generated according to NPRR. Then for each $t \in \{1, 2, \dots\}$, Z_t is a conditionally ε_t-LDP view of X_t with*

$$\varepsilon_t := \log\left(1 + \frac{(G_t + 1) r_t}{1 - r_t}\right). \tag{5}$$

The proof in Section A.2 proceeds by bounding the conditional likelihood ratio for any two data points $x, \widetilde{x} \in [0, 1]$ similar to (1). In all of the results that follow in the following sections, we will write expressions in terms of $(r_t)_{t=1}^{n}$, but these can always be chosen given desired $(\varepsilon_t)_{t=1}^{n}$ levels via the relationship

$$r_t = \frac{\exp\{\varepsilon_t\} - 1}{\exp\{\varepsilon_t\} + G_t}. \tag{6}$$

In the familiar special case of $r_t = r \in (0, 1]$ and $G_t = G \in \{1, 2, \dots\}$ for each t, we have that $(Z_t)_{t=1}^{\infty}$ satisfy ε-LDP with $\varepsilon := \log(1 + (G + 1)r/(1 - r))$. Notice that when $G_t = 1$ for each t, we have that NPRR satisfies ε-LDP with the same value of ε as Warner's RR. Consequently, there is no privacy lost from instantiating the more general NPRR to the binary case.

Remark 2 (Who chooses ε_t, r_t, or G_t, and how?). Due to the sequential interactivity of NPRR, individuals can specify their own levels of privacy, or the parameters $(r_t, G_t)_{t=1}^{\infty}$ can be adjusted over time (e.g. if the data collector chooses to decrease ε_t for regulatory reasons, or increase ε_t to obtain sharper inference). Formally, (r_t, G_t) can be chosen in any way as long as they are *predictable*, meaning that they can depend on Z_1^{t-1}. Nevertheless, sequential interactivity is completely optional, and the data collector is free to set $(r_t, G_t) = (r, G)$ for every t to recover the familiar notion of ε-LDP.

Why introduce NPRR as an alternative? While RR is limited to privatizing binary data, the Laplace mechanism can handle bounded data, so why introduce NPRR as an alternative to the two? The reason stems from our original motivation: to derive locally private nonparametric, nonasymptotic confidence sets for means of bounded random variables. To achieve this, we will ultimately use modern concentration techniques from the literature on (non-private) confidence sets, many of which exploit boundedness in clever ways to yield clean, closed-form expressions and/or empirically tight confidence intervals. Since the Laplace mechanism does not preserve the boundedness of its input, it is not clear how those techniques can be used for Laplace-privatized data (though we do derive novel Laplace-based solutions using a different approach in Appendix B.4, but they are ultimately outperformed by those that we derive based on NPRR). NPRR on the other hand, preserves the input's boundedness, making it possible to apply analogues of these modern concentration techniques for NPRR-privatized data. The efficiency gains that result from this approach are illustrated in Figures 4 and 5.

In addition to being useful for deriving simple and efficient confidence sets, NPRR has some other orthogonal advantages over the Laplace mechanism. First, NPRR has reduced storage requirements: Once a $[0, 1]$-bounded random variable has been privatized via Laplace, the output is a floating-point number, requiring 64 bits to store as a double-precision float. In contrast, NPRR outputs one of $(G + 1)$ different values, hence requiring only $\lceil \log_2(G + 1) \rceil$ bits to store. Moreover, storing the NPRR-privatized view of x will never require more memory than storing x itself (unless G is set to nonsensical values larger than 2^{64}), while Laplace-privatized views will always require at least enough memory to represent floating point numbers.

Second, NPRR is automatically resistant to the floating-point attacks that the Laplace mechanism suffers from. Mironov

(2012) showed that storing Laplace output as a floating-point number can leak information about the input x, thereby compromising its LDP guarantees. While Mironov (2012) discusses remedies to this issue, practitioners may still naively apply the Laplace mechanism using common software packages and remain vulnerable to these so-called "floating-point attacks". In contrast, the discrete representation of NPRR's output is not vulnerable to such attacks, without the need for remedies at all. Note that while NPRR may have to deal with floating point numbers as input, they are transformed into discrete random variables *before* any ε-LDP guarantees are added. The privatization step (transforming Y_t into Z_t in Algorithm 2) takes one of $G_t + 1$ values as input and produces one of $G_t + 1$ values as output, thereby sidestepping any need to handle floating point numbers.

The remainder of this paper will focus solely on constructing efficient locally private confidence sets, but the above benefits can be seen as "free" byproducts of NPRR's design.

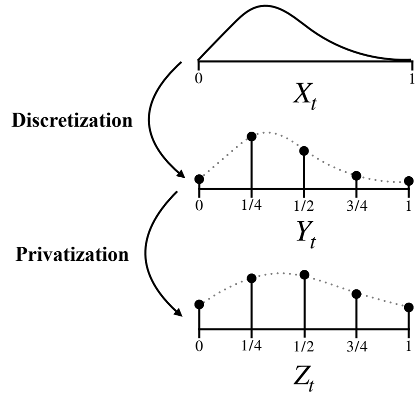

Figure 2. An illustration of how a distribution $Q \in \mathcal{Q}^n_{\mu^\star}$ can arise from applying NPRR with $G_t = 4$ to draws from the input distribution $P \in \mathcal{P}^n_{\mu^\star}$. Raw data X_t are discretized into Y_t so that Y_t has finite support but so that $\mu^\star = \mathbb{E}(X_t) = \mathbb{E}(Y_t)$. The discrete Y_t are then privatized into Z_t with conditional mean $\mathbb{E}(Z_t \mid Z_1^{t-1}) = \zeta_t(\mu^\star) = r_t\mu^\star + (1 - r_t)/2$ by being mixed with independent uniform noise $\mathcal{U}_t \sim \text{Unif}\{0, 1/4, 1/2, 3/4, 1\}$.

3. Private CIs for Bounded Data

Making matters formal, let $\mathcal{P}_\mu$ be the set of distributions on $[0, 1]$ with population mean $\mu \in [0, 1]$. $\mathcal{P}_\mu$ is a convex set of distributions with no common dominating measure, since it consists of discrete and continuous distributions, as well as their mixtures. We will consider sequences of random variables $(X_i)_{i=1}^n$ drawn from the product distribution $\prod_{i=1}^n P_i$ where $n \in \{1, 2, \dots, \infty\}$ and each $P_i \in \mathcal{P}_\mu$. For succinctness, define the following set of distributions,

$$\mathcal{P}^n_\mu := \left\{ \prod_{i=1}^n P_i \text{ such that each } P_i \in \mathcal{P}_\mu \right\}, \tag{7}$$

for $n \in \{1, 2, \dots, \infty\}$. In words, $\mathcal{P}^n_\mu$ contains distributions for which the random variables are independent and $[0, 1]$-bounded with mean μ but need not be identically distributed. We use the notation $(X_t)_{t=1}^n \sim P$ for some $P \in \mathcal{P}^n_{\mu^\star}$ to indicate that $(X_t)_{t=1}^n$ are independent with mean $\mu^\star$. The goal is now to derive sharp CIs and time-uniform CSs for $\mu^\star$ given NPRR-privatized views of $(X_t)_{t=1}^n$.

Let us write $\mathcal{Q}^n_{\mu^\star}$ to denote the set of joint distributions on NPRR's output, where we have left the dependence on each G_t and r_t implicit. In other words, given $(X_t)_{t=1}^n \sim P$ for some $P \in \mathcal{P}^n_{\mu^\star}$, their NPRR-induced privatized views $(Z_t)_{t=1}^n$ have a joint distribution from Q for some $Q \in \mathcal{Q}^n_{\mu^\star}$.

3.1. What is a Locally Private Confidence Set?

Let first define what we mean by locally private confidence intervals (LPCI) and sequences (LPCS), and subsequently derive them for means of bounded random variables.

Definition 3 (Locally private confidence sets)**.** Let $\varepsilon \equiv (\varepsilon_t)_{t=1}^n \equiv (\varepsilon_t)_t$. We say that L_n is a lower $(1 - \alpha, \varepsilon)$-LPCI for a parameter $\theta^\star$, and with respect to the raw data $(X_t)_{t=1}^n$ if L_n is a lower $(1 - \alpha)$-CI for $\theta^\star$, meaning

$$\mathbb{P}(\theta^\star \geqslant L_n) \geqslant 1 - \alpha, \tag{8}$$

and if $L_n \equiv L(Z_1, \dots, Z_n)$ is only a function of the ε_t-LDP view Z_t of X_t for each t, but not of $(X_t)_{t=1}^n$ directly.

Similarly, we say that $(L_t)_{t=1}^\infty$ is a lower $(1 - \alpha, \varepsilon)$-LPCS for $\theta^\star$ if (8) is replaced with the time-uniform guarantee

$$\mathbb{P}(\forall t,\ \theta^\star \geqslant L_t) \geqslant 1 - \alpha. \tag{9}$$

Upper CIs and CSs are defined analogously.

Note that LPCIs and LPCSs also satisfy ε-LDP, since DP is closed under post-processing (Dwork & Roth, 2014).

3.2. A Locally Private Hoeffding CI via NPRR

First, we present a private generalization of Hoeffding's inequality under NPRR.

Theorem 4 (NPRR-H)**.** *Suppose $(X_t)_{t=1}^n \sim P$ for some $P \in \mathcal{P}^n_{\mu^\star}$, and let $(Z_t)_{t=1}^n \sim Q \in \mathcal{Q}^n_{\mu^\star}$ be their privatized views via NPRR. Define the NPRR-adjusted sample mean*

$$\widehat{\mu}_n := \frac{\frac{1}{n}\sum_{i=1}^n (Z_i - (1 - r_i)/2)}{\frac{1}{n}\sum_{i=1}^n r_i}. \tag{10}$$

Then,

$$\dot{L}^{\mathrm{H}}_n := \widehat{\mu}_n - \sqrt{\frac{\log(1/\alpha)}{2n(\frac{1}{n}\sum_{i=1}^n r_i)^2}} \tag{11}$$

is a lower $(1 - \alpha, (\varepsilon_t)_t)$-LPCI for $\mu^\star$.

The proof in Section A.3 uses a locally private supermartingale variant of the Cramér-Chernoff bound. We recommend setting r_t for the desired ε_t-LDP level via the relationship in (6) and $G_t := 1$ for all t (the reason behind which we will discuss in Remark 7). Notice that in the non-private setting where we set $r_i = 1$ for all i, then $\dot{L}_n^{\mathrm{H}}$ recovers the classical Hoeffding inequality exactly (Hoeffding, 1963). Moreover, notice that if $(X_t)_{t=1}^n$ took values in $[a, b]$ instead of $[0, 1]$, then (11) would simply scale with $(b - a)$ in the same manner as Hoeffding (1963). Recall as discussed in Remark 2 that $(r_t)_{t=1}^n$ could be chosen either by the data collector or by the subject whose data are being collected, but that sequential interactivity is optional.

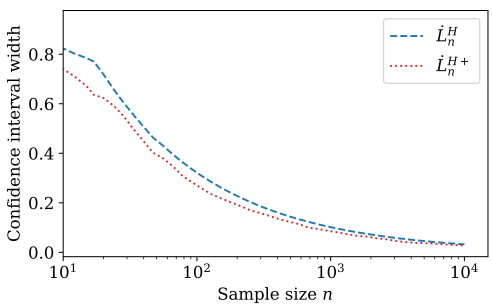

Figure 3. Two $(90\%, 2)$-LPCIs: $\dot{L}_n^H$ given in (11) and $\dot{L}_n^{H+}$ given in (12) — i.e. these are $(1-\alpha, \varepsilon)$-LPCIs with $\alpha = 0.1$ and $\varepsilon = 2$. Notice that the latter can be tighter than the former. Indeed this is because L_n^{H+} is never looser than L_n^H (by definition) but strictly tighter with positive probability.

In fact, we can strictly improve on (11) by exploiting the martingale dependence of this problem. Indeed, under the same assumptions as Theorem 4, we have that

$$\dot{L}_n^{\mathrm{H}+} := \max_{1 \leqslant t \leqslant n} \left\{ \widehat{\mu}_t - \frac{\log(1/\alpha) + t\lambda_n^2/8}{\lambda_n \sum_{i=1}^t r_i} \right\} \tag{12}$$

is also a lower $(1-\alpha, (\varepsilon_t)_{t=1}^n)$-LPCI for $\mu^\star$, where $\lambda_n := \sqrt{8\log(1/\alpha)/n}$. Notice that $\dot{L}_n^{\mathrm{H}+}$ is at least as tight as $\dot{L}_n^{\mathrm{H}}$ since the n^{th} term in the above $\max_{1 \leqslant t \leqslant n}$ recovers $\dot{L}_n^{\mathrm{H}}$ exactly. Moreover, $\dot{L}_n^{\mathrm{H}+}$ is *strictly* tighter than $\dot{L}_n^{\mathrm{H}}$ with positive probability, and hence strictly tighter in expectation: $\mathbb{E}(\dot{L}_n^{\mathrm{H}+}) > \mathbb{E}(\dot{L}_n^{\mathrm{H}})$.

Remark 5 (Minimax rate optimality of (11)). In the case of $\varepsilon_1 = \cdots = \varepsilon_n = \varepsilon \in (0, 1]$, Duchi et al. (2013a, Proposition 1) give minimax estimation rates for the problem of nonparametric mean estimation. Their lower bounds say that for any ε-LDP mechanism and estimator $\widehat{\mu}_n$ for $\mu^\star$, the root mean squared error $\sqrt{\mathbb{E}(\widehat{\mu}_n - \mu^\star)^2}$ cannot scale faster than $O(1/\sqrt{n\varepsilon^2})$. Since NPRR is ε-LDP with $\varepsilon = \log(1 + 2r/(1-r))$, we have that $r \asymp \varepsilon$ up to constants on $\varepsilon \in (0, 1]$. It follows that $\dot{L}_n^{\mathrm{H}} \asymp 1/\sqrt{n\varepsilon^2}$, matching the minimax estimation rate. Of course, the midpoint of a CI for $\mu^\star$ can always be used as an estimator for $\mu^\star$, and hence we cannot expect the width of the CI to shrink faster than the minimax estimation rate. While explicit minimax lower bounds do not exist for the setting where $\varepsilon_i \neq \varepsilon_j$ for some i, j, notice that instead of scaling with r^{-1} (which we would have if $r_i = r_j$ for $i \neq j$), $\dot{L}_n^{\mathrm{H}}$ scales with $(\frac{1}{n}\sum_{i=1}^n r_i)^{-1}$, and hence our bounds seem to be of the right order when ε is permitted to change.

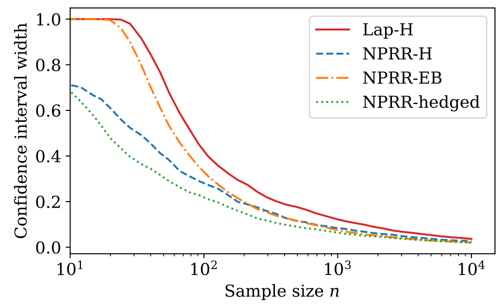

Figure 4. Widths of $(90\%, 2)$-LPCIs for the mean of a Beta(50, 50) distribution. Hoeffding-based methods (Lap-H and NPRR-H found in Corollary 4 and Theorem 4) do slightly worse than the variance-adaptive ones (NPRR-EB and NPRR-hedged in Proposition 2 and Theorem 10), but in all cases, CIs that rely on NPRR seem to outperform Lap-H in both small and large n regimes.

Remark 6 (The relationship between ε and (11) for practical levels of privacy). As mentioned in the introduction and in Figure 1, Apple uses values of $\varepsilon \in \{2, 4, 8\}$ for various ε-LDP data collection tasks on iPhones (Apple Inc., 2022). Note that for $G = 1$, having ε take values of 2, 4, and 8 corresponds to r being roughly 0.762, 0.964, and 0.999, respectively, via the relationship $r = (\exp(\varepsilon) - 1)/(\exp(\varepsilon) + 1)$. As such, (11) simply inflates the width of the non-private Hoeffding CI by 0.762^{-1}, 0.964^{-1}, and 0.999^{-1}, respectively. Hence larger ε (e.g. $\varepsilon \geqslant 4$) leads to CIs that are nearly indistinguishable from the non-private case (Figure 1).

Remark 7. Since Hoeffding-type bounds are not variance-adaptive (meaning they use a worst-case upper-bound on the variance of bounded random variables as in Hoeffding (1963)), they do not benefit from the additional granularity when setting $G_t \geqslant 2$ (see Section B.3 for a detailed mathematical explanation). As such, we set $G_t = 1$ for each t when running NPRR-H. Nevertheless, other CIs are capable of adapting to the variance with $G_t \geqslant 2$, and these are discussed in Appendix B.5, with some suggestions for how to choose $G_t \geqslant 2$ in Appendix B.6. Nevertheless, the empirical performance of our variance-adaptive CIs is illustrated in Figure 4.

3.3. Time-uniform Confidence Sequences for $\mu^\star$

Previously, we focused on constructing a (lower) CI L_n for $\mu^\star$, meaning that L_n satisfies the high-probability guarantee $\mathbb{P}(\mu^\star \geqslant L_n) \geqslant 1 - \alpha$ for the prespecified sample size

n. We will now derive CSs — i.e. entire *sequences* of CIs $(L_t)_{t=1}^\infty$ — which have the stronger *time-uniform* coverage guarantee $\mathbb{P}(\forall t,\ \mu^\star \geqslant L_t) \geqslant 1-\alpha$, enabling anytime-valid inference in sequential regimes. See Section 1.2 for a review of the mathematical and practical differences between CIs and CSs. In summary, if $(L_t)_{t=1}^\infty$ is a lower $(1-\alpha)$-CS, then L_τ forms a valid $(1-\alpha)$-CI at arbitrary stopping times τ (including random and data-dependent times) and hence a practitioner can continuously update inferences as new data are collected, without any penalties for "peeking" at the data early. Let us now present a Hoeffding-type CS for $\mu^\star$, serving as a time-uniform analogue of Theorem 4.

Theorem 8 (`NPRR-H-CS`)**.** *Let $(Z_t)_{t=1}^\infty \sim Q$ for some $Q \in \mathcal{Q}_{\mu^\star}^\infty$. Define the modified mean estimator under `NPRR`:*

$$\widehat{\mu}_t(\lambda_1^t) := \frac{\sum_{i=1}^t \lambda_i \cdot (Z_i - (1-r_i)/2)}{\sum_{i=1}^t r_i \lambda_i}, \tag{13}$$

and let $(\lambda_t)_{t=1}^\infty$ be a real-valued sequence of tuning parameters (discussed in (34)). Then,

$$\bar{L}_t^{\mathrm{H}} := \widehat{\mu}_t(\lambda_1^t) - \frac{\log(1/\alpha) + \sum_{i=1}^t \lambda_i^2/8}{\sum_{i=1}^t r_i \lambda_i} \tag{14}$$

forms a lower $(1-\alpha, (\varepsilon_t)_t)$-LPCS for $\mu^\star$.

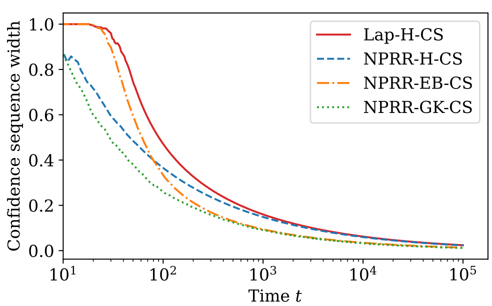

Figure 5. Widths of (90%, 2)-LPCSs for the mean of a Beta(50, 50) distribution. Like Figure 4, Hoeffding-based methods (Lap-H-CS and NPRR-H-CS found in Proposition 1 and Theorem 8) do worse than the variance-adaptive ones (NPRR-EB-CS and NPRR-GK-CS in Proposition 3 and Theorem 11) for large t, though NPRR-H-CS does outperform NPRR-EB-CS for small t. Nevertheless, in all cases, we find that `NPRR`-based CSs outperform Lap-H-CS in both small and large t regimes.

The proof can be found in Section A.4. Unlike Theorem 4, we suggest setting

$$\lambda_t := \sqrt{\frac{8\log(1/\alpha)}{t\log(t+1)}} \wedge 1, \tag{15}$$

to ensure that $\bar{L}_t^{\mathrm{H}} \asymp O(\sqrt{\log t/t})$ up to $\log\log t$ factors. Waudby-Smith & Ramdas (2023, Section 3.3) give a derivation and discussion of λ_t and the $O(\sqrt{\log t/t})$ rate. Similar to Theorem 4, we recommend setting r_t for the desired ε_t-LDP level via (6) and $G_t := 1$ for all t.

The similarity between Theorem 8 and Theorem 4 is no coincidence: indeed, Theorem 4 is a corollary of Theorem 8 where we instantiated a CS at a fixed sample size n and set $\lambda_1 = \cdots = \lambda_n = \sqrt{8\log(1/\alpha)/n}$. In fact, every Cramér-Chernoff bound (even in the non-private regime) has an underlying supermartingale and CS that are rarely exploited (Howard et al., 2020), but setting λ's as in Theorem 4 tightens these CSs for the fixed time n — yielding $O(1/\sqrt{n})$ rates but only for a fixed n — while tuning λ_t as in (15) allows them to spread their efficiency over all t — yielding $O(\sqrt{\log t/t})$ rates but for all t simultaneously. Notice that both the time-uniform and fixed-time bounds in Theorems 4 and 8 cover an unchanging real-valued mean $\mu^\star \in \mathbb{R}$ — in the following section, we will relax this assumption and allow for the mean of each X_i to change over time in an arbitrary matter, but still derive CSs for sensible parameters.

3.4. Confidence Sequences for Time-varying Means

All of the bounds derived thus far have been concerned with estimating some common $\mu^\star$ under the nonparametric assumption $(X_t)_{t=1}^\infty \sim P$ for some $P \in \mathcal{P}_{\mu^\star}^\infty$ and hence $(Z_t)_{t=1}^\infty \sim Q$ for some $Q \in \mathcal{Q}_{\mu^\star}^\infty$. Let us now consider the more general (and challenging) task of constructing CSs for *the average mean so far* $\widetilde{\mu}_t^\star := \frac{1}{t}\sum_{i=1}^t \mu_i^\star$ under the assumption that each X_t has a different mean $\mu_t^\star$. In what follows, we require that `NPRR` is non-interactive, i.e. $r_t = r \in (0,1]$ and $G_t = G \in \{1,2,\dots\}$ for each t.

Theorem 9 (Confidence sequences for time-varying means)**.** *Suppose $X_1, X_2, \dots$ are independent $[0,1]$-bounded random variables with individual means $\mathbb{E}X_t = \mu_t^\star$ for each t, and let $Z_1, Z_2 \dots$ be their privatized views according to `NPRR` without sequential interactivity. Define*

$$\widehat{\mu}_t := \frac{\sum_{i=1}^t (Z_i - (1-r)/2)}{tr}, \tag{16}$$

$$\text{and } \widetilde{B}_t^\pm := \sqrt{\frac{t\beta^2+1}{2(tr\beta)^2}\log\left(\frac{\sqrt{t\beta^2+1}}{\alpha}\right)}, \tag{17}$$

for any $\beta > 0$. Then, $\widetilde{C}_t^\pm := (\widehat{\mu}_t \pm \widetilde{B}_t^\pm)$ forms a two-sided $(1-\alpha, \varepsilon)$-LPCS for $\widetilde{\mu}_t^\star$, where $\varepsilon = \log(1 + \frac{2r}{1-r})$.

The proof in Section A.5 uses a sub-Gaussian mixture supermartingale technique similar to Robbins (1970) and Howard et al. (2020; 2021). The parameter $\beta > 0$ is a tuning parameter dictating a time for which the CS boundary is optimized. Regardless of how β is chosen, $\widetilde{C}_t^\pm$ has the time-uniform coverage guarantee given in Theorem 9 but finite-sample performance can be improved near a particular time t_0 by selecting

$$\beta_\alpha(t_0) := \sqrt{\frac{-2\log\alpha + \log(-2\log\alpha + 1)}{t_0}}, \tag{18}$$

which approximately minimizes $\widetilde{B}_{t_0}$; see Howard et al. (2021, Section 3.5) for details.

Notice that in the non-private case where $r = 1$, we have that $\widetilde{C}_t^{\pm}$ recovers Robbins' sub-Gaussian mixture CS (Robbins, 1970; Howard et al., 2021). Notice that while Theorem 9 handles a strictly more general and challenging problem than the previous sections (by tracking a time-varying mean $(\widetilde{\mu}_t)_{t=1}^{\infty}$), it has the restriction that NPRR must be non-interactive. There is a technical reason for this that boils down to it being difficult to combine time-varying *tuning parameters* (such as those in Theorem 8) with time-varying *estimands* in the same CS. This challenge has appeared in other (non-private) works on CSs (Waudby-Smith & Ramdas, 2023; Howard et al., 2021). In short, this paper has methods for tracking a time-varying mean under non-interactive NPRR or a fixed mean under sequentially interactive NPRR, but not both simultaneously — this would be an interesting direction to explore in future work.

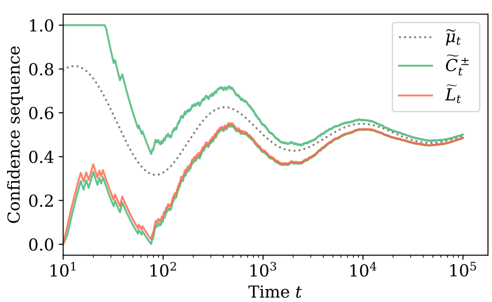

Figure 6. $(90\%, 2)$-LPCSs for the average time-varying mean so far $\widetilde{\mu}_t^{\star}$ with the boundary optimized for $t_0 = 100$. In this example, we set $\mu_t^{\star} = \frac{1}{2}\left[1 - \sin(2\log(e+t))/\log(e+0.01t)\right]$ to produce the displayed sinusoidal behavior. Notice that $\widetilde{L}_t$ is tighter at the expense of only being one-sided. In either case, however, the CSs adapt to non-stationarity and capture $\widetilde{\mu}_t^{\star}$ uniformly over time.

A one-sided analogue of Theorem 9 is presented in Appendix B.7 via slightly different techniques.

4. Illustration: Private Online A/B Testing

Our methods can be used to conduct locally private *online A/B tests* (sequential randomized experiments). Broadly, an A/B test is a statistically principled way of comparing two different *treatments* — e.g. administering drug A versus drug B in a clinical trial. In its simplest form, A/B testing proceeds by (i) randomly assigning subjects to receive treatment A with some probability $\pi \in (0, 1)$ and treatment B with probability $1 - \pi$, (ii) collecting some outcome measurement Y_t for each subject $t \in \{1, 2, \dots\}$ — e.g. severity of headache after taking drug A or B — and (iii) measuring the difference in that outcome between the two groups. An *online* A/B test is one that is conducted sequentially over time — e.g. a sequential clinical trial where patients are recruited one after the other or in batches.

We now illustrate how to sequentially test for the mean difference in outcomes between groups A and B when only given access to locally private data. To set the stage, suppose that $(A_1, Y_1), (A_2, Y_2), \dots$ are random variables such that $A_t \sim \text{Bernoulli}(\pi)$ is 1 if subject t received treatment A and 0 if they received treatment B, and Y_t is a $[0, 1]$-bounded outcome of interest after being assigned treatment A_t.

Using the techniques of Section 3.4, we will construct $(1 - \alpha)$-CSs for the *time-varying mean* $\widetilde{\Delta}_t := \frac{1}{t}\sum_{i=1}^{t} \Delta_i$ where $\Delta_i := \mathbb{E}(Y_i \mid A_i = 1) - \mathbb{E}(Y_i \mid A_i = 0)$ is the mean difference in the outcomes at time i. In words, $\widetilde{\Delta}_t$ is the mean difference in outcomes *among the subjects so far*.

Unlike Section 3.4, however, we will not directly privatize $(Y_t)_{t=1}^{\infty}$, but instead will apply NPRR to some "pseudo-outcomes" $\varphi_t \equiv \varphi_t(Y_t, A_t)$ — functions of Y_t and A_t,

$$\varphi_t := \frac{f_t + \frac{1}{1-\pi}}{\frac{1}{\pi} + \frac{1}{1-\pi}}, \quad \text{where } f_t := \left[\frac{Y_t A_t}{\pi} - \frac{Y_t(1 - A_t)}{1 - \pi}\right].$$

Notice that due to the fact that $Y_t, A_t \in [0, 1]$, we have $f_t \in [-1/(1-\pi), 1/\pi]$, and hence $\varphi_t \in [0, 1]$. Now that we have $[0, 1]$-bounded random variables $(\varphi_t)_{t=1}^{\infty}$, we can obtain their NPRR-induced ε-LDP views $(\psi_t)_{t=1}^{\infty}$ by setting $G_t = 1$ and $r_t = \exp\{\varepsilon - 1\}/\exp\{\varepsilon + 1\}$ for each t. Notice that we are privatizing φ_t which is a function of both Y_t and A_t, so both the outcome *and* the treatment are protected with ε-LDP.

Corollary 1 (Locally private online A/B estimation)**.** *Following the setup above, let $(\psi_t)_{t=1}^{\infty}$ be the NPRR-induced privatized views of $(\varphi_t)_{t=1}^{\infty}$. Define the estimator*

$$\widehat{\varphi}_t := \frac{\sum_{i=1}^{t}(\psi_i - (1-r)/2)}{tr}, \tag{19}$$

and set $\widetilde{B}_t$ as in (46). Then,

$$\widetilde{L}_t^{\Delta} := -\frac{1}{1-\pi} + \left(\frac{1}{\pi} + \frac{1}{1-\pi}\right)\left(\widehat{\varphi}_t - \widetilde{B}_t\right) \tag{20}$$

is a lower $(1 - \alpha, \varepsilon)$-LPCS for $\widetilde{\Delta}_t$.

The proof is an immediate consequence of the well-known fact about "inverse-probability-weighted" estimators that $\mathbb{E}f_t = \Delta_t$ for every t (Horvitz & Thompson, 1952; Robins et al., 1994), combined with Proposition 4. Similarly, a two-sided CS can be obtained by replacing $\widehat{\varphi}_t - \widetilde{B}_t$ in (20) with $\widehat{\varphi}_t \pm \widetilde{B}_t^{\pm}$, where $\widetilde{B}_t^{\pm}$ is given in (17).

Practical implications. The implications of Corollary 1 for the practitioner are threefold:

1. The CSs can be continuously monitored from the start of the A/B test and for an indefinite amount of time;

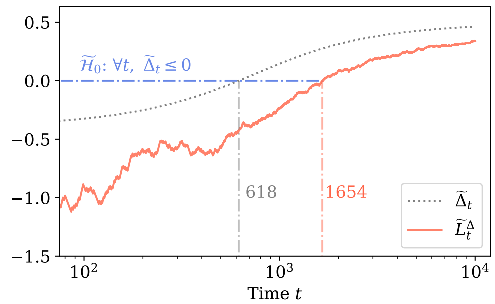

Figure 7. An example of Corollary 1 applied to the time-varying mean given by $\Delta_t := 1.8(\exp\{t/300\}/(1+\exp\{t/300\}) - 1/2)$. In this particular example, we have that $\widetilde{\Delta}_t := \frac{1}{t}\sum_{i=1}^t \Delta_i$ changes from negative to positive at time 618, and yet our lower CS $\widetilde{L}_t^{\Delta}$ later detects this change at time 1654, at which point the weak null $\widetilde{\mathcal{H}}_0$: $\forall t,\ \widetilde{\Delta}_t \leqslant 0$ can be rejected (see Appendix B.9 for details regarding the composite hypothesis $\widetilde{\mathcal{H}}_0$ and how to test it).

2. Inferences made from $\widetilde{L}_\tau^{\Delta}$ are valid at any stopping time τ, regardless of why the test is stopped; and
3. $\widetilde{L}_t^{\Delta}$ adapts to non-stationarity: if the treatment differences Δ_t drift over time, $\widetilde{L}_t^{\Delta}$ still forms an LPCS for $\widetilde{\Delta}_t$. But if $\Delta_1 = \Delta_2 = \cdots = \Delta^\star$ is constant, then $\widetilde{L}_t^{\Delta}$ forms an LPCS for $\Delta^\star$.

5. Additional Results & Summary

Both NPRR and our proof techniques are general-purpose tools with several other implications for locally private statistical inference, including confidence sets via the Laplace mechanism, variance-adaptive inference, and sequential hypothesis testing. We briefly expand on these implications here, and leave their details to the appendix.

- §B.4: **Confidence sets via the Laplace mechanism.** We introduced NPRR as an extension of randomized response for arbitrary bounded data (rather than just binary), but of course the Laplace mechanism also handles bounded data. While NPRR enjoys advantages over Laplace as discussed in Section 2, it may still be of interest to derive confidence sets from data that are privatized via Laplace, given its ubiquity and simplicity. Appendix B.4 presents new nonparametric CIs and CSs for population means under the Laplace mechanism.
- §B.5: **Variance-adaptive inference.** Notice that the CIs and CSs presented in Section 3 were not variance-adaptive due to the fact that they relied on sub-Gaussianity of bounded random variables. However, this is not necessary, and we present other locally private *variance-adaptive* CIs and CSs in Appendix B.5.
- §B.8: **Sequential hypothesis testing.** While the statistical procedures of this paper have taken the form of CIs and CSs rather than hypothesis tests, there is a deep relationship between the two, and our results have analogues that could have been presented in the language of the latter. Appendix B.8 articulates this relationship and presents explicit (sequential) tests.
- §B.10: **Adaptive online A/B testing.** Corollary 1 assumes a common propensity score π among all subjects for simplicity of exposition, but it is also possible to derive CSs for $\widetilde{\Delta}_t$ under an adaptive framework where propensity scores $(\pi_t(X_t))_{t=1}^\infty$ can change over time in a data-dependent fashion, and be functions of some measured covariates $(X_t)_{t=1}^\infty$. The details of this more complex setup are left to Appendix B.10.

Another followup problem that we do not explicitly address here but that can be solved using our techniques is locally private *variance* estimation. Notice that the variance $\mathrm{Var}(X) := \mathbb{E}(X^2) - (\mathbb{E}(X))^2$ is a function of two expectations, $\mathbb{E}(X^2)$ and $\mathbb{E}(X)$. Since X^2 is also $[0,1]$-bounded if X is, we can use all of the techniques in this paper to derive two separate $(1-\alpha/2, \varepsilon/2)$-LPCIs (or LPCSs) to derive a $(1-\alpha, \varepsilon)$-LPCI for $\mathrm{Var}(X)$. Of course this requires collecting privatized views of both X^2 and X separately. As a further generalization, a similar argument can be made for the construction of LPCIs for the covariance of X and Y since $\mathrm{Cov}(X,Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$ (though here we would need to construct $(1-\alpha/3, \varepsilon/3)$-LPCIs, etc.).

A limitation of the present paper is that we have only discussed confidence sets for univariate parameters. Indeed, it is not immediately clear to us what is the right way to generalize NPRR to the multivariate case, or how to derive LPCIs and LPCSs for means of random *vectors* given such a generalization. This is an open direction for future work.

With the growing interest in protecting user privacy, an increasingly important addition to the statistician's toolbox are methods that can extract population information from privatized data. In this paper, we derived nonparametric confidence intervals and time-uniform confidence sequences for population means from locally private data. We introduced, NPRR a nonparametric and sequentially interactive extension of Warner's randomized response for bounded data. The privatized output from NPRR can then be harnessed to produce confidence sets for the mean of the raw data distribution. Importantly, our confidence sets are sharp, some attaining optimal theoretical convergence rates and others simply having excellent empirical performance, not only making private nonparametric (sequential) inference possible, but practical. In future work, we aim to apply these general-purpose tools to changepoint detection, two-sample testing, and (conditional) independence testing.

Acknowledgments. ZSW was supported in part by NSF CNS2120667. AR acknowledges NSF DMS2310718.

References

Acharya, J., Canonne, C. L., Han, Y., Sun, Z., and Tyagi, H. Domain compression and its application to randomness-optimal distributed goodness-of-fit. In Abernethy, J. and Agarwal, S. (eds.), *Proceedings of Thirty Third Conference on Learning Theory*, volume 125 of *Proceedings of Machine Learning Research*, pp. 3–40. PMLR, 09–12 Jul 2020a. URL https://proceedings.mlr.press/v125/acharya20a.html.

Acharya, J., Canonne, C. L., and Tyagi, H. Inference under information constraints I: Lower bounds from Chi-square contraction. *IEEE Transactions on Information Theory*, 66(12):7835–7855, 2020b.

Acharya, J., Canonne, C. L., Freitag, C., Sun, Z., and Tyagi, H. Inference under information constraints III: Local privacy constraints. *IEEE Journal on Selected Areas in Information Theory*, 2(1):253–267, 2021a.

Acharya, J., Sun, Z., and Zhang, H. Differentially private Assouad, Fano, and Le Cam. In *Algorithmic Learning Theory*, pp. 48–78. PMLR, 2021b.

Acharya, J., Canonne, C. L., Liu, Y., Sun, Z., and Tyagi, H. Interactive inference under information constraints. *IEEE Transactions on Information Theory*, 68(1):502–516, 2022.

Amin, K., Joseph, M., and Mao, J. Pan-private uniformity testing. In Abernethy, J. and Agarwal, S. (eds.), *Proceedings of Thirty Third Conference on Learning Theory*, volume 125 of *Proceedings of Machine Learning Research*, pp. 183–218. PMLR, 09–12 Jul 2020. URL https://proceedings.mlr.press/v125/amin20a.html.

Apple Inc. Differential privacy overview. https://www.apple.com/privacy/docs/Differential_Privacy_Overview.pdf, 2022. Accessed: 2022-02-01.

Audibert, J.-Y., Munos, R., and Szepesvári, C. Tuning bandit algorithms in stochastic environments. In *International conference on algorithmic learning theory*, pp. 150–165. Springer, 2007.

Awan, J. and Slavković, A. Differentially private uniformly most powerful tests for binomial data. *Advances in Neural Information Processing Systems*, 31:4208–4218, 2018.

Barnes, L. P., Chen, W.-N., and Özgür, A. Fisher information under local differential privacy. *IEEE Journal on Selected Areas in Information Theory*, 1(3):645–659, 2020.

Barnes, R., Cooke-Yarborough, E., and Thomas, D. An electronic digital computor using cold cathode counting tubes for storage. *Electronic Engineering*, 23:286–91, 1951.

Bentkus, V. On Hoeffding's inequalities. *The Annals of Probability*, 32(2):1650–1673, 2004.

Berrett, T. and Butucea, C. Locally private non-asymptotic testing of discrete distributions is faster using interactive mechanisms. In Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M., and Lin, H. (eds.), *Advances in Neural Information Processing Systems*, volume 33, pp. 3164–3173. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper/2020/file/20b02dc95171540bc52912baf3aa709d-Paper.pdf.

Berrett, T. and Yu, Y. Locally private online change point detection. *Advances in Neural Information Processing Systems*, 34, 2021.

Butucea, C. and ISSARTEL, Y. Locally differentially private estimation of functionals of discrete distributions. In Ranzato, M., Beygelzimer, A., Dauphin, Y., Liang, P., and Vaughan, J. W. (eds.), *Advances in Neural Information Processing Systems*, volume 34, pp. 24753–24764. Curran Associates, Inc., 2021. URL https://proceedings.neurips.cc/paper/2021/file/cf8c9be2a4508a24ae92c9d3d379131d-Paper.pdf.

Butucea, C., Dubois, A., Kroll, M., and Saumard, A. Local differential privacy: Elbow effect in optimal density estimation and adaptation over Besov ellipsoids. *Bernoulli*, 26(3):1727–1764, 2020.

Canonne, C. L., Kamath, G., McMillan, A., Smith, A., and Ullman, J. The structure of optimal private tests for simple hypotheses. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, pp. 310–321, 2019.

Couch, S., Kazan, Z., Shi, K., Bray, A., and Groce, A. Differentially private nonparametric hypothesis testing. In *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security*, pp. 737–751, 2019.

Covington, C., He, X., Honaker, J., and Kamath, G. Unbiased statistical estimation and valid confidence intervals under differential privacy. *arXiv preprint arXiv:2110.14465*, 2021.

Darling, D. A. and Robbins, H. Confidence sequences for mean, variance, and median. *Proceedings of the National Academy of Sciences of the United States of America*, 58 (1):66, 1967.

Ding, B., Kulkarni, J., and Yekhanin, S. Collecting telemetry data privately. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, pp. 3574–3583, 2017.

Drechsler, J., Globus-Harris, I., McMillan, A., Sarathy, J., and Smith, A. Non-parametric differentially private confidence intervals for the median. *arXiv preprint arXiv:2106.10333*, 2021.

Duchi, J. C. and Ruan, F. The right complexity measure in locally private estimation: It is not the fisher information. *arXiv preprint arXiv:1806.05756*, 2018.

Duchi, J. C., Jordan, M. I., and Wainwright, M. J. Local privacy and statistical minimax rates. In *2013 IEEE 54th Annual Symposium on Foundations of Computer Science*, pp. 429–438. IEEE, 2013a.

Duchi, J. C., Jordan, M. I., and Wainwright, M. J. Local privacy and minimax bounds: sharp rates for probability estimation. In *Proceedings of the 26th International Conference on Neural Information Processing Systems-Volume 1*, pp. 1529–1537, 2013b.

Duchi, J. C., Jordan, M. I., and Wainwright, M. J. Minimax optimal procedures for locally private estimation. *Journal of the American Statistical Association*, 113(521):182–201, 2018.

Dwork, C. and Roth, A. The algorithmic foundations of differential privacy. *Found. Trends Theor. Comput. Sci.*, 9(3-4):211–407, 2014.

Dwork, C., McSherry, F., Nissim, K., and Smith, A. Calibrating noise to sensitivity in private data analysis. In *Theory of cryptography conference*, pp. 265–284. Springer, 2006.

Fan, X., Grama, I., and Liu, Q. Exponential inequalities for martingales with applications. *Electronic Journal of Probability*, 20:1–22, 2015.

Ferrando, C., Wang, S., and Sheldon, D. Parametric bootstrap for differentially private confidence intervals. In *International Conference on Artificial Intelligence and Statistics*, pp. 1598–1618. PMLR, 2022.

Forsythe, G. E. Reprint of a note on rounding-off errors. *SIAM review*, 1(1):66, 1959.

Gaboardi, M., Lim, H., Rogers, R., and Vadhan, S. Differentially private chi-squared hypothesis testing: Goodness of fit and independence testing. In *International conference on machine learning*, pp. 2111–2120. PMLR, 2016.

Gaboardi, M., Rogers, R., and Sheffet, O. Locally private mean estimation: z-test and tight confidence intervals. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pp. 2545–2554. PMLR, 2019.

Grünwald, P., de Heide, R., and Koolen, W. Safe testing. *arXiv preprint arXiv:1906.07801*, 2019.

Hoeffding, W. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963.

Horvitz, D. G. and Thompson, D. J. A generalization of sampling without replacement from a finite universe. *Journal of the American statistical Association*, 47(260):663–685, 1952.

Howard, S. R., Ramdas, A., McAuliffe, J., and Sekhon, J. Time-uniform Chernoff bounds via nonnegative supermartingales. *Probability Surveys*, 17:257–317, 2020.

Howard, S. R., Ramdas, A., McAuliffe, J., and Sekhon, J. Time-uniform, nonparametric, nonasymptotic confidence sequences. *The Annals of Statistics*, 49(2):1055–1080, 2021.

Hull, T. E. and Swenson, J. R. Tests of probabilistic models for propagation of roundoff errors. *Communications of the ACM*, 9(2):108–113, 1966.

Johari, R., Koomen, P., Pekelis, L., and Walsh, D. Peeking at A/B tests: Why it matters, and what to do about it. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1517–1525, 2017.

Joseph, M., Kulkarni, J., Mao, J., and Wu, S. Z. Locally private Gaussian estimation. *Advances in Neural Information Processing Systems*, 32:2984–2993, 2019.

Jun, K.-S. and Orabona, F. Parameter-free online convex optimization with sub-exponential noise. In *Conference on Learning Theory*, pp. 1802–1823. PMLR, 2019.

Kairouz, P., Oh, S., and Viswanath, P. Extremal mechanisms for local differential privacy. *Advances in neural information processing systems*, 27, 2014.

Kairouz, P., Bonawitz, K., and Ramage, D. Discrete distribution estimation under local privacy. In *International Conference on Machine Learning*, pp. 2436–2444. PMLR, 2016.

Kamath, G., Singhal, V., and Ullman, J. Private mean estimation of heavy-tailed distributions. In *Conference on Learning Theory*, pp. 2204–2235. PMLR, 2020.

Karwa, V. and Vadhan, S. Finite sample differentially private confidence intervals. In *9th Innovations in Theoretical Computer Science Conference (ITCS 2018)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2018.

Kasiviswanathan, S. P., Lee, H. K., Nissim, K., Raskhodnikova, S., and Smith, A. What can we learn privately? *SIAM Journal on Computing*, 40(3):793–826, 2011.

Kuchibhotla, A. K. and Zheng, Q. Near-optimal confidence sequences for bounded random variables. In *International Conference on Machine Learning*, pp. 5827–5837. PMLR, 2021.

Li, Z., Wang, T., Lopuhaä-Zwakenberg, M., Li, N., and Škoric, B. Estimating numerical distributions under local differential privacy. In *Proceedings of the 2020 ACM SIGMOD International Conference on Management of Data*, pp. 621–635, 2020.

Maurer, A. and Pontil, M. Empirical Bernstein bounds and sample variance penalization. In *Conference on Learning Theory*, pp. 2372–2387. PMLR, 2009.

Mironov, I. On significance of the least significant bits for differential privacy. In *Proceedings of the 2012 ACM conference on Computer and communications security*, pp. 650–661, 2012.

Orabona, F. and Jun, K.-S. Tight concentrations and confidence sequences from the regret of universal portfolio. *arXiv preprint arXiv:2110.14099*, 2021.

Ramdas, A., Ruf, J., Larsson, M., and Koolen, W. M. Testing exchangeability: Fork-convexity, supermartingales and e-processes. *International Journal of Approximate Reasoning*, 2021.

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

Robins, J. M., Rotnitzky, A., and Zhao, L. P. Estimation of regression coefficients when some regressors are not always observed. *Journal of the American statistical Association*, 89(427):846–866, 1994.

Shafer, G. Testing by betting: A strategy for statistical and scientific communication. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 184(2):407–431, 2021.

Shafer, G., Shen, A., Vereshchagin, N., and Vovk, V. Test martingales, Bayes factors and p-values. *Statistical Science*, 26(1):84–101, 2011.

ter Schure, J. and Grünwald, P. ALL-IN meta-analysis: breathing life into living systematic reviews. *arXiv preprint arXiv:2109.12141*, 2021.

Ville, J. Etude critique de la notion de collectif. *Bull. Amer. Math. Soc*, 45(11):824, 1939.

Vovk, V. and Wang, R. E-values: Calibration, combination and applications. *The Annals of Statistics*, 49(3):1736–1754, 2021.

Vu, D. and Slavkovic, A. Differential privacy for clinical trial data: Preliminary evaluations. In *2009 IEEE International Conference on Data Mining Workshops*, pp. 138–143. IEEE, 2009.

Wald, A. Sequential tests of statistical hypotheses. *The annals of mathematical statistics*, 16(2):117–186, 1945.

Wang, N., Xiao, X., Yang, Y., Zhao, J., Hui, S. C., Shin, H., Shin, J., and Yu, G. Collecting and analyzing multidimensional data with local differential privacy. In *2019 IEEE 35th International Conference on Data Engineering (ICDE)*, pp. 638–649. IEEE, 2019.

Wang, R. and Ramdas, A. False discovery rate control with e-values. *Journal of the Royal Statistical Society, Series B*, 2022.

Wang, Y., Lee, J., and Kifer, D. Differentially private hypothesis testing, revisited. *arXiv preprint arXiv:1511.03376*, 2015.

Wang, Y., Sibai, H., Yen, M., Mitra, S., and Dullerud, G. E. Differentially private algorithms for statistical verification of cyber-physical systems. *IEEE Open Journal of Control Systems*, 1:294–305, 2022.

Warner, S. L. Randomized response: A survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association*, 60(309):63–69, 1965.

Wasserman, L. and Zhou, S. A statistical framework for differential privacy. *Journal of the American Statistical Association*, 105(489):375–389, 2010.

Waudby-Smith, I. and Ramdas, A. Confidence sequences for sampling without replacement. *Advances in Neural Information Processing Systems*, 33, 2020.

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

Waudby-Smith, I., Arbour, D., Sinha, R., Kennedy, E. H., and Ramdas, A. Time-uniform central limit theory, asymptotic confidence sequences, and anytime-valid causal inference. *arXiv preprint arXiv:2103.06476*, 2021.

Waudby-Smith, I., Wu, L., Ramdas, A., Karampatziakis, N., and Mineiro, P. Anytime-valid off-policy inference for contextual bandits. *arXiv preprint arXiv:2210.10768*, 2022.

Zhao, S., Zhou, E., Sabharwal, A., and Ermon, S. Adaptive concentration inequalities for sequential decision problems. *Advances in Neural Information Processing Systems*, 29, 2016.

A. Proofs of main results

A.1. Prelude: filtrations, supermartingales, and Ville's inequality

By far the most common way to derive a CS is by constructing a nonnegative supermartingales and then applying Ville's maximal inequality to it. Indeed, all of the proofs for our CS and CI results employ this technique. However, in order to discuss supermartingales we must first review *filtrations*. A filtration $\mathcal{F} \equiv (\mathcal{F}_t)_{t=0}^{\infty}$ is a nondecreasing sequence of sigma-algebras $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \cdots$, and a stochastic process $(M_t)_{t=0}^{\infty}$ is said to be *adapted* to $\mathcal{F}$ if M_t is $\mathcal{F}_t$-measurable for all $t \in \mathbb{N}$. On the other hand, $(M_t)_{t=1}^{\infty}$ is said to be $\mathcal{F}$-predictable if each M_t is $\mathcal{F}_{t-1}$-measurable — informally "M_t depends on the past".

For example, the canonical filtration $\mathcal{X}$ generated by a sequence of random variables $(X_t)_{t=1}^{\infty}$ is given by the sigma-algebra generated by X_1^t, i.e. $\mathcal{X}_t := \sigma(X_1^t)$ for each $t \in \{1, 2, \dots\}$, and $\mathcal{X}_0$ is the trivial sigma-algebra. A function $M_t \equiv M(X_1^t)$ depending only on X_1^t forms a $\mathcal{X}$-adapted process, while $(M_{t-1})_{t=1}^{\infty}$ is $\mathcal{X}$-predictable. Likewise, if we obtain a privatized view $(Z_t)_{t=1}^{\infty}$ of $(X_t)_{t=1}^{\infty}$ using some locally private mechanism, a different filtration $\mathcal{Z}$ emerges, given by $\mathcal{Z}_t := \sigma(Z_1^t)$. Throughout our proofs, $\mathcal{Z}$-adapted and $\mathcal{Z}$-predictable processes will be central mathematical objects.

A process $(M_t)_{t=0}^{\infty}$ adapted to $\mathcal{F}$ is a *supermartingale* if

$$\mathbb{E}(M_t \mid \mathcal{F}_{t-1}) \leqslant M_{t-1} \text{ for each } t \geqslant 1. \tag{21}$$

If the above inequality is replaced by an equality, then $(M_t)_{t=0}^{\infty}$ is a *martingale*. The methods in this paper will involve derivations of (super)martingales which are nonnegative and begin at one — often referred to as "test (super)martingales" (Shafer et al., 2011) or simply "nonnegative (super)martingales" (NMs or NSMs for martingales and supermartingales, respectively) (Robbins, 1970; Howard et al., 2020). NSMs $(M_t)_{t=0}^{\infty}$ satisfy the following powerful concentration inequality due to Ville (1939):

$$\mathbb{P}(\exists t \in \mathbb{N} : M_t \geqslant 1/\alpha) \leqslant \alpha. \tag{22}$$

In other words, they are unlikely to ever grow too large.

In the CS proofs that follow, we will focus on deriving processes $(M_t(\mu))_{t=1}^{\infty}$ for any $\mu \in [0,1]$ such that when μ is equal to the true mean of interest $\mu^\star$, we have that $M_t(\mu^\star)$ forms a NSM. In this case, it turns out that the set of μ such that $M_t(\mu)$ is less than $1/\alpha$ — i.e. $C_t := \{\mu \in [0,1] : M_t(\mu) < 1/\alpha\}$ — forms a $(1-\alpha)$-CS for $\mu^\star$. This is easy to see since $\mu^\star \notin C_t$ if and only if $M_t(\mu^\star) \geqslant 1/\alpha$, and thus

$$\mathbb{P}(\exists t \in \mathbb{N} : \mu^\star \notin C_t) = \mathbb{P}(\exists t \in \mathbb{N} : M_t(\mu^\star) \geqslant 1/\alpha) \leqslant \alpha, \tag{23}$$

where the last inequality is precisely (22). The CS proofs that follow will make the exact processes $(M_t(\mu))_{t=1}^{\infty}$ explicit.

A.2. Proof of Theorem 1

Theorem 1 (NPRR satisfies LDP)**.** *Suppose $(Z_t)_{t=1}^{\infty}$ are generated according to* NPRR*. Then for each $t \in \{1, 2, \dots\}$, Z_t is a conditionally ε_t-LDP view of X_t with*

$$\varepsilon_t := \log\left(1 + \frac{(G_t+1)r_t}{1-r_t}\right). \tag{5}$$

Proof. We will prove the result for fixed $r \in (0,1), G \geqslant 1$ but it is straightforward to generalize the proof for r_t depending on Z_1^{t-1}. It suffices to verify that the likelihood ratio $L(x, \widetilde{x})$ is bounded above by $\exp(\varepsilon)$ for any $x, \widetilde{x} \in [0,1]$. Writing out the likelihood ratio $L(x, \widetilde{x})$, we have

$$L(x,\widetilde{x}) := \frac{\frac{1-r}{G+1} + rG \cdot \left\{\mathbb{1}(Z = x^{\text{ceil}})(x - x^{\text{floor}}) + \mathbb{1}(Z = x^{\text{floor}})\left[1/G - (x - x^{\text{floor}})\right]\right\}}{\frac{1-r}{G+1} + rG \cdot \{\mathbb{1}(Z = \widetilde{x}^{\text{ceil}})(\widetilde{x} - \widetilde{x}^{\text{floor}}) + \mathbb{1}(Z = \widetilde{x}^{\text{floor}})\left[1/G - (\widetilde{x} - \widetilde{x}^{\text{floor}})\right]\}},$$

which is dominated by the counting measure. Notice that the numerator of L is maximized when x already lies in the discretized range, i.e. $Z = x = x^{\text{ceil}} = x^{\text{floor}}$ so that the numerator becomes $\frac{1-r}{G+1} + r$, while the denominator is minimized when $Z \neq \widetilde{x}^{\text{ceil}}$ and $Z \neq \widetilde{x}^{\text{floor}}$ so that the denominator becomes $\frac{1-r}{G+1}$. Therefore, we have that with probability one,

$$L(x,\widetilde{x}) \leqslant \frac{\frac{1-r}{G+1} + r}{\frac{1-r}{G+1}} = 1 + \frac{(G+1)r}{1-r},$$

and thus NPRR is ε-locally DP with $\varepsilon := \log(1 + (G+1)r/(1-r))$. □

A.3. Proof of Theorem 4

Theorem 4 (NPRR-H)**.** *Suppose* $(X_t)_{t=1}^n \sim P$ *for some* $P \in \mathcal{P}_{\mu^\star}^n$*, and let* $(Z_t)_{t=1}^n \sim Q \in \mathcal{Q}_{\mu^\star}^n$ *be their privatized views via* NPRR*. Define the* NPRR*-adjusted sample mean*

$$\widehat{\mu}_n := \frac{\frac{1}{n}\sum_{i=1}^n (Z_i - (1-r_i)/2)}{\frac{1}{n}\sum_{i=1}^n r_i}. \tag{10}$$

Then,

$$\dot{L}_n^{\mathrm{H}} := \widehat{\mu}_n - \sqrt{\frac{\log(1/\alpha)}{2n(\frac{1}{n}\sum_{i=1}^n r_i)^2}} \tag{11}$$

is a lower $(1-\alpha, (\varepsilon_t)_t)$*-LPCI for* $\mu^\star$*.*

Proof. The proof proceeds in two steps. First we note that $\bar{L}_t^{\mathrm{H}}$ forms a $(1-\alpha)$-lower confidence *sequence*, and then instantiate this fact at the sample size n.

Step 1. $\bar{L}_t^{\mathrm{H}}$ **forms a** $(1-\alpha)$**-lower CS.** This is exactly the statement of Theorem 8.

Step 2. $\dot{L}_n^{\mathrm{H}}$ **is a lower-CI.** By Step 1, we have that $\bar{L}_t^{\mathrm{H}}$ forms a $(1-\alpha)$-lower CS, meaning

$$\mathbb{P}(\forall t \in \{1, \dots, n\},\ \mu^\star \geqslant \bar{L}_t^{\mathrm{H}}) \geqslant 1 - \alpha.$$

Therefore,

$$\mathbb{P}\left(\mu^\star \geqslant \max_{1 \leqslant t \leqslant n} \bar{L}_t^{\mathrm{H}}\right) = \mathbb{P}(\mu^\star \geqslant \dot{L}_n^{\mathrm{H}}) \geqslant 1 - \alpha,$$

which completes the proof. □

A.4. Proof of Theorem 8

Theorem 8 (NPRR-H-CS)**.** *Let* $(Z_t)_{t=1}^\infty \sim Q$ *for some* $Q \in \mathcal{Q}_{\mu^\star}^\infty$*. Define the modified mean estimator under* NPRR*:*

$$\widehat{\mu}_t(\lambda_1^t) := \frac{\sum_{i=1}^t \lambda_i \cdot (Z_i - (1-r_i)/2)}{\sum_{i=1}^t r_i \lambda_i}, \tag{13}$$

and let $(\lambda_t)_{t=1}^\infty$ *be a real-valued sequence of tuning parameters (discussed in (34)). Then,*

$$\bar{L}_t^{\mathrm{H}} := \widehat{\mu}_t(\lambda_1^t) - \frac{\log(1/\alpha) + \sum_{i=1}^t \lambda_i^2/8}{\sum_{i=1}^t r_i \lambda_i} \tag{14}$$

forms a lower $(1-\alpha, (\varepsilon_t)_t)$*-LPCS for* $\mu^\star$*.*

Proof. The proof proceeds in two steps. First, we construct an NSM adapted to the private filtration $\mathcal{Z} \equiv (\mathcal{Z}_t)_{t=0}^\infty$. Second and finally, we apply Ville's inequality to obtain a high-probability upper bound on the NSM, and show that this inequality results in the CS given in Theorem 8.

Step 1. Consider the nonnegative process starting at one given by

$$M_t(\mu^\star) := \prod_{i=1}^t \exp\left\{\lambda_i (Z_i - \zeta_i(\mu^\star)) - \lambda_i^2/8\right\}, \tag{24}$$

where $(\lambda_t)_{t=1}^\infty$ is a real-valued sequence[2] and $\zeta_t(\mu^\star) := r_t\mu^\star + (1-r_t)/2$ as usual. We claim that $(M_t(\mu^\star))_{t=0}^\infty$ is a supermartingale, meaning $\mathbb{E}(M_t(\mu^\star) \mid \mathcal{Z}_{t-1}) \leqslant M_{t-1}(\mu^\star)$. Writing out the conditional expectation of $M_t(\mu^\star)$, we have

$$\begin{aligned}
&\mathbb{E}\left(M_t(\mu^\star) \mid \mathcal{Z}_{t-1}\right) \\
&= \mathbb{E}\left(\prod_{i=1}^t \exp\left\{\lambda_i(Z_i - \zeta_i(\mu^\star)) - \lambda_i^2/8\right\} \;\middle|\; \mathcal{Z}_{t-1}\right) \\
&= \underbrace{\prod_{i=1}^{t-1} \exp\left\{\lambda_i(Z_i - \zeta_i(\mu^\star)) - \lambda_i^2/8\right\}}_{M_{t-1}(\mu^\star)} \cdot \underbrace{\mathbb{E}\left(\exp\left\{\lambda_t(Z_t - \zeta_t(\mu^\star)) - \lambda_t^2/8\right\} \;\middle|\; \mathcal{Z}_{t-1}\right)}_{(\dagger)},
\end{aligned}$$

since $M_{t-1}(\mu^\star)$ is $\mathcal{Z}_{t-1}$-measurable, and thus it can be written outside of the conditional expectation. It now suffices to show that $(\dagger) \leqslant 1$. To this end, note that Z_t is a $[0,1]$-bounded random variable with conditional mean $\mathbb{E}(Z_t \mid \mathcal{Z}_{t-1}) = \zeta_t(\mu^\star)$ by design of NPRR (Algorithm 2). Since bounded random variables are sub-Gaussian (Hoeffding, 1963), we have that

$$\mathbb{E}(\lambda_t(Z_t - \zeta_t(\mu^\star)) \mid \mathcal{Z}_{t-1}) \leqslant \exp\left\{\lambda_t^2/8\right\},$$

and hence $(\dagger) \leqslant 1$. Therefore, $(M_t(\mu^\star))_{t=0}^\infty$ is a $\mathcal{Q}_{\mu^\star}^\infty$-NSM.

Step 2. By Ville's inequality for NSMs (Ville, 1939), we have that

$$\mathbb{P}(\exists t : M_t(\mu^\star) \geqslant 1/\alpha) \leqslant \alpha.$$

In other words, we have that $M_t(\mu^\star) < 1/\alpha$ for all $t \in \mathbb{N}$ with probability at least $1-\alpha$. Using some algebra to rewrite the inequality $M_t(\mu^\star) < 1/\alpha$, we have

$$\begin{aligned}
M_t(\mu^\star) < 1/\alpha &\iff \prod_{i=1}^t \exp\left\{\lambda_i(Z_i - \zeta_i(\mu^\star)) - \lambda_i^2/8\right\} < \frac{1}{\alpha} \\
&\iff \sum_{i=1}^t \left[\lambda_i(Z_i - \zeta_i(\mu^\star)) - \lambda_i^2/8\right] < \log(1/\alpha) \\
&\iff \sum_{i=1}^t \lambda_i Z_i - \mu^\star \sum_{i=1}^t \lambda_i r_i - \sum_{i=1}^t \lambda_i \cdot (1-r_i)/2 - \sum_{i=1}^t \lambda_i^2/8 < \log(1/\alpha) \\
&\iff \mu^\star > \underbrace{\frac{\sum_{i=1}^t \lambda_i \cdot (Z_i - (1-r_i)/2)}{\sum_{i=1}^t r_i\lambda_i}}_{\widehat{\mu}_t(\lambda_1^t)} - \underbrace{\frac{\log(1/\alpha) + \sum_{i=1}^t \lambda_i^2/8}{\sum_{i=1}^t r_i\lambda_i}}_{\bar{B}_t(\lambda_1^t)}
\end{aligned}$$

Therefore, $\bar{L}_t := \widehat{\mu}_t(\lambda_1^t) - \bar{B}_t(\lambda_1^t)$ forms a lower $(1-\alpha)$-CS for $\mu^\star$. The upper CS $\bar{U}_t := \widehat{\mu}_t(\lambda_1^t) + \bar{B}_t(\lambda_1^t)$ can be derived by applying the above proof to $(-Z_t)_{t=1}^\infty$ and their conditional means $(-\zeta_i(\mu^\star))_{t=1}^\infty$. This completes the proof

□

A.5. Proof of Theorem 9

Theorem 9 (Confidence sequences for time-varying means)**.** *Suppose $X_1, X_2, \ldots$ are independent $[0,1]$-bounded random variables with individual means $\mathbb{E}X_t = \mu_t^\star$ for each t, and let $Z_1, Z_2 \ldots$ be their privatized views according to* NPRR *without sequential interactivity. Define*

$$\widehat{\mu}_t := \frac{\sum_{i=1}^t (Z_i - (1-r)/2)}{tr}, \tag{16}$$

$$\textit{and } \widetilde{B}_t^\pm := \sqrt{\frac{t\beta^2+1}{2(tr\beta)^2}\log\left(\frac{\sqrt{t\beta^2+1}}{\alpha}\right)}, \tag{17}$$

for any $\beta > 0$. Then, $\widetilde{C}_t^\pm := (\widehat{\mu}_t \pm \widetilde{B}_t^\pm)$ forms a two-sided $(1-\alpha, \varepsilon)$-LPCS for $\widetilde{\mu}_t^\star$, where $\varepsilon = \log(1 + \frac{2r}{1-r})$.

[2] The proof also works if $(\lambda_t)_{t=1}^\infty$ is $\mathcal{Z}$-predictable but we omit this detail since we typically recommend using real-valued sequences anyway.

Proof. The proof proceeds in three steps. First, we derive a sub-Gaussian NSM indexed by a parameter $\lambda \in \mathbb{R}$. Second, we mix this NSM over λ using the density of a Gaussian distribution, and justify why the resulting process is also an NSM. Third and finally, we apply Ville's inequality and invert the NSM to obtain $(\widetilde{C}_t^{\pm})_{t=1}^{\infty}$.

Step 1: Constructing the λ-indexed NSM. Let $(X_t)_{t=1}^{\infty}$ be independent $[0,1]$-bounded random variables with individual means given by $\mathbb{E}X_t = \mu_t^\star$, and let $(Z_t)_{t=1}^{\infty}$ be the $\texttt{NPRR}$-induced private views of $(X_t)_{t=1}^{\infty}$. Define $\zeta(\mu) := r\mu + (1-r)/2$ for any $\mu \in [0,1]$, and $r \in (0,1]$. Let $\lambda \in \mathbb{R}$ and consider the process,

$$M_t(\lambda) := \prod_{i=1}^{t} \exp\left\{\lambda(Z_i - \zeta(\mu_i^\star)) - \lambda^2/8\right\}, \tag{25}$$

with $M_0(\lambda) \equiv 0$. We claim that (25) forms an NSM with respect to the private filtration $\mathcal{Z}$. The proof technique is nearly identical to that of Theorem 8 but with changing means and $\lambda = \lambda_1 = \lambda_2 = \cdots \in \mathbb{R}$. Indeed, $M_t(\lambda)$ is nonnegative with initial value one by construction, so it remains to show that $(M_t(\lambda))_{t=0}^{\infty}$ is a supermartingale. That is, we need to show that for every t, we have $\mathbb{E}(M_t(\lambda) \mid \mathcal{Z}_{t-1}) \leqslant M_{t-1}(\lambda)$. Writing out the conditional expectation of $M_t(\lambda)$, we have

$$\begin{aligned}
\mathbb{E}(M_t(\lambda) \mid \mathcal{Z}_{t-1}) &= \mathbb{E}\left(\prod_{i=1}^{t} \exp\left\{\lambda(Z_i - \zeta(\mu_i^\star)) - \lambda^2/8\right\} \;\middle|\; Z_1^{t-1}\right) \\
&= \underbrace{\prod_{i=1}^{t-1} \exp\left\{\lambda(Z_i - \zeta(\mu_i^\star)) - \lambda^2/8\right\}}_{M_{t-1}(\lambda)} \cdot \mathbb{E}\left(\exp\left\{\lambda(Z_t - \zeta(\mu_t^\star)) - \lambda^2/8\right\} \mid Z_1^{t-1}\right) \\
&= M_{t-1}(\lambda) \cdot \underbrace{\mathbb{E}\left(\exp\left\{\lambda(Z_t - \zeta(\mu_t^\star)) - \lambda^2/8\right\}\right)}_{(\dagger)},
\end{aligned}$$

where the last inequality follows by independence of $(Z_t)_{t=1}^{\infty}$, and hence the conditional expectation becomes a marginal expectation. Therefore, it now suffices to show that $(\dagger) \leqslant 1$. Indeed, Z_t is a $[0,1]$-bounded, mean-$\zeta(\mu_t^\star)$ random variable. By Hoeffding's sub-Gaussian inequality for bounded random variables (Hoeffding, 1963), we have that $\mathbb{E}[\exp\{\lambda(Z_t - \zeta(\mu_t^\star))\}] \leqslant \exp\{\lambda^2/8\}$, and thus

$$(\dagger) = \mathbb{E}\left[\exp\left\{\lambda(Z_t - \zeta(\mu_t^\star))\right\}\right] \cdot \exp\left\{-\lambda^2/8\right\} \leqslant 1.$$

It follows that $(M_t(\lambda))_{t=0}^{\infty}$ is an NSM.

Step 2. Let us now construct a sub-Gaussian mixture NSM. Note that the mixture of an NSM with respect to a probability distribution is itself an NSM (Robbins, 1970; Howard et al., 2020) — a straightforward consequence of Fubini's theorem. Concretely, let $f_{\rho^2}(\lambda)$ be the probability density function of a mean-zero Gaussian random variable with variance ρ^2,

$$f_{\rho^2}(\lambda) := \frac{1}{\sqrt{2\pi\rho^2}} \exp\left\{\frac{-\lambda^2}{2\rho^2}\right\}.$$

Then, since mixtures of NSMs are themselves NSMs, the process $(M_t)_{t=0}^{\infty}$ given by

$$M_t := \int_{\lambda \in \mathbb{R}} M_t(\lambda) f_{\rho^2}(\lambda) d\lambda \tag{26}$$

is an NSM. We will now find a closed-form expression for M_t. To ease notation, define the partial sum $S_t^\star := \sum_{i=1}^{t}(Z_i - \zeta(\mu_i^\star))$. Writing out the definition of M_t, we have

$$
\begin{aligned}
M_t &:= \int_{\lambda \in \mathbb{R}} \prod_{i=1}^{t} \exp\left\{\lambda(Z_i - \zeta(\mu_i^\star)) - \lambda^2/8\right\} f_{\rho^2}(\lambda) d\lambda \\
&= \int_\lambda \exp\left\{ \lambda \underbrace{\sum_{i=1}^{t}(Z_i - \zeta(\mu_i^\star))}_{S_t^\star} - t\lambda^2/8 \right\} f_{\rho^2}(\lambda) d\lambda \\
&= \int_\lambda \exp\left\{\lambda S_t^\star - t\lambda^2/8\right\} \frac{1}{\sqrt{2\pi\rho^2}} \exp\left\{\frac{-\lambda^2}{2\rho^2}\right\} d\lambda \\
&= \frac{1}{\sqrt{2\pi\rho^2}} \int_\lambda \exp\left\{\lambda S_t^\star - t\lambda^2/8\right\} \exp\left\{\frac{-\lambda^2}{2\rho^2}\right\} d\lambda \\
&= \frac{1}{\sqrt{2\pi\rho^2}} \int_\lambda \exp\left\{\lambda S_t^\star - \frac{\lambda^2(t\rho^2/4+1)}{2\rho^2}\right\} d\lambda \\
&= \frac{1}{\sqrt{2\pi\rho^2}} \int_\lambda \exp\left\{\frac{-\lambda^2(t\rho^2/4+1) + 2\lambda\rho^2 S_t^\star}{2\rho^2}\right\} d\lambda \\
&= \frac{1}{\sqrt{2\pi\rho^2}} \int_\lambda \exp\left\{\frac{-a(\lambda^2 - \frac{b}{a}2\lambda)}{2\rho^2}\right\} d\lambda,
\end{aligned}
$$

where we have set $a := t\rho^2/4 + 1$ and $b := \rho^2 S_t^\star$. Completing the square in the exponent, we have that

$$
\begin{aligned}
\exp\left\{\frac{-\lambda^2 - 2\lambda\frac{b}{a} + \left(\frac{b}{a}\right)^2 - \left(\frac{b}{a}\right)^2}{2\rho^2/a}\right\} &= \exp\left\{\frac{-(\lambda - b/a)^2}{2\rho^2/a} + \frac{a\,(b/a)^2}{2\rho^2}\right\} \\
&= \underbrace{\exp\left\{\frac{-(\lambda - b/a)^2}{2\rho^2/a}\right\}}_{(\star)} \exp\left\{\frac{b^2}{2a\rho^2}\right\}.
\end{aligned}
$$

Now notice that $(\star)$ is proportional to the density of a Gaussian random variable with mean b/a and variance ρ^2/a. Plugging the above back into the integral and multiplying the entire quantity by $a^{-1/2}/a^{-1/2}$, we obtain the closed-form expression of the mixture NSM,

$$
\begin{aligned}
M_t &:= \underbrace{\frac{1}{\sqrt{2\pi\rho^2/a}} \int_{\lambda \in \mathbb{R}} \exp\left\{\frac{-(\lambda - b/a)^2}{2\rho^2/a}\right\} d\lambda}_{=1} \frac{\exp\left\{\frac{b^2}{2a\rho^2}\right\}}{\sqrt{a}} \\
&= \frac{1}{\sqrt{t\rho^2/4+1}} \exp\left\{\frac{\rho^2(S_t^\star)^2}{2(t\rho^2/4+1)}\right\}.
\end{aligned} \tag{27}
$$

Step 3. Now that we have computed the mixture NSM $(M_t)_{t=0}^\infty$, we are ready to apply Ville's inequality and invert the process. Since $(M_t)_{t=0}^\infty$ is an NSM, we have by Ville's inequality (Ville, 1939),

$$
\mathbb{P}(\exists t : M_t \geqslant 1/\alpha) \leqslant \alpha \quad \text{or equivalently,} \quad \mathbb{P}(\forall t,\ M_t < 1/\alpha) \geqslant 1 - \alpha.
$$

Therefore, with probability at least $(1-\alpha)$, we have that for all $t \in \{1, 2, \dots\}$,

$$
\begin{aligned}
M_t < 1/\alpha &\iff \frac{1}{\sqrt{t\rho^2/4+1}} \exp\left\{ \frac{\rho^2 (S_t^\star)^2}{2(t\rho^2/4+1)} \right\} < 1/\alpha \\
&\iff \frac{\rho^2 (S_t^\star)^2}{2(t\rho^2/4+1)} - \log\left(\sqrt{t\rho^2/4+1} \right) < \log(1/\alpha) \\
&\iff \frac{\rho^2 (S_t^\star)^2}{2(t\rho^2/4+1)} < \log\left(\frac{\sqrt{t\rho^2/4+1}}{\alpha} \right) \\
&\iff (S_t^\star)^2 < \frac{2(t\rho^2/4+1)}{\rho^2} \log\left(\frac{\sqrt{t\rho^2/4+1}}{\alpha} \right) \\
&\iff \frac{(S_t^\star)^2}{t^2 r^2} < \underbrace{\frac{2(t(\rho/2)^2+1)}{(tr\rho)^2} \log\left(\frac{\sqrt{t(\rho/2)^2+1}}{\alpha} \right)}_{(\star\star)}.
\end{aligned}
$$

Set $\beta := \rho/2$ and notice that $(\star\star) = (\widetilde{B}_t^\pm)^2$ where $\widetilde{B}_t^\pm$ is the boundary given by (17) in the statement of Theorem 9. Also recall from Theorem 9 the private estimator $\widehat{\mu}_t := \frac{1}{tr}\sum_{i=1}^t [Z_i - (1-r)/2]$ and the quantity we wish to capture — the moving average of *population means* $\widetilde{\mu}_t^\star := \frac{1}{t}\sum_{i=1}^t \mu_i^\star$, where $\mu_i^\star = \mathbb{E}X_i$. Putting these together with the above high-probability bound, we have that with probability $\geqslant (1-\alpha)$, for all t,

$$
\begin{aligned}
M_t < 1/\alpha &\iff \frac{(S_t^\star)^2}{t^2 r^2} < (\widetilde{B}_t^\pm)^2 \\
&\iff -\widetilde{B}_t^\pm < \frac{S_t^\star}{tr} < \widetilde{B}_t^\pm. \\
&\iff -\widetilde{B}_t^\pm < \frac{\sum_{i=1}^t [Z_i - \zeta(\mu_i^\star)]}{tr} < \widetilde{B}_t^\pm. \\
&\iff -\widetilde{B}_t^\pm < \frac{\sum_{i=1}^t [Z_i - (r\mu_i^\star + (1-r)/2)]}{tr} < \widetilde{B}_t^\pm. \\
&\iff -\widetilde{B}_t^\pm < \frac{\sum_{i=1}^t [Z_i - (1-r)/2]}{tr} - \frac{\cancel{r}\sum_{i=1}^t \mu_i^\star}{t\cancel{r}} < \widetilde{B}_t^\pm. \\
&\iff -\frac{\sum_{i=1}^t [Z_i - (1-r)/2]}{tr} - \widetilde{B}_t^\pm < -\frac{\sum_{i=1}^t \mu_i^\star}{t} < -\frac{\sum_{i=1}^t [Z_i - (1-r)/2]}{tr} + \widetilde{B}_t^\pm. \\
&\iff -\widehat{\mu}_t - \widetilde{B}_t^\pm < -\widetilde{\mu}_t^\star < -\widehat{\mu}_t + \widetilde{B}_t^\pm. \\
&\iff \widehat{\mu}_t - \widetilde{B}_t^\pm < \widetilde{\mu}_t^\star < \widehat{\mu}_t + \widetilde{B}_t^\pm.
\end{aligned}
$$

In summary, we have that $\widetilde{C}_t^\pm := (\widehat{\mu}_t \pm \widetilde{B}_t^\pm)$ forms a $(1-\alpha)$-CS for the time-varying parameter $\widetilde{\mu}_t^\star$, meaning

$$
\mathbb{P}\left(\forall t,\ \widetilde{\mu}_t^\star \in \widetilde{C}_t^\pm \right) \geqslant 1 - \alpha.
$$

This completes the proof.

□

B. Additional results

B.1. Confidence sets under randomized response

Since `NPRR` is a strict generalization for bounded random variables, it can be used to construct confidence sets for the mean of Bernoulli random variables which are privatized via randomized response (RR). The following corollary provides a Hoeffding-type CI for the mean under RR.

Corollary 2 (Locally private Hoeffding inequality under RR)**.** *Let* $X_1, \dots, X_n \sim \text{Bernoulli}(p^\star)$*, and let* $Z_1, \dots, Z_n$ *be their privatized views according to RR for some fixed* $r \in (0, 1]$*. Then,*

$$\dot{L}_n^{\text{H}} := \frac{\sum_{i=1}^n (Z_i - (1-r)/2)}{nr} - \sqrt{\frac{\log(1/\alpha)}{2nr^2}} \tag{28}$$

is a $(1-\alpha, \varepsilon)$*-lower LPCI for* $p^\star$*, where* $\varepsilon = \log(1 + 2r/(1-r))$*.*

Corollary 2 is a special case of Theorem 4. Notice that in the non-private setting when $r = 1$, Corollary 2 recovers Hoeffding's inequality exactly (Hoeffding, 1963).

B.2. Confidence sets for sample means

While we primarily focused on deriving CIs and CSs for population means, our techniques can also be applied to the construction of CIs and CSs for the *sample mean*. Indeed, in the non-interactive case, the proof of Theorem 4 can be modified so that the bound (11) is a lower $(1-\alpha)$-CI for the sample mean $\mu^\star := \frac{1}{n}\sum_{i=1}^n x_i$, recovering essentially the same result as Ding et al. (2017, Theorem 1).[3] However, implicit in our results are also time-uniform CSs for the *running sample mean so far*. Concretely, we have the following corollary.

Corollary 3 (A confidence sequence for the running sample mean)**.** *Let* $(x_t)_{t=1}^\infty$ *be a sequence of* $[0, 1]$*-bounded numbers and let* $(Z_t)_{t=1}^\infty$ *be their privatized views according to* `NPRR` *without sequential interactivity. Then, the same bound as given in Theorem 9, i.e.*

$$\widetilde{C}_t := \left(\frac{\sum_{i=1}^t (Z_i - (1-r)/2)}{tr} \pm \sqrt{\frac{t\beta^2 + 1}{2(tr\beta)^2} \log\left(\frac{\sqrt{t\beta^2 + 1}}{\alpha} \right)} \right) \tag{29}$$

forms a $(1-\alpha, \varepsilon)$*-LPCS for the running sample mean* $\widetilde{\mu}_t^\star := \frac{1}{t}\sum_{i=1}^t x_i$*, i.e.*

$$\mathbb{P}\left(\forall t,\ \widetilde{\mu}_t^\star \in \widetilde{C}_t \right) \geqslant 1 - \alpha. \tag{30}$$

The above corollary is an immediate consequence of Theorem 9 instantiated for random variables $(X_t)_{t=1}^\infty$ with degenerate distributions. (and hence $\mathbb{E}X_t = X_t = \mu_t^\star$).

Corollary 3 also sheds some light on how the two estimands (population vs sample means) are related but fundamentally different. Both the (a) stochastic setting with data $X_1, X_2, \dots$ that have a constant mean $\mathbb{E}X_1 = \mu^\star \in [0, 1]$ and (b) nonstochastic setting with deterministic data $x_1, x_2, \dots$ are special cases of the stochastic setting with data that have time-varying means $\mathbb{E}X_t = \mu_t$ for $t \geqslant 1$. Setting (a) is recovered by assuming that $\mu_1 = \mu_2 = \cdots = \mu^\star$, while setting (b) is recovered by assuming $(X_t)_{t=1}^\infty$ have degenerate distributions (or by conditioning on them). Clearly, neither is a special case of the other, and hence we cannot expect CIs/CSs for one to work for the other in general (though in this case, $(\widetilde{C}_t)_{t=1}^\infty$ works for both).

B.3. Why one should set $G = 1$ for Hoeffding-type methods

In Section 3, we recommended setting G to the smallest possible value of 1 because Hoeffding-type bounds cannot benefit from larger values. We will now justify mathematically where this recommendation came from.

Suppose $(X_t)_{t=1}^n \sim P$ for some $\mathcal{P}_{\mu^\star}^n$ where we have chosen $r \in (0, 1]$ and an integer $G \geqslant 1$ to satisfy ε-LDP with

$$\varepsilon := \log\left(1 + \frac{(G+1)r}{1-r}\right). \tag{31}$$

Recall the `NPRR`-Hoeffding lower LPCI given (11),

$$\dot{L}_n^{\text{H}} := \frac{\sum_{i=1}^n (Z_i - (1-r)/2)}{nr} - \underbrace{\sqrt{\frac{\log(1/\alpha)}{2nr^2}}}_{\dot{B}_n^{\text{H}}}, \tag{32}$$

[3] Technically, a one-sided CI is more general than Ding et al. (2017)'s since theirs is a two-sided CI that we recover after taking a union bound over lower and upper CIs, but the lower CI is also implicit in their proof.

and take particular notice of $\dot{B}_n^{\mathrm{H}}$, the "boundary". Making this bound as sharp as possible amounts to minimizing $\dot{B}_n^{\mathrm{H}}$, which is clearly when $r = 1$ — the non-private case — but what if we want to minimize $\dot{B}_n^{\mathrm{H}}$ *subject to* ε-LDP? Given the relationship between ε, r, and G, we have that r can be written as

$$r := \frac{\exp\{\varepsilon\} - 1}{\exp\{\varepsilon\} + G}.$$

Plugging this into $\dot{B}_n^{\mathrm{H}}$, we have

$$\dot{B}_n^{\mathrm{H}} := \sqrt{\frac{\log(1/\alpha)}{2n\left(\frac{\exp\varepsilon - 1}{\exp\varepsilon + G}\right)^2}} = \left(\frac{\exp\{\varepsilon\} + G}{\exp\{\varepsilon\} - 1}\right) \cdot \sqrt{\frac{\log(1/\alpha)}{2n}},$$

which is a strictly increasing function of G. It follows that G should be set to the minimal value of 1 to make $\dot{L}_n^{\mathrm{H}}$ as sharp as possible.

B.4. Confidence sets under the sequentially interactive Laplace mechanism

Proposition 1 (`Lap-H-CS`)**.** *Suppose $(X_t)_{t=1}^{\infty} \sim P$ for some $\mathcal{P}_{\mu^\star}^{\infty}$ and let $(Z_t)_{t=1}^{\infty}$ be their privatized views according to Algorithm 1. Let $\psi_t^{\mathcal{L}}(\lambda) := -\log(1 - \lambda^2/\varepsilon_t^2)$ be the (conditional) cumulant generating function of a mean-zero Laplace random variable with scale $1/\varepsilon_t$. Let $(\lambda_t)_{t=1}^{\infty}$ be a sequence of random variables such that λ_t depends on Z_1^{t-1} — formally $\sigma(Z_1^{t-1})$-measurable — and $[0, \varepsilon_t)$-valued. Then,*

$$\bar{L}_t^{\mathcal{L}} := \frac{\sum_{i=1}^t \lambda_i Z_i}{\sum_{i=1}^t \lambda_i} - \frac{\log(1/\alpha) + \sum_{i=1}^t \left(\lambda_i^2/8 + \psi_i^{\mathcal{L}}(\lambda_i)\right)}{\sum_{i=1}^t \lambda_i} \tag{33}$$

forms a lower $(1-\alpha, (\varepsilon_t)_{t=1}^{\infty})$-LPCS for $\mu^\star$.

To obtain sharp CSs for $\mu^\star$, we recommend setting

$$\lambda_t := \sqrt{\frac{\log(1/\alpha)}{\sum_{i=1}^t (1/8 + 1/\varepsilon_i^2)\log(t+1)}} \wedge c \cdot \varepsilon_t, \tag{34}$$

for some prespecified truncation scale $c \in (0, 1)$. We choose λ_t as scaling like $1/\sqrt{t \log t}$ so that the CS $\bar{L}_t^{\mathcal{L}}$ is $O(\sqrt{\log t / t})$ up to $\log\log$ factors (see Waudby-Smith & Ramdas (2023, Table 1) for more details).[4] The constants provided in (34) arise from approximating $\psi^{\mathcal{L}}(\lambda)$ by λ^2/ε^2 for λ near 0 — an approximation that can be justified by a simple application of L'Hopital's rule — and attempting to minimize the CI width.

Similar to Section 3, we can choose $(\lambda_t)_{t=1}^{\infty}$ so that $\bar{L}_t^{\mathcal{L}}$ is tight for a fixed sample size n. Indeed, we have the following Laplace-Hoeffding CIs for $\mu^\star$.

Corollary 4 (`Lap-H`)**.** *Given the same assumptions as Proposition 1 for a fixed sample size n, define*

$$\lambda_{t,n} := \sqrt{\frac{\log(1/\alpha)}{\frac{n}{t}\sum_{i=1}^t (1/8 + 1/\varepsilon_i^2)}} \wedge c \cdot \varepsilon_t, \tag{35}$$

and plug it into $\bar{L}_t^{\mathcal{L}}$ as given above. Then,

$$\dot{L}_n^{\mathcal{L}} := \max_{1 \leqslant t \leqslant n} \bar{L}_t$$

is a $(1-\alpha, (\varepsilon_t)_t)$-lower LPCI for $\mu^\star$.

The proof of Proposition 1 (and hence Corollary 4) can be found in Section C.1. Note that any prespecified value of $c \in (0, 1)$ yields valid CSs and CIs, we find that smaller values (e.g. near 0.1) yield tighter intervals, and we set $c = 0.1$ in our simulations (Figures 4 and 5).

[4] This specific rate assumes $\varepsilon_t = \varepsilon \in (0, 1)$ for each t.

B.5. Variance-adaptive confidence intervals and sequences

B.5.1. VARIANCE-ADAPTIVE CONFIDENCE INTERVALS

Notice that if $G_t = 1$ for each t, then regardless of how low-variance $(X_t)_{t=1}^n$ are, the observations that are ultimately used for confidence set construction are still Bernoulli. In other words, it does not matter whether $(X_t)_{t=1}^n$ are Bernoulli(1/2), Uniform[0, 1], or Beta(100, 100) — with variances of roughly 0.25, 0.083, and 0.0012, respectively — the privatized observations $(Z_t)_{t=1}^n$ are all Bernoulli(1/2) with a maximal variance 0.25. Unfortunately, this means that variance-adaptive techniques cannot be used to derive tighter CIs from $(Z_t)_{t=1}^n$ directly. The story changes, however, when $G_t \geqslant 2$. Concretely, for the same value of r_t, setting G_t to be very large does not change the conditional mean of Z_t but it can substantially lower its conditional variance (e.g. if X_t has a continuous distribution, such as Beta(α, β)). Of course, given the fact that `NPRR` satisfies ε_t-LDP with $\varepsilon_t = \log\left(1 + \frac{(G_t+1)r_t}{1-r_t}\right)$, there are privacy implications to increasing G_t, and hence there is a tradeoff that must be carefully navigated when choosing (r_t, G_t) to satisfy ε_t when attempting to derive variance-adaptive CIs. We will leave that delicate discussion for later — for now, it is just important to keep in mind that larger G_t can lower the variance of $(Z_t)_{t=1}^n$, and our goal will be to exploit this fact for the sake of tighter CIs.

We will proceed by turning to the literature on nonasymptotic CIs for bounded random variables, focusing on the (super)martingale-based CIs of Waudby-Smith & Ramdas (2023) and adapting their techniques to the locally private setting. Specifically, we will derive private analogues of the product martingales outlined in Waudby-Smith & Ramdas (2023, Section 4) as well as the so-called "predictable plug-in" supermartingales of Waudby-Smith & Ramdas (2023, Section 3). As we will see in Theorem 10 and Proposition 2, the former product martingales yield tighter CIs but at the expense of a closed-form expression, while the latter supermartingales are looser (but still variance-adaptive) and are available in closed-form.

Product "betting" martingales. Beginning with the former, we follow the discussions in Waudby-Smith & Ramdas (2023, Remark 1 & Section 5.1) and set

$$\lambda_{t,n}(\mu) := \sqrt{\frac{2\log(1/\alpha)}{\widehat{\gamma}_{t-1}^2 n}} \wedge \frac{c}{\zeta_t(\mu)}, \text{ where} \tag{36}$$
$$\widehat{\gamma}_t^2 := \frac{1/4 + \sum_{i=1}^t (Z_i - \widehat{\zeta}_i)^2}{t+1}, \ \widehat{\zeta}_t := \frac{1/2 + \sum_{i=1}^t Z_i}{t+1},$$

and $c \in (0, 1)$ is some prespecified truncation scale (e.g. 1/2 or 3/4). Given the above, we have the following variance-adaptive CI for $\mu^\star$ under `NPRR`.

Theorem 10 (`NPRR-hedged`). *Suppose $(X_t)_{t=1}^n \sim P$ for some $P \in \mathcal{P}_{\mu^\star}^n$ and let $(Z_t)_{t=1}^n \sim Q$ be their `NPRR`-privatized views where $Q \in \mathcal{Q}_{\mu^\star}^n$. Define*

$$\mathcal{K}_{t,n}(\mu) := \prod_{i=1}^t \left[1 + \lambda_{i,n}(\mu) \cdot (Z_i - \zeta_i(\mu))\right] \tag{37}$$

with $\lambda_{t,n}(\mu)$ given by (36). Then, $\mathcal{K}_{t,n}(\mu)$ is a nonincreasing function of $\mu \in [0, 1]$, and $\mathcal{K}_{t,n}(\mu^\star)$ forms a $\mathcal{Q}_{\mu^\star}^n$-NM. Consequently,

$$\dot{L}_n := \max_{1 \leqslant t \leqslant n} \inf \{\mu \in [0,1] : \mathcal{K}_{t,n}(\mu) < 1/\alpha\} \tag{38}$$

forms a lower $(1-\alpha, (\varepsilon_t)_t)$-LPCI for $\mu^\star$, meaning $\mathbb{P}(\mu^\star \geqslant \dot{L}_n) \geqslant 1 - \alpha$.

The proof in Section C.3 follows a similar technique to that of Theorem 11. As is apparent in the proof, $\mathcal{K}_{t,n}(\mu^\star)$ forms a $\mathcal{Q}_{\mu^\star}^n$-NM regardless of how $\lambda_{t,n}(\mu)$ is chosen, in which case the resulting $\dot{L}_n$ would still be a bona fide lower confidence bound. However, the choice of $\lambda_{t,n}(\mu)$ given in (36) provides excellent empirical performance for the reasons discussed in (Waudby-Smith & Ramdas, 2023, Section 5.1) and guarantees that $\dot{L}_n$ is an interval (rather than a union of disjoint sets, for example). We find that Theorem 10 has the best empirical performance out of the private CIs in our paper (see Figure 4). In our simulations (Figure 4), we set $c = 0.8$, and (r, G) were chosen using the technique outlined in Section B.6.

Empirical Bernstein supermartingales. While Theorem 10 improves on Theorem 4 in terms of variance-adaptivity, the resulting bounds given in (38) are *implicit*, and hence require numerical methods (e.g. root-finding algorithms) to compute the downstream CI. The numerical operations required are both computationally efficient and straightforward to implement in code, but closed-form bounds may nevertheless be preferable for the sake of simplicity. Empirical Bernstein CIs occupy a middle ground between the Hoeffding-style CIs of Theorem 8 and the implicit CIs of Theorem 10 by being both closed-form and variance-adaptive. To this end, consider the following tuning parameters which are similar (but not identical) to (36):

$$\lambda_{t,n}^{\mathrm{EB}}(\mu) := \sqrt{\frac{2\log(1/\alpha)}{\widehat{\gamma}_{t-1}^2 n}} \wedge c, \text{ where} \tag{39}$$
$$\widehat{\gamma}_t^2 := \frac{1/4 + \sum_{i=1}^t (Z_i - \widehat{\zeta}_i)^2}{t+1}, \ \widehat{\zeta}_t := \frac{1/2 + \sum_{i=1}^t Z_i}{t+1},$$

and $c \in (0,1)$. Then, we have the following variance-adaptive empirical Bernstein CIs under NPRR.

Proposition 2 (NPRR-EB)**.** *Under the same assumptions as Theorem 10, let $(\lambda_{t,n}^{\mathrm{EB}})_{t=1}^n$ be the $[0,1)$-valued $\mathcal{Z}$-predictable sequence given in* (39) *and define*

$$\widehat{\mu}_t(\lambda_1^t) := \frac{\sum_{i=1}^t \lambda_i \cdot (Z_i - (1-r_i)/2)}{\sum_{i=1}^t r_i \lambda_i},$$
$$\bar{B}_t^{\mathrm{EB}}(\lambda_1^t) := \frac{\log(1/\alpha) + \sum_{i=1}^t 4(Z_i - \widehat{\zeta}_{i-1})^2 \psi_E(\lambda_i)}{\sum_{i=1}^t r_i \lambda_i}.$$

where $\psi_E(\lambda) := (-\log(1-\lambda) - \lambda)/4$. Then,

$$\dot{L}_t^{\mathrm{EB}} := \max_{1 \leqslant t \leqslant n} \left\{ \widehat{\mu}_t - \bar{B}_t^{\mathrm{EB}} \right\} \tag{40}$$

forms a lower $(1-\alpha, (\varepsilon_t)_t)$-LPCI for $\mu^\star$, meaning $\mathbb{P}(\mu^\star \geqslant \dot{L}_t^{\mathrm{EB}}) \geqslant 1-\alpha$.

Proposition 2 is a corollary of Proposition 3 whose proof can be found in Appendix C.5. Similar to Theorem 10, one can use any $(\lambda_{t,n}^{\mathrm{EB}})_{t=1}^n$ as long as they are predictable and $[0,1)$-valued, but we presented (39) as it tends to exhibit good empirical performance for the reasons discussed in Waudby-Smith & Ramdas (2023). As previously alluded to, the essential difference between Theorem 10 and Proposition 2 is that the former tends to be tighter in practice, while the latter has the advantage of having a computationally and analytically simple closed-form expression. In principle, the proof and techniques of Theorem 10 and Proposition 2 may be adapted to many other variance-adaptive CIs for bounded random variables, including Bentkus (2004), Audibert et al. (2007), Maurer & Pontil (2009), Orabona & Jun (2021), or other bounds in Waudby-Smith & Ramdas (2023), but we presented the aforementioned two for simplicity and illustration. Let us now turn our attention to a more challenging but related problem of constructing time-uniform *confidence sequences* instead of fixed-time *confidence intervals*.

B.5.2. Variance-adaptive time-uniform confidence sequences

In Section 3, we presented Hoeffding-type CSs for $\mu^\star$ under NPRR. As discussed in Section B.5.1, Hoeffding-type inequalities are not variance-adaptive. In this section, we will derive a simple-to-compute, variance-adaptive CS at the expense of a closed-form expression. Adapting the so-called "grid Kelly capital process" (GridKelly) of Waudby-Smith & Ramdas (2023, Section 5.6) to the locally private setting, consider the family of processes for each $\mu \in [0,1]$, and for any user-chosen integer $D \geqslant 2$,

$$\mathcal{K}_t^+(\mu) := \sum_{d=1}^D \prod_{i=1}^t \left[1 + \lambda_{i,d}^+ \cdot (Z_i - \zeta_i(\mu))\right],$$
$$\text{and } \mathcal{K}_t^-(\mu) := \sum_{d=1}^D \prod_{i=1}^t \left[1 - \lambda_{i,d}^- \cdot (Z_i - \zeta_i(\mu))\right],$$

where $\lambda_{i,d}^+ := \frac{d}{(D+1)\zeta_i(\mu)}$ and $\lambda_{i,d}^- := \frac{d}{(D+1)(1-\zeta_i(\mu))}$ for each i. Then we have the following locally private CSs for $\mu^\star$.

Theorem 11 (NPRR-GK-CS)**.** *Let $(Z_t)_{t=1}^\infty \sim Q$ for some $Q \in \mathcal{Q}_{\mu^\star}^\infty$ be the output of NPRR as described in Section 2. For any prespecified $\theta \in [0,1]$, define the process $(\mathcal{K}_t^{\mathrm{GK}}(\mu))_{t=0}^\infty$ given by*

$$\mathcal{K}_t^{\mathrm{GK}}(\mu) := \theta \mathcal{K}_t^+(\mu) + (1-\theta)\mathcal{K}_t^-(\mu),$$

with $\mathcal{K}_0^{\mathrm{GK}}(\mu) \equiv 1$. *Then,* $\mathcal{K}_t^{\mathrm{GK}}(\mu^\star)$ *forms a* $\mathcal{Q}_{\mu^\star}^\infty$*-NM, and*

$$\bar{C}_t^{\mathrm{GK}} := \left\{ \mu \in [0,1] : \mathcal{K}_t^{\mathrm{GK}}(\mu) < \frac{1}{\alpha} \right\}$$

forms a $(1-\alpha, (\varepsilon_t)_t)$*-LPCS for* $\mu^\star$*, meaning* $\mathbb{P}(\forall t,\ \mu^\star \in \bar{C}_t^{\mathrm{GK}}) \geqslant 1-\alpha$. *Moreover,* $\bar{C}_t^{\mathrm{GK}}$ *forms an interval almost surely.*

The proof of Theorem 11 is given in Section C.4 and follows from Ville's inequality for nonnegative supermartingales (Ville, 1939; Howard et al., 2020). If a lower or upper CS is desired, one can set $\theta = 1$ or $\theta = 0$, respectively, with $\theta = 1/2$ yielding a two-sided CS. In our simulations (Figure 5), we set $D = 30$, and (r, G) were chosen using the technique outlined in Section B.6.

In Proposition 2, we presented a closed-form empirical Bernstein CI for $\mu^\star$ under `NPRR`. Similar to the relationship between the fixed-time NPRR-Hoeffding CI (Theorem 8) and the time-uniform NPRR-Hoeffding CS (Theorem 4), Proposition 2 is a corollary of a more general closed-form empirical Bernstein *CS* instantiated at a fixed sample size. We omitted this CS from the main discussion for brevity, but provide its details here.

Proposition 3 (`NPRR-EB-CS`)**.** *Given* $(Z_t)_{t=1}^\infty \sim \mathcal{Q}_{\mu^\star}^\infty$ *and let* $\widehat{\mu}_t(\lambda_1^t)$ *and* $\bar{B}_t^{\mathrm{EB}}(\lambda_1^t)$ *be as in Proposition 2:*

$$\widehat{\mu}_t(\lambda_1^t) := \frac{\sum_{i=1}^t \lambda_i \cdot (Z_i - (1-r_i)/2)}{\sum_{i=1}^t r_i \lambda_i}, \text{ and}$$

$$\bar{B}_t^{\mathrm{EB}}(\lambda_1^t) := \frac{\log(1/\alpha) + \sum_{i=1}^t 4(Z_i - \widehat{\zeta}_{i-1})^2 \psi_E(\lambda_i)}{\sum_{i=1}^t r_i \lambda_i}.$$

where $\psi_E(\lambda) := (-\log(1-\lambda) - \lambda)/4$. *Then,*

$$\bar{L}_t^{\mathrm{EB}} := \widehat{\mu}_t(\lambda_1^t) - \bar{B}_t^{\mathrm{EB}}(\lambda_1^t) \tag{41}$$

forms a lower $(1-\alpha, (\varepsilon_t)_t)$*-LPCS for* $\mu^\star$*, meaning* $\mathbb{P}(\forall t \geqslant 1,\ \mu^\star \geqslant \bar{L}_t^{\mathrm{EB}}) \geqslant 1-\alpha$.

The proof can be found in Appendix C.5, and combines the techniques for deriving private concentration inequalities (such as in Theorem 8) with those for deriving predictable plug-in empirical Bernstein inequalities (such as in Waudby-Smith & Ramdas (2023, Theorem 2)).

Similar to Theorem 10 and Proposition 2, the proofs and techniques of Theorem 11 and Proposition 3 could potentially be adapted to many other variance-adaptive CSs for bounded random variables, including other bounds contained in Waudby-Smith & Ramdas (2023), Kuchibhotla & Zheng (2021), or Orabona & Jun (2021).

B.6. Choosing (r, G) for variance-adaptive confidence sets

Unlike Hoeffding-type bounds, it is not immediately clear how we should choose (r, G) to satisfy ε-LDP and obtain sharp confidence sets using Theorems 11 and 10, since there is no closed form bound to optimize. Nevertheless, certain heuristic calculations can be performed to choose (r, G) in a principled way.[5]

One approach is to view the raw-to-private data mapping $X \mapsto Z$ as a channel over which information is lost, and we would like to choose the mapping so that as much information is preserved as possible. We will aim to measure "information lost" by the conditional entropy $H(Z \mid X)$ and minimize a surrogate of this value.

For the sake of illustration, suppose that X has a continuous uniform distribution. This is a reasonable starting point because it captures the essence of preserving information about a continuous random variable X on a discretely supported output space $\mathcal{G} := \{0, 1/G, \ldots, G/G\}$. Then, the entropy $H(Z \mid X = x)$ conditioned on $X = x$ is given by

$$H(Z \mid X = x) := \sum_{z \in \mathcal{G}} \mathbb{P}(Z = z \mid X = x) \log_2 \mathbb{P}(Z = z \mid X = x), \tag{42}$$

and we know that by definition of `NPRR`, the conditional probability mass function of $(Z \mid X)$ is

$$\mathbb{P}(Z = z \mid X = x) = \frac{1-r}{G+1} + rG \cdot \left[\mathbb{1}(z = x^{\mathrm{ceil}})(x - x^{\mathrm{floor}}) + \mathbb{1}(z = x^{\mathrm{floor}})(x^{\mathrm{ceil}} - x) \right].$$

[5]Note that "heuristics" do not invalidate the method — no matter what (r, G) are chosen to be, ε-LDP and confidence set coverage are preserved. We are just using heuristic to choose (r, G) in a smart way for the sake of gaining power.

We will use the heuristic approximation $x - x^{\text{floor}} \approx x^{\text{ceil}} - x \approx 1/(2G)$, which would hold with equality if x were at the midpoint between x^{floor} and x^{ceil}. With this approximation in mind, we can write

$$\begin{aligned}\mathbb{P}(Z = z \mid X = x) &\approx \frac{1-r}{G+1} + rG \cdot \left[\frac{1}{2G}\mathbb{1}(z = x^{\text{ceil}} \text{ or } z = x^{\text{floor}})\right] \\ &= \frac{1-r}{G+1} + \frac{r}{2}\mathbb{1}(z = x^{\text{ceil}} \text{ or } z = x^{\text{floor}}) \end{aligned} \tag{43}$$

Given (43), we can heuristically compute $H(Z \mid X = x)$ because for exactly two terms in the sum $\sum_{z \in \mathcal{G}} \mathbb{P}(Z = z \mid X = x) \log_2 \mathbb{P}(Z = z \mid X = x)$, we will have $\mathbb{1}(z = x^{\text{ceil}} \text{ or } z = x^{\text{floor}}) = 1$ and the other $G - 1$ terms will have the indicator set to 0. Simplifying notation slightly, let $p_1(r, G) := (1 - r)/(G + 1) + r/2$ be (43) for those whose indicator is 1, and $p_0(r, G) := (1 - r)/(G + 1)$ for those whose indicator is 0. Therefore, we can write

$$H(Z \mid X = x) \approx (G-1)p_0(r,G)\log_2 p_0(r,G) + 2p_1(r,G)\log_2 p_1(r,G). \tag{44}$$

Finally, the conditional entropy $H(Z \mid X)$ can be approximated by

$$H(Z \mid X) = \int_0^1 H(Z \mid X = x)dx \approx (G-1)p_0(r,G)\log_2 p_0(r,G) + 2p_1(r,G)\log_2 p_1(r,G), \tag{45}$$

since we assumed that X was uniform on [0, 1].

Given a fixed privacy level $\varepsilon \in (0, \infty)$, the approximation (45) gives us an objective function to minimize with respect to r (since G is completely determined by r once ε is fixed). This can be done using standard numerical minimization solvers. Once an optimal $(r_{\text{opt}}, \widetilde{G}_{\text{opt}})$ pair is determined numerically, $\widetilde{G}_{\text{opt}}$ may not be an integer (but we require $G \geqslant 1$ to be an integer for NPRR). As such, one can then choose the final G_{opt} to be $\lfloor \widetilde{G}_{\text{opt}} \rfloor$ or $\lceil \widetilde{G}_{\text{opt}} \rceil$, depending on which one minimizes $H(Z \mid X)$ while keeping ε fixed. If the numerically determined $\widetilde{G}_{\text{opt}}$ is $\leqslant 1$, then one can simply set $G_{\text{opt}} := 1$ and adjust r_{opt} accordingly.

B.7. One-sided time-varying

The following one-sided analogue of Theorem 9 can be derived via slightly different techniques; the details can be found in its proof.

Proposition 4. *Given the same setup as Theorem 9, define*

$$\widetilde{B}_t := \sqrt{\frac{t\beta^2 + 1}{2(tr\beta)^2} \log\left(1 + \frac{\sqrt{t\beta^2 + 1}}{2\alpha}\right)}. \tag{46}$$

Then, $\widetilde{L}_t := \widehat{\mu}_t - \widetilde{B}_t$ *forms a lower* $(1 - \alpha, \varepsilon)$*-LPCS for* $\widetilde{\mu}_t^\star := \frac{1}{t}\sum_{i=1}^t \mu_i^\star$*, meaning*

$$\mathbb{P}\left(\forall t,\ \widetilde{\mu}_t^\star \geqslant \widetilde{L}_t\right) \geqslant 1 - \alpha. \tag{47}$$

The proof is provided in Section C.7 and uses a one-sided sub-Gaussian mixture supermartingale technique similar to Howard et al. (2021, Proposition 6). Since $\widetilde{B}_t$ resembles $\widetilde{B}_t^{\pm}$ but with α doubled, we suggest choosing β using (18) but with $\beta_{2\alpha}(t_0)$. We display $\widetilde{L}_t$ alongside the two-sided bound $\widetilde{C}_t^{\pm}$ of Theorem 9 in Figure 6.

B.8. Private hypothesis testing and p-values

So far, we have focused on the use of *confidence sets* for statistical inference, but another closely related perspective is through the lens of hypothesis testing and p-values (and their sequential counterparts). Fortunately, we do not need any additional techniques to derive methods for testing, since they are byproducts of our previous results.

Following the nonparametric conditions[6] outlined in Section 3, suppose that $(X_t)_{t=1}^\infty \sim P$ for some $P \in \mathcal{P}_{\mu^\star}^\infty$ which are then privatized into $(Z_t)_{t=1}^\infty \sim Q \in \mathcal{Q}_{\mu^\star}^\infty$ via NPRR. The goal now — "locally private sequential testing" — is to use the

[6] The discussion that follows also applies to the parametric case.

private data $(Z_t)_{t=1}^\infty$ to test some null hypothesis $\mathcal{H}_0$. For example, to test $\mu^\star = \mu_0$, we set $\mathcal{H}_0 = \mathcal{Q}_{\mu_0}^\infty$ or to test $\mu^\star \leqslant \mu_0$, we set $\mathcal{H}_0 = \{Q \in \mathcal{Q}_\mu^\infty : \mu \leqslant \mu_0\}$.

Concretely, we are tasked with designing a binary-valued function $\bar\phi_t \equiv \bar\phi(Z_1, \dots, Z_t) \to \{0, 1\}$ with outputs of 1 and 0 being interpreted as "reject $\mathcal{H}_0$" and "fail to reject $\mathcal{H}_0$", respectively, so that

$$\sup_{Q \in \mathcal{H}_0} Q(\exists t : \bar\phi_t = 1) \leqslant \alpha. \tag{48}$$

A sequence of functions $(\bar\phi_t)_{t=1}^\infty$ satisfying (48) is known as a *level-α sequential test*. Another common tool in hypothesis testing is the p-value, which also has a sequential counterpart, known as the *anytime p-value* (Johari et al., 2017; Howard et al., 2021). We say that a sequence of p-values $(\bar p_t)_{t=1}^\infty$ is an *anytime p-value* if

$$\sup_{Q \in \mathcal{H}_0} Q(\exists t : \bar p_t \leqslant \alpha) \leqslant \alpha. \tag{49}$$

There are at least two ways to achieve (48) and (49): (a) by using CSs to reject non-intersecting null hypotheses, and (b) by explicitly deriving e-processes. We will first discuss (a) and leave (b) to Appendix B.8.2 as the discussion is more involved.

B.8.1. Private hypothesis testing using confidence sets.

The simplest and most direct way to test hypotheses using the results of this paper is to exploit the duality between CSs and sequential tests (or CIs and fixed-time tests). Suppose $(\bar C_t(\alpha))_{t=1}^\infty$ is an LDP $(1-\alpha)$-CS for $\mu^\star$, and let $\mathcal{H}_0 : \{Q \in \mathcal{Q}_\mu^\infty : \mu \in \Theta_0\}$ be a null hypothesis that we wish to test. Then, for any $\alpha \in (0, 1)$,

$$\bar\phi_t := \mathbb{1}\left(\bar C_t(\alpha) \cap \Theta_0 = \varnothing\right) \tag{50}$$

forms an LDP level-α sequential test for $\mathcal{H}_0$, meaning it satisfies (48). In particular, if $\bar C_t(\alpha)$ shrinks to a single point as $t \to \infty$, then $(\bar\phi_t)_{t=1}^\infty$ has asymptotic power one. Furthermore, $\inf\{\alpha : \bar C_t(\alpha) \cap \Theta_0 = \varnothing\}$ forms an anytime p-value for $\mathcal{H}_0$, meaning it satisfies (49).

Similarly, if $\dot C_n(\alpha)$ is a $(1-\alpha)$ CI for $\mu^\star$, then $\dot\phi_n := \mathbb{1}(\dot C_n(\alpha) \cap \Theta_0 = \varnothing)$ is a level-α test: $\sup_{Q \in \mathcal{H}_0} Q(\dot\phi_n = 1) \leqslant \alpha$, and $\dot p_n := \inf\{\alpha : \dot C_n(\alpha) \cap \Theta_0 = \varnothing\}$ is a p-value for $\mathcal{H}_0$: $\sup_{Q \in \mathcal{H}} Q(\dot p_n \leqslant \alpha) \leqslant \alpha$.

One can also derive sequential tests using so-called *e-processes* — processes that are upper-bounded by nonnegative supermartingales under a given null hypothesis. In fact, every single one of our CSs is derived by first deriving an explicit e-process. Let us now discuss how one can derive sequential tests and CSs using e-processes.

B.8.2. Testing via e-processes

To achieve (48) and (49), it is also sufficient to derive an *e-process* $(\bar E_t)_{t=1}^\infty$ — a $\mathcal{Z}$-adapted process that is upper bounded by an NSM *for every element of* $\mathcal{H}_0$. Formally, $(\bar E_t)_{t=1}^\infty$ is an e-process for $\mathcal{H}_0$ if for every $Q \in \mathcal{H}_0$, there exists a Q-NSM $(M_t^Q)_{t=1}^\infty$ such that

$$\forall t,\ \bar E_t \leqslant M_t^Q,\ \ Q\text{-almost surely.} \tag{51}$$

Here, $(M_t^Q)_{t=1}^\infty$ being a Q-NSM means that $\mathbb{E}_Q M_t^Q \leqslant M_{t-1}^Q$, and $M_0^Q \equiv 1$, and $M_t^Q \geqslant 0$, Q-almost surely. Note that these upper-bounding NSMs need not be the same, i.e. $(\bar E_t)_{t=1}^\infty$ can be upper bounded by a different Q-NSM for each $Q \in \mathcal{H}_0$.

Importantly, if $(\bar E_t)_{t=1}^\infty$ is an e-process under $\mathcal{H}_0$, then $\phi_t := \mathbb{1}(\bar E_t \geqslant 1/\alpha)$ forms a level-α sequential test satisfying (48) by applying Ville's inequality to the NSM that upper bounds $(\bar E_t)_{t=1}^\infty$:

$$\sup_{Q \in \mathcal{H}_0} Q(\exists t : \bar E_t \geqslant 1/\alpha) \leqslant \alpha. \tag{52}$$

Using the same technique it is easy to see that, $\bar p_t := 1/\bar E_t$ forms an anytime p-value satisfying (49). Similarly to Section 3, if we are only interested in inference at a fixed sample size n, we can still leverage e-processes to obtain sharp finite-sample p-values from private data by simply taking

$$\dot p_n := \min_{1 \leqslant t \leqslant n} 1/\bar E_t. \tag{53}$$

As an immediate consequence of (52), we have $\sup_{Q \in \mathcal{H}_0} Q(\dot p_n \leqslant \alpha) \leqslant \alpha$.

With all of this in mind, the question becomes: where can we find e-processes? The answer is simple: every single CS and CI in this paper was derived by first constructing an e-process under a point null, and Table B.8.2 explicitly links all of these CSs to their corresponding e-processes.[7] For more complex composite nulls however, there may exist e-processes that are not NSMs (Ramdas et al., 2021), and we touch on one such example in Proposition 5.

Confidence sequence	e-process
Theorem 8	(24)
Theorem 9	(27)
Proposition 4	(75)

A note on locally private e-values. Similar to how the p-value is the fixed-time version of an anytime p-value, the so-called *e-value* is the fixed-time version of an e-process. An e-value for a null $\mathcal{H}_0$ is a nonnegative random variable $\dot{E}$ with Q-expectation at most one, meaning $\mathbb{E}_Q(\dot{E}) \leqslant 1$ for any $Q \in \mathcal{H}_0$ (Grünwald et al., 2019; Vovk & Wang, 2021), and clearly by Markov's inequality, $1/\dot{E}$ is a p-value for $\mathcal{H}_0$. Indeed, the time-uniform property (52) for the e-process $(E_t)_{t=1}^\infty$ is *equivalent* to saying E_τ is an e-value for any stopping time τ (Howard et al., 2021, Lemma 3); (Zhao et al., 2016, Proposition 1).

Given the shared goals between e- and p-values, a natural question arises: "Should one use e-values or p-values for inference?". While p-values are the canonical measure of evidence in hypothesis testing, there are several reasons why one may prefer to work with e-values directly; some practical, and others philosophical. From a purely practical perspective, e-values make it simple to combine evidence across several studies (Grünwald et al., 2019; ter Schure & Grünwald, 2021; Vovk & Wang, 2021) or to control the false discovery rate under arbitrary dependence (Wang & Ramdas, 2022). They have also received considerable attention for philosophical reasons including how they relate testing to betting (Shafer, 2021) and connect frequentist and Bayesian notions of uncertainty (Grünwald et al., 2019; Waudby-Smith & Ramdas, 2020). While the details of these advantages are well outside the scope of this paper, they are advantages that can now be enjoyed in locally private inference using our methods.

B.9. A/B testing the weak null

As described in Section B.8, there is a close connection between CSs and sequential hypothesis tests. The lower CS $(\widetilde{L}_t^\Delta)_{t=1}^\infty$ presented in Proposition 5 is no exception, and can be used to test the weak null hypothesis, $\widetilde{\mathcal{H}}_0$: $\forall t,\ \widetilde{\Delta}_t \leqslant 0$ (see Figure 7). In words, $\widetilde{\mathcal{H}}_0$ is testing "is the new treatment as bad or worse than placebo *among the patients so far*?". Indeed, adapting (76) from the proof of Proposition 4 to the current setting, we have the following anytime p-value for the weak null under locally private online A/B tests.

Proposition 5. *Consider the same setup as Corollary 1, and let $\Phi(\cdot)$ be the cumulative distribution function of a standard Gaussian. Define for any $\beta > 0$,*

$$\widetilde{E}_t^\Delta := \frac{2}{\sqrt{t\beta^2 + 1}} \exp\left\{ \frac{2\beta^2 (S_{t,0}^\Delta)^2}{t\beta^2 + 1} \right\} \Phi\left(\frac{2\beta S_{t,0}^\Delta}{\sqrt{t\beta^2+1}} \right),$$

where $S_{t,0}^\Delta := \sum_{i=1}^t (\psi_i - (1-r)/2) - tr\frac{1/(1-\pi)}{1/\pi + 1/(1-\pi)}$ and $\beta > 0$. Then, $\widetilde{E}_t^\Delta$ forms an e-process and hence $\widetilde{p}_t^\Delta := 1/\widetilde{E}_t^\Delta$ forms an anytime p-value, and $\widetilde{\phi}_t^\Delta := \mathbb{1}(\widetilde{p}_t^\Delta \leqslant \alpha)$ forms a level-α sequential test for the weak null $\widetilde{\mathcal{H}}_0$.

The proof provided in Section C.6 relies on the simple observation that under $\widetilde{\mathcal{H}}_0$, $(\widetilde{E}_t^\Delta)_{t=1}^\infty$ is upper bounded by a nonnegative supermartingale, and is hence an "e-process". We suggest choosing $\beta > 0$ in a similar manner to Proposition 4.

B.10. Locally private adaptive online A/B testing

In Section 4, we demonstrated how our techniques can be used to conduct online A/B tests. However, those A/B tests were non-adaptive, in the sense that the propensity score $\pi \in (0, 1)$ was required to be the same constant for all individuals (e.g. in a Bernoulli experiment). In this section, we briefly describe an alternative CS that can be used to conduct *adaptive* online A/B tests, where the propensity scores $(\pi_t(X_t))_{t=1}^\infty$ can change over time in a data-dependent fashion and be a function of

[7] We do not link CIs to e-processes since all of our CIs are built using the aforementioned CSs.

some measured baseline covariates $(X_t)_{t=1}^\infty$. Note that while we will still consider private tests in the sense of the outcomes $(Y_t)_{t=1}^\infty$ being privatized, we will not be privatizing the covariates $(X_t)_{t=1}^\infty$ (though this is an interesting direction for future work).

To set the stage, suppose that $(X_1, A_1, Y_1), (X_2, A_2, Y_2), \dots$ are joint random variables such that covariates $X_t \sim p_X(\cdot)$, are drawn according to some common distribution, treatments $A_t \sim \text{Bernoulli}(\pi_t(X_t))$ are drawn from a conditional distribution π_t (called the propensity score) which can be chosen based on $(X_i, A_i, Y_i)_{i=1}^{t-1}$, and $Y_t \sim p_Y(\cdot \mid A_t, X_t)$ is drawn from a common conditional distribution.[8] In words, we have that for each subject t, covariates X_t are observed, a propensity score π_t is chosen based on all previous subjects, a binary treatment A_t is drawn with probability $\pi_t(X_t)$, and a $[0, 1]$-bounded outcome Y_t is observed based on subject t's covariates and their treatment A_t. Of course, if $\pi(X_t) \equiv \pi$ for each t, then the above setup recovers the classical (non-adaptive) A/B testing setup considered in Section 4.

Similarly to Section 4, we will construct $(1-\alpha)$-CSs for the *time-varying mean* $\widetilde{\Delta}_t := \frac{1}{t}\sum_{i=1}^t \Delta_i$ where

$$\Delta_i := \mathbb{E}\left\{\mathbb{E}(Y_i \mid X_i, A_i = 1) - \mathbb{E}(Y_i \mid X_i, A_i = 0)\right\} \tag{54}$$

is the individual treatment effect for subject i. To state our main result, we need to prepare some notation. Let $w_t^{(1)} := \frac{\mathbb{1}(A_t=1)}{\pi_t(X_t)}$ and $w_t^{(0)} := \frac{\mathbb{1}(A_t=0)}{1-\pi_t(X_t)}$ denote the inverse propensity score weights for treatment and control groups, respectively, and define the following pseudo-outcomes $\theta_t := [w_t^{(1)} Z_t - (1 - w_t^{(0)}(1 - Z_t))]/r$, and the resulting variance process

$$V_t := \frac{1}{t}\sum_{i=1}^t \left(\theta_i - \widehat{\theta}_{i-1}\right)^2, \text{ where } \widehat{\theta}_t := \left(\frac{1}{t}\sum_{i=1}^t \theta_i\right) \wedge 1 \tag{55}$$

We are now ready to state the main result of this section.

Theorem 12 (Locally private adaptive A/B estimation). *Let $S_t(\widetilde{\Delta}_t') := (\sum_{i=1}^t \theta_i - t\widetilde{\Delta}_t')/2$ for any $\widetilde{\Delta}_t' \in [0, 1]$ and define for any $\rho > 0$,*

$$\widetilde{M}_t^{\text{EB}}(\widetilde{\Delta}_t') := \left(\frac{\rho^\rho e^{-\rho}}{\Gamma(\rho) - \Gamma(\rho, \rho)}\right)\left(\frac{1}{V_t + \rho}\right) F_t(\widetilde{\Delta}_t'), \tag{56}$$

where $F_t(\widetilde{\Delta}_t') := {}_1F_1(1, V_t + \rho + 1, S_t(\widetilde{\Delta}_t') + V_t + \rho)$, and ${}_1F_1$ is Kummer's confluent hypergeometric function, and $\Gamma(\cdot, \cdot)$ is the upper incomplete gamma function. Then, when evaluated at the true $\widetilde{\Delta}_t$, we have that $\widetilde{M}_t^{\text{EB}}(\widetilde{\Delta}_t)$ forms a nonnegative supermartingale. Consequently,

$$\widetilde{L}_t^\Delta := \inf\left\{\widetilde{\Delta}_t \in [0, 1] : \widetilde{M}_t^{\text{EB}}(\widetilde{\Delta}_t) < 1/\alpha\right\} \tag{57}$$

forms a lower $(1-\alpha)$-CS for the running ATE $\widetilde{\Delta}_t$.

The proof can be found in Appendix C.8. Readers familiar with the semiparametric causal inference literature will notice that θ_t takes the form of a modified inverse-probability-weighted (IPW) influence function, and that doubly robust (also known as "augmented IPW") approaches are often superior both theoretically and empirically. In principle, the above discussion can be modified to handle doubly robust pseudo-outcomes and CSs using the ideas contained in Waudby-Smith et al. (2022, Section 2.1), but we presented the IPW-based approach instead for the sake of simplicity.

C. Proofs of additional results

C.1. Proof of Proposition 1

Proposition 1 (`Lap-H-CS`). *Suppose $(X_t)_{t=1}^\infty \sim P$ for some $\mathcal{P}_{\mu^\star}^\infty$ and let $(Z_t)_{t=1}^\infty$ be their privatized views according to Algorithm 1. Let $\psi_t^{\mathcal{L}}(\lambda) := -\log(1 - \lambda^2/\varepsilon_t^2)$ be the (conditional) cumulant generating function of a mean-zero Laplace random variable with scale $1/\varepsilon_t$. Let $(\lambda_t)_{t=1}^\infty$ be a sequence of random variables such that λ_t depends on Z_1^{t-1} — formally $\sigma(Z_1^{t-1})$-measurable — and $[0, \varepsilon_t)$-valued. Then,*

$$\bar{L}_t^{\mathcal{L}} := \frac{\sum_{i=1}^t \lambda_i Z_i}{\sum_{i=1}^t \lambda_i} - \frac{\log(1/\alpha) + \sum_{i=1}^t \left(\lambda_i^2/8 + \psi_i^{\mathcal{L}}(\lambda_i)\right)}{\sum_{i=1}^t \lambda_i} \tag{33}$$

forms a lower $(1-\alpha, (\varepsilon_t)_{t=1}^\infty)$-LPCS for $\mu^\star$.

[8] These distributional assumptions can be substantially weakened as in Waudby-Smith et al. (2022), but we present this simplified setting for the sake of exposition.

Proof. The proof proceeds in two steps. First, we construct an exponential NSM using the cumulant generating function of a Laplace distribution. Second and finally, we apply Ville's inequality to the NSM and invert it to obtain the lower CS.

Step 1. Consider the following process for any $\mu \in [0, 1]$,

$$M_t^{\mathcal{L}}(\mu) := \prod_{i=1}^{t} \exp\left\{\lambda_i(Z_i - \mu) - \lambda_i^2/8 - \psi_i^{\mathcal{L}}(\lambda_i)\right\},$$

with $M_t^{\mathcal{L}}(\mu) \equiv 1$. We claim that $(M_t^{\mathcal{L}}(\mu^\star))_{t=0}^{\infty}$ forms an NSM with respect to the private filtration $\mathcal{Z}$. Indeed, $(M_t^{\mathcal{L}}(\mu^\star))_{t=0}^{\infty}$ is nonnegative and starts at one by construction. It remains to prove that $(M_t^{\mathcal{L}}(\mu^\star))$ is a supermartingale, meaning $\mathbb{E}(M_t^{\mathcal{L}}(\mu^\star) \mid \mathcal{Z}_{t-1}) \leqslant M_{t-1}^{\mathcal{L}}(\mu^\star)$. Writing out the conditional expectation of $M_t^{\mathcal{L}}(\mu^\star)$, we have

$$\begin{aligned}
&\mathbb{E}\left(M_t^{\mathcal{L}}(\mu^\star) \mid \mathcal{Z}_{t-1}\right) \\
&= \mathbb{E}\left(\prod_{i=1}^{t} \exp\left\{\lambda_i(Z_i - \mu^\star) - \lambda_i^2/8 - \psi_i^{\mathcal{L}}(\lambda_i)\right\} \,\middle|\, \mathcal{Z}_{t-1}\right) \\
&= \underbrace{\prod_{i=1}^{t-1} \exp\left\{\lambda_i(Z_i - \mu^\star) - \lambda_i^2/8 - \psi_i^{\mathcal{L}}(\lambda_i)\right\}}_{M_{t-1}(\mu^\star)} \cdot \underbrace{\mathbb{E}\left(\exp\left\{\lambda_t(Z_t - \mu^\star) - \lambda_t^2/8 - \psi_t^{\mathcal{L}}(\lambda_t)\right\} \,\middle|\, \mathcal{Z}_{t-1}\right)}_{(\dagger)},
\end{aligned}$$

since $M_{t-1}^{\mathcal{L}}(\mu^\star)$ is $\mathcal{Z}_{t-1}$-measurable, and thus it can be written outside of the conditional expectation. It now suffices to show that $(\dagger) \leqslant 1$. To this end, note that $Z_t = X_t + \mathcal{L}_t$ where X_t is a $[0, 1]$-bounded, mean-$\mu^\star$ random variable, and $\mathcal{L}_t$ is a mean-zero Laplace random variable (conditional on $\mathcal{Z}_{t-1}$). Consequently, $\mathbb{E}(\exp\{\lambda_t X_t\} \mid \mathcal{Z}_{t-1}) \leqslant \exp\{\lambda_t^2/8\}$ by Hoeffding's inequality (Hoeffding, 1963), and $\mathbb{E}(\exp\{\lambda_t \mathcal{L}_t\} \mid \mathcal{Z}_{t-1}) = \exp\left\{\lambda_t^{\mathcal{L}}(\lambda_t)\right\}$ by definition of a Laplace random variable. Moreover, note that by design of Algorithm 1, X_t and $\mathcal{L}_t$ are conditionally independent. It follows that

$$\begin{aligned}
(\dagger) &= \mathbb{E}\left(\exp\left\{\lambda_t(Z_t - \mu^\star) - \lambda_t^2/8 - \psi_t^{\mathcal{L}}(\lambda_t)\right\} \,\middle|\, \mathcal{Z}_{t-1}\right) \\
&= \mathbb{E}\left(\exp\left\{\lambda_t(X_t - \mu^\star) + \lambda_t \mathcal{L}_t - \lambda_t^2/8 - \psi_t^{\mathcal{L}}(\lambda_t)\right\} \,\middle|\, \mathcal{Z}_{t-1}\right) \\
&= \underbrace{\mathbb{E}\left(\exp\left\{\lambda_t(X_t - \mu^\star) - \lambda_t^2/8\right\} \,\middle|\, \mathcal{Z}_{t-1}\right)}_{\leqslant 1} \cdot \underbrace{\mathbb{E}\left(\exp\left\{\lambda_t \mathcal{L}_t - \psi_t^{\mathcal{L}}(\lambda_t)\right\} \,\middle|\, \mathcal{Z}_{t-1}\right)}_{=1} \leqslant 1,
\end{aligned}$$

where the third equality follows from the conditional independence of X_t and $\mathcal{L}_t$. Therefore, $(M_t^{\mathcal{L}}(\mu^\star))_{t=0}^{\infty}$ is an NSM.

Step 2. By Ville's inequality, we have that

$$\mathbb{P}\left(\forall t,\ M_t^{\mathcal{L}}(\mu^\star) < 1/\alpha\right) \geqslant 1 - \alpha.$$

Let us rewrite the inequality $M_t^{\mathcal{L}}(\mu^\star) < 1/\alpha$ so that we obtain the desired lower CS.

$$\begin{aligned}
M_t^{\mathcal{L}}(\mu^\star) < 1/\alpha \iff &\prod_{i=1}^{t} \exp\left\{\lambda_t(Z_t - \mu^\star) - \lambda_i^2/8 - \psi_i^{\mathcal{L}}(\lambda_i)\right\} < 1/\alpha \\
\iff &\sum_{i=1}^{t} \left[\lambda_t(Z_t - \mu^\star) - \lambda_i^2/8 - \psi_i^{\mathcal{L}}(\lambda_i)\right] < \log(1/\alpha) \\
\iff &\sum_{i=1}^{t} \lambda_t Z_t - \sum_{i=1}^{t} [\lambda_i^2/8 + \psi_i^{\mathcal{L}}(\lambda_i)] - \mu^\star \sum_{i=1}^{t} \lambda_i < \log(1/\alpha) \\
\iff &\mu^\star > \frac{\sum_{i=1}^{t} \lambda_t Z_t}{\sum_{i=1}^{t} \lambda_i} - \frac{\log(1/\alpha) + \sum_{i=1}^{t}(\lambda_i^2/8 + \psi_i^{\mathcal{L}}(\lambda_i))}{\sum_{i=1}^{t} \lambda_i}.
\end{aligned}$$

In summary, the above inequality holds uniformly for all $t \in \{1, 2, \dots\}$ with probability at least $(1 - \alpha)$. In other words,

$$\frac{\sum_{i=1}^t \lambda_t Z_t}{\sum_{i=1}^t \lambda_i} - \frac{\log(1/\alpha) + \sum_{i=1}^t (\lambda_i^2/8 + \psi_i^{\mathcal{L}}(\lambda_i))}{\sum_{i=1}^t \lambda_i}$$

forms a $(1 - \alpha)$-lower CS for $\mu^\star$. An analogous upper-CS can be derived by applying the same technique to $-Z_1, -Z_2, \dots$ and their mean $-\mu^\star$. This completes the proof. □

C.2. A lemma for Theorems 10 and 11

To prove Theorems 11 and 10, we will prove a more general result (Lemma 1), and use it to instantiate both theorems as immediate consequences. The proof follows a similar technique to Waudby-Smith & Ramdas (2023) but adapted to the locally private setting.

Lemma 1. *Suppose $(X_t)_{t=1}^\infty \sim P$ for some $P \in \mathcal{P}_{\mu^\star}^\infty$ and let $(Z_t)_{t=1}^\infty$ be their* NPRR*-induced privatized views. Let $\theta_1, \dots, \theta_D \in [0, 1]$ be convex weights satisfying $\sum_{d=1}^D \theta_d = 1$ and let $(\lambda_t^{(d)})_{t=1}^\infty$ be a $\mathcal{Z}$-predictable sequence for each $d \in \{1, \dots, D\}$ such that $\lambda_t \in (-(1 - \zeta_t(\mu^\star))^{-1}, \zeta_t(\mu^\star)^{-1})$. Then the process formed by*

$$M_t := \sum_{d=1}^D \theta_d \prod_{i=1}^t (1 + \lambda_i^{(d)} \cdot (Z_i - \zeta_i(\mu^\star))) \tag{58}$$

is a nonnegative martingale starting at one. Further suppose that $(\breve{M}_t(\mu))_{t=0}^\infty$ is a process for any $\mu \in (0, 1)$ that when evaluated at $\mu^\star$, satisfies $\breve{M}_t(\mu^\star) \leqslant M_t$ almost surely for each t. Then

$$\breve{C}_t := \left\{ \mu \in (0, 1) : \breve{M}_t(\mu) < 1/\alpha \right\} \tag{59}$$

forms a $(1 - \alpha)$-CS for $\mu^\star$.

Proof. The proof proceeds in three steps. First, we will show that the product processes given by $\prod_{i=1}^t (1 + \lambda_i \cdot (Z_i - \zeta_i(\mu^\star)))$ form nonnegative martingales with respect to $\mathcal{Z}$. Second, we argue that $\sum_{d=1}^D \theta_d M_t^{(d)}$ forms a martingale for any $\mathcal{Z}$-adapted martingales. Third and finally, we argue that $\breve{C}_t$ forms a $(1 - \alpha)$-CS despite not being constructed from a martingale directly.

Step 1. We wish to show that $M_t^{(d)} := \prod_{i=1}^t (1 + \lambda_i^{(d)} \cdot (Z_i - \zeta_i(\mu^\star)))$ forms a nonnegative martingale starting at one given a fixed $d \in \{1, \dots, D\}$. Nonnegativity follows immediately from the fact that $\lambda_t \in (-[1 - \zeta_t(\mu^\star)]^{-1},\ \zeta_t(\mu^\star)^{-1})$, and $M_t^{(d)}$ begins at one by design. It remains to show that $M_t^{(d)}$ forms a martingale. To this end, consider the conditional expectation of $M_t^{(d)}$ for any $t \in \{1, 2, \dots\}$,

$$\begin{aligned}
\mathbb{E}\left(M_t^{(d)} \mid \mathcal{Z}_{t-1} \right) &= \mathbb{E}\left(\prod_{i=1}^t (1 + \lambda_i^{(d)} \cdot (Z_i - \zeta_i(\mu^\star))) \;\middle|\; \mathcal{Z}_{t-1} \right) \\
&= \underbrace{\prod_{i=1}^{t-1} (1 + \lambda_i^{(d)} \cdot (Z_i - \zeta_i(\mu^\star)))}_{M_{t-1}^{(d)}} \cdot \mathbb{E}\left(1 + \lambda_t^{(d)} \cdot (Z_t - \zeta_t(\mu^\star)) \mid \mathcal{Z}_{t-1} \right) \\
&= M_{t-1}^{(d)} \cdot \left(1 + \lambda_t^{(d)} \cdot \underbrace{[\mathbb{E}(Z_t \mid \mathcal{Z}_{t-1}) - \zeta_t(\mu^\star)]}_{=0} \right) \\
&= M_{t-1}^{(d)}.
\end{aligned}$$

Therefore, $(M_t^{(d)})_{t=0}^\infty$ forms a martingale.

Step 2. Now, suppose that $M_t^{(1)}, \dots, M_t^{(D)}$ are test martingales with respect to the private filtration $\mathcal{Z}$, and let $\theta_1, \dots, \theta_D \in [0,1]$ be convex weights, i.e. satisfying $\sum_{d=1}^D \theta_d = 1$. Then $M_t := \sum_{d=1}^D M_t^{(d)}$ also forms a martingale since

$$\begin{aligned}\mathbb{E}(M_t \mid \mathcal{Z}_{t-1}) &= \mathbb{E}\left(\sum_{d=1}^D \theta_d M_t^{(d)} \;\middle|\; \mathcal{Z}_{t-1} \right) \\ &= \sum_{d=1}^D \theta_d \mathbb{E}\left(M_t^{(d)} \mid \mathcal{Z}_{t-1} \right) \\ &= \sum_{d=1}^D \theta_d M_{t-1}^{(d)} \\ &= M_{t-1}.\end{aligned}$$

Moreover, $(M_t)_{t=0}^\infty$ starts at one since $M_0 := \sum_{d=1}^D \theta_d M_0^{(d)} = \sum_{d=1}^D \theta_d = 1$. Finally, nonnegativity follows from the fact that $\theta_1, \dots, \theta_D$ are convex and each $(M_t^{(d)})_{t=0}^\infty$ is almost-surely nonnegative. Therefore, $(M_t)_{t=0}^\infty$ is a test martingale.

Step 3. Now, suppose $(\breve{M}_t(\mu))_{t=0}^\infty$ is a process that is almost-surely upper-bounded by $(M_t)_{t=0}^\infty$. Define $\breve{C}_t := \left\{ \mu \in (0,1) : \breve{M}_t(\mu) < 1/\alpha \right\}$. Writing out the probability of $\breve{C}_t$ miscovering $\mu^\star$ for any t, we have

$$\begin{aligned}\mathbb{P}(\exists t : \mu^\star \notin \breve{C}_t) &= \mathbb{P}(\exists t : \breve{M}_t(\mu^\star) \geqslant 1/\alpha) \\ &\leqslant \mathbb{P}(\exists t : M_t \geqslant 1/\alpha) \\ &\leqslant \alpha,\end{aligned}$$

where the first inequality follows from the fact that $\breve{M}_t(\mu^\star) \leqslant M_t$ almost surely for each t, and the second follows from Ville's inequality (Ville, 1939). This completes the proof of Lemma 1. □

In fact, a more general "meta-algorithm" extension of Lemma 1 holds, following the derivation of the "Sequentially Rebalanced Portfolio" in Waudby-Smith & Ramdas (2023, Section 5.8) but we omit these details for the sake of simplicity.

C.3. Proof of Theorem 10

Theorem 10 (`NPRR-hedged`)**.** *Suppose $(X_t)_{t=1}^n \sim P$ for some $P \in \mathcal{P}_{\mu^\star}^n$ and let $(Z_t)_{t=1}^n \sim Q$ be their `NPRR`-privatized views where $Q \in \mathcal{Q}_{\mu^\star}^n$. Define*

$$\mathcal{K}_{t,n}(\mu) := \prod_{i=1}^t \left[1 + \lambda_{i,n}(\mu) \cdot (Z_i - \zeta_i(\mu))\right] \tag{37}$$

with $\lambda_{t,n}(\mu)$ given by (36). Then, $\mathcal{K}_{t,n}(\mu)$ is a nonincreasing function of $\mu \in [0,1]$, and $\mathcal{K}_{t,n}(\mu^\star)$ forms a $\mathcal{Q}_{\mu^\star}^n$-NM. Consequently,

$$\dot{L}_n := \max_{1 \leqslant t \leqslant n} \inf \left\{ \mu \in [0,1] : \mathcal{K}_{t,n}(\mu) < 1/\alpha \right\} \tag{38}$$

forms a lower $(1-\alpha, (\varepsilon_t)_t)$-LPCI for $\mu^\star$, meaning $\mathbb{P}(\mu^\star \geqslant \dot{L}_n) \geqslant 1 - \alpha$.

Proof. The proof of Theorem 10 proceeds in three steps. First, we show that $\mathcal{K}_{t,n}$ is nonincreasing and continuous in $\mu \in [0,1]$, making $\dot{L}_n$ simple to compute via line/grid search. Second, we show that $\mathcal{K}_{t,n}(\mu^\star)$ forms a $\mathcal{Q}_{\mu^\star}^\infty$-NM. Third and finally, we show that $\dot{L}_n$ is a lower CI by constructing a lower CS that yields $\dot{L}_n$ when instantiated at n.

Step 1. $\mathcal{K}_{t,n}(\mu)$ is nonincreasing and continuous. To simplify the notation that follows, write $g_{i,n}(\mu) := 1 + \lambda_{i,n}(\mu) \cdot (Z_i - \zeta_i(\mu))$ so that

$$\mathcal{K}_{t,n}(\mu) \equiv \prod_{i=1}^t g_{i,n}(\mu).$$

Now, recall the definition of $\lambda_{i,n}(\mu)$,

$$\lambda_{t,n}(\mu) := \underbrace{\sqrt{\frac{2\log(1/\alpha)}{\widehat{\gamma}_{t-1}^2 n}}}_{\eta} \wedge \frac{c}{\zeta_t(\mu)}, \text{ where}$$

$$\widehat{\gamma}_t^2 := \frac{1/4 + \sum_{i=1}^t (Z_i - \widehat{\zeta}_i)^2}{t+1}, \; \widehat{\zeta}_t := \frac{1/2 + \sum_{i=1}^t Z_i}{t+1}.$$

Notice that $\lambda_{t,n}(\mu) \equiv \eta \wedge c/\zeta_t(\mu)$ is nonnegative and does not depend on μ except through the truncation with $c/\zeta_t(\mu)$. In particular we can write $g_{i,n}(\mu)$ as

$$\begin{aligned} g_{i,n}(\mu) &\equiv 1 + \left(\eta \wedge \frac{c}{\zeta_i(\mu)}\right)(Z_i - \zeta_i(\mu)) \\ &= 1 + (\eta Z_i) \wedge \frac{cZ_i}{\zeta_i(\mu)} - \eta\zeta_i(\mu) \wedge c, \end{aligned}$$

which is a nonincreasing (and continuous) function of $\zeta_i(\mu)$. Since $\zeta_i(\mu) := r_i\mu + (1-r_i)/2$ is an increasing (and continuous) function of μ, we have that $g_{i,n}(\mu)$ is nonincreasing and continuous in μ.

Moreover, we have that $g_{i,n}(\mu) \geqslant 0$ by design, and the product of nonnegative nonincreasing functions is also nonnegative and nonincreasing, so $\mathcal{K}_{t,n} = \prod_{i=1}^t g_{i,n}(\mu)$ is nonincreasing.

Step 2. $\mathcal{K}_{t,n}(\mu^\star)$ is a $\mathcal{Q}_{\mu^\star}^\infty$-NM. Recall the definition of $\mathcal{K}_{t,n}(\mu^\star)$

$$\mathcal{K}_{t,n}(\mu^\star) := \prod_{i=1}^t [1 + \lambda_{i,n}(\mu^\star) \cdot (Z_i - \zeta_i(\mu^\star))]$$

Then by Lemma 1 with $D = 1$ and $\theta_1 = 1$, $\mathcal{K}_{t,n}(\mu^\star)$ is a $\mathcal{Q}_{\mu^\star}^n$-NM.

Step 3. $\dot{L}_n$ is a lower CI. First, note that by Lemma 1, we have that

$$C_t := \{\mu \in [0,1] : \mathcal{K}_{t,n}(\mu) < 1/\alpha\}$$

forms a $(1-\alpha)$-CS for $\mu^\star$. In particular, define

$$\bar{L}_{t,n} := \inf\{\mu \in [0,1] : \mathcal{K}_{t,n}(\mu) < 1/\alpha\}.$$

Then, $[\bar{L}_{t,n}, 1]$ forms a $(1-\alpha)$-CS for $\mu^\star$, meaning $\mathbb{P}(\forall t,\ \mu^\star \geqslant L_{t,n}) \geqslant 1-\alpha$, and hence

$$\mathbb{P}\left(\mu^\star \geqslant \max_{1\leqslant t\leqslant n} L_{t,n}\right) = \mathbb{P}\left(\mu^\star \geqslant \dot{L}_n\right) \geqslant 1-\alpha.$$

This completes the proof. □

C.4. Proof of Theorem 11

Theorem 11 (NPRR-GK-CS)**.** *Let $(Z_t)_{t=1}^\infty \sim Q$ for some $Q \in \mathcal{Q}_{\mu^\star}^\infty$ be the output of* NPRR *as described in Section 2. For any prespecified $\theta \in [0,1]$, define the process $(\mathcal{K}_t^{\mathrm{GK}}(\mu))_{t=0}^\infty$ given by*

$$\mathcal{K}_t^{\mathrm{GK}}(\mu) := \theta\mathcal{K}_t^+(\mu) + (1-\theta)\mathcal{K}_t^-(\mu),$$

with $\mathcal{K}_0^{\mathrm{GK}}(\mu) \equiv 1$. Then, $\mathcal{K}_t^{\mathrm{GK}}(\mu^\star)$ forms a $\mathcal{Q}_{\mu^\star}^\infty$-NM, and

$$\bar{C}_t^{\mathrm{GK}} := \left\{\mu \in [0,1] : \mathcal{K}_t^{\mathrm{GK}}(\mu) < \frac{1}{\alpha}\right\}$$

forms a $(1-\alpha, (\varepsilon_t)_t)$-LPCS for $\mu^\star$, meaning $\mathbb{P}(\forall t,\ \mu^\star \in \bar{C}_t^{\mathrm{GK}}) \geqslant 1-\alpha$. Moreover, $\bar{C}_t^{\mathrm{GK}}$ forms an interval almost surely.

Proof. The proof will proceed in two steps. First, we will invoke Lemma 1 to justify that $\bar{C}_t^{\mathrm{GK}}$ indeed forms a CS. Second and finally, we prove that $\bar{C}_t^{\mathrm{GK}}$ forms an interval almost surely for each $t \in \{1, 2, \dots\}$ by showing that $\mathcal{K}_t^{\mathrm{GK}}(\mu)$ is a convex function.

Step 1. $\bar{C}_t^{\mathrm{GK}}$ forms a CS. Notice that by Lemma 1, we have that $\mathcal{K}_t^+(\mu^\star)$ and $\mathcal{K}_t^-(\mu^\star)$ defined in Theorem 11 are both test martingales. Consequently, their convex combination

$$\mathcal{K}_t^{\mathrm{GK}}(\mu^\star) := \theta \mathcal{K}_t^+(\mu^\star) + (1-\theta)\mathcal{K}_t^-(\mu^\star)$$

is also a test martingale. Therefore, $\bar{C}_t^{\mathrm{GK}} := \left\{\mu \in [0,1] : \mathcal{K}_t^{\mathrm{GK}}(\mu) < 1/\alpha\right\}$ indeed forms a $(1-\alpha)$-CS.

Step 2. $\bar{C}_t^{\mathrm{GK}}$ is an interval almost surely. We will now justify that $\bar{C}_t^{\mathrm{GK}}$ forms an interval by proving that $\mathcal{K}_t^{\mathrm{GK}}(\mu)$ is a convex function of $\mu \in [0,1]$ and noting that the sublevel sets of convex functions are themselves convex.

To ease notation, define the multiplicands $g_i^+(\mu) := 1 + \lambda_{i,d}^+ \cdot (Z_i - \zeta_i(\mu))$ so that

$$\mathcal{K}_t^+(\mu) \equiv \prod_{i=1}^t g_i(\mu).$$

Rewriting $g_i(\mu)$, we have that

$$1 + \lambda_{i,d}^+ \cdot (Z_i - \zeta_i(\mu)) = 1 + \frac{d}{D+1} \cdot \left(\frac{Z_i}{r_i \mu + (1-r_i)/2} - 1 \right),$$

from which it is clear that each $g_i(\mu)$ is (a) nonnegative, (b) nonincreasing, and (c) convex in $\mu \in [0,1]$. Now, note that properties (a)–(c) are preserved under products (Waudby-Smith & Ramdas, 2023, Section A.7), meaning

$$\mathcal{K}_t^+(\mu) \equiv \prod_{i=1}^t g_i(\mu)$$

also satisfies (a)–(c).

A similar argument goes through for $\mathcal{K}_t^-(\mu)$, except that this function is nonincreasing rather than nondecreasing, but it is nevertheless nonnegative and convex. Since convexity of functions is preserved under convex combinations, we have that

$$\mathcal{K}_t^{\mathrm{GK}}(\mu) := \theta \mathcal{K}_t^+(\mu) + (1-\theta)\mathcal{K}_t^-(\mu)$$

is a convex function of $\mu \in [0,1]$.

Finally, observe that $\bar{C}_t^{\mathrm{GK}}$ is the $(1/\alpha)$-sublevel set of $\mathcal{K}_t^{\mathrm{GK}}(\mu)$ by definition, and the sublevel sets of convex functions are convex. Therefore, $\bar{C}_t^{\mathrm{GK}}$ is an interval almost surely. This completes the proof of Theorem 11.

□

C.5. Proof of Proposition 3

Proposition 3 (NPRR-EB-CS)**.** *Given $(Z_t)_{t=1}^\infty \sim \mathcal{Q}_{\mu^\star}^\infty$ and let $\widehat{\mu}_t(\lambda_1^t)$ and $\bar{B}_t^{\mathrm{EB}}(\lambda_1^t)$ be as in Proposition 2:*

$$\widehat{\mu}_t(\lambda_1^t) := \frac{\sum_{i=1}^t \lambda_i \cdot (Z_i - (1-r_i)/2)}{\sum_{i=1}^t r_i \lambda_i}, \quad \text{and}$$

$$\bar{B}_t^{\mathrm{EB}}(\lambda_1^t) := \frac{\log(1/\alpha) + \sum_{i=1}^t 4(Z_i - \widehat{\zeta}_{i-1})^2 \psi_E(\lambda_i)}{\sum_{i=1}^t r_i \lambda_i}.$$

where $\psi_E(\lambda) := (-\log(1-\lambda) - \lambda)/4$. Then,

$$\bar{L}_t^{\mathrm{EB}} := \widehat{\mu}_t(\lambda_1^t) - \bar{B}_t^{\mathrm{EB}}(\lambda_1^t) \tag{41}$$

forms a lower $(1-\alpha, (\varepsilon_t)_t)$-LPCS for $\mu^\star$, meaning $\mathbb{P}(\forall t \geqslant 1,\ \mu^\star \geqslant \bar{L}_t^{\mathrm{EB}}) \geqslant 1-\alpha$.

Proof. The proof proceeds in two steps. First, we derive a sub-exponential NSM. Second and finally, we apply Ville's inequality to the NSM and invert it to obtain $(\dot{L}_t^{\mathrm{EB}})_{t=1}^\infty$.

Step 1: Deriving a sub-exponential nonnegative supermartingale. Consider the process $(M_t^{\mathrm{EB}}(\mu^\star))_{t=1}^\infty$ given by

$$M_t^{\mathrm{EB}}(\mu^\star) := \prod_{i=1}^t \exp\left\{\lambda_i \cdot (Z_i - \zeta_i(\mu^\star)) - 4(Z_i - \widehat{\zeta}_{i-1}(\mu^\star))^2 \psi_E(\lambda_i)\right\}, \tag{60}$$

and defined as $M_0^{\mathrm{EB}}(\mu^\star) \equiv 1$. Clearly, $M_t^{\mathrm{EB}} > 0$, and hence in order to show that $(M_t^{\mathrm{EB}}(\mu^\star))_{t=1}^\infty$ is an NSM, it suffices to show that $\mathbb{E}(M_t^{\mathrm{EB}}(\mu^\star) \mid \mathcal{Z}_{t-1}) = M_{t-1}^{\mathrm{EB}}(\mu^\star)$ for each $t \geqslant 1$. To this end, we have that

$$\mathbb{E}(M_t^{\mathrm{EB}}(\mu^\star) \mid \mathcal{Z}_{t-1}) = \mathbb{E}\left(\prod_{i=1}^t \exp\left\{\lambda_i \cdot (Z_i - \zeta_i(\mu^\star)) - 4(Z_i - \widehat{\zeta}_{i-1}(\mu^\star))^2 \psi_E(\lambda_i)\right\} \mid \mathcal{Z}_{t-1}\right) \tag{61}$$

$$= M_{t-1}^{\mathrm{EB}}(\mu^\star) \underbrace{\mathbb{E}\left(\exp\left\{\lambda_t \cdot (Z_t - \zeta_t(\mu^\star)) - 4(Z_t - \widehat{\zeta}_{t-1}(\mu^\star))^2 \psi_E(\lambda_t)\right\} \mid \mathcal{Z}_{t-1}\right)}_{(\star)}, \tag{62}$$

and hence it suffices to show that $(\star) \leqslant 1$. Following the proof of Waudby-Smith & Ramdas (2023, Theorem 2), denote

$$Y_t := Z_t - \zeta_t(\mu^\star) \quad \text{and} \quad \delta_t := \widehat{\zeta}_t(\mu^\star) - \zeta_t(\mu^\star). \tag{63}$$

Note that $\mathbb{E}(Y_t \mid \mathcal{Z}_{t-1}) = 0$. and thus it suffices to prove that for any $[0, 1)$-bounded, $\mathcal{Z}_{t-1}$-measurable λ_t,

$$\mathbb{E}\left(\exp\left\{\lambda_t Y_t - 4(Y_t - \delta_{t-1})^2 \psi_E(\lambda_t)\right\} \Bigm| \mathcal{F}_{t-1}\right) \leqslant 1.$$

Indeed, in the proof of Fan et al. (2015, Proposition 4.1), $\exp\{\xi\lambda - 4\xi^2\psi_E(\lambda)\} \leqslant 1 + \xi\lambda$ for any $\lambda \in [0, 1)$ and $\xi \geqslant -1$. Setting $\xi := Y_t - \delta_{t-1} = Z_t - \widehat{\zeta}_{t-1}(\mu^\star)$,

$$\begin{aligned}
&\mathbb{E}\left(\exp\left\{\lambda_t Y_t - 4(Y_t - \delta_{t-1})^2 \psi_E(\lambda_t)\right\} \Bigm| \mathcal{Z}_{t-1}\right) \\
&= \mathbb{E}\left(\exp\left\{\lambda_t (Y_t - \delta_{t-1}) - 4(Y_t - \delta_{t-1})^2 \psi_E(\lambda_t)\right\} \mid \mathcal{Z}_{t-1}\right) \exp(\lambda_t \delta_{t-1}) \\
&\leqslant \mathbb{E}\left(1 + (Y_t - \delta_{t-1})\lambda_t \mid \mathcal{Z}_{t-1}\right) \exp(\lambda_t \delta_{t-1}) \overset{(i)}{=} \mathbb{E}\left(1 - \delta_{t-1}\lambda_t \mid \mathcal{Z}_{t-1}\right) \exp(\lambda_t \delta_{t-1}) \overset{(ii)}{\leqslant} 1,
\end{aligned}$$

where (i) follows from the fact that Y_t is conditionally mean zero, and (ii) follows from the inequality $1 - x \leqslant \exp(-x)$ for all $x \in \mathbb{R}$. This completes the proof of Step 1.

Step 2: Applying Ville's inequality and inverting. Now that we have established that $(M_t^{\mathrm{EB}}(\mu^\star))_{t=1}^\infty$ is an NSM, we have by Ville's inequality (Ville, 1939) that

$$\mathbb{P}(\exists t \geqslant 1 : M_t^{\mathrm{EB}}(\mu^\star) \geqslant 1/\alpha) \leqslant \alpha, \tag{64}$$

or equivalently, $\mathbb{P}(\forall t \geqslant 1,\ M_t^{\mathrm{EB}}(\mu^\star) < 1/\alpha) \geqslant 1-\alpha$. Consequently, we have that with probability at least $(1-\alpha)$,

$$M_t^{\mathrm{EB}}(\mu^\star) > 1/\alpha \prod_{i=1}^t \exp\left\{\lambda_i \cdot (Z_i - \zeta_i(\mu^\star)) - 4(Z_i - \widehat{\zeta}_{i-1}(\mu^\star))^2 \psi_E(\lambda_i)\right\} < 1/\alpha \tag{65}$$

$$\iff \sum_{i=1}^t \lambda_i (Z_i - \zeta_i(\mu^\star)) - \sum_{i=1}^t 4(Z_i - \widehat{\zeta}_{i-1}(\mu^\star))^2 \psi_E(\lambda_i) < \log(1/\alpha) \tag{66}$$

$$\iff \sum_{i=1}^t \lambda_i \zeta_i(\mu^\star) > \sum_{i=1}^t \lambda_i Z_i - \sum_{i=1}^t 4(Z_i - \widehat{\zeta}_{i-1}(\mu^\star))^2 \psi_E(\lambda_i) - \log(1/\alpha) \tag{67}$$

$$\iff \sum_{i=1}^t \lambda_i \left(r_i \mu^\star + \frac{1-r_i}{2} \right) > \sum_{i=1}^t \lambda_i Z_i - \sum_{i=1}^t 4(Z_i - \widehat{\zeta}_{i-1}(\mu^\star))^2 \psi_E(\lambda_i) - \log(1/\alpha) \tag{68}$$

$$\iff \mu^\star \sum_{i=1}^t \lambda_i r_i + \sum_{i=1}^t \lambda_i \frac{1-r_i}{2} > \sum_{i=1}^t \lambda_i Z_i - \sum_{i=1}^t 4(Z_i - \widehat{\zeta}_{i-1}(\mu^\star))^2 \psi_E(\lambda_i) - \log(1/\alpha) \tag{69}$$

$$\iff \mu^\star > \underbrace{\frac{\sum_{i=1}^t \lambda_i (Z_i - (1-r_i)/2)}{\sum_{i=1}^t \lambda_i r_i}}_{\widehat{\mu}_t(\lambda_1^t)} - \underbrace{\frac{\sum_{i=1}^t 4(Z_i - \widehat{\zeta}_{i-1}(\mu^\star))^2 \psi_E(\lambda_i) + \log(1/\alpha)}{\sum_{i=1}^t \lambda_i r_i}}_{\bar{B}_t(\lambda_1^t)}, \tag{70}$$

$$\tag{71}$$

and hence $\widehat{\mu}_t(\lambda_1^t) - \bar{B}_t(\lambda_1^t)$ forms a lower $(1-\alpha)$-CS. This completes the proof of Proposition 3.

□

C.6. Proof of Proposition 5

Proposition 5. *Consider the same setup as Corollary 1, and let $\Phi(\cdot)$ be the cumulative distribution function of a standard Gaussian. Define for any $\beta > 0$,*

$$\widetilde{E}_t^\Delta := \frac{2}{\sqrt{t\beta^2+1}} \exp\left\{ \frac{2\beta^2 (S_{t,0}^\Delta)^2}{t\beta^2+1} \right\} \Phi\left(\frac{2\beta S_{t,0}^\Delta}{\sqrt{t\beta^2+1}} \right),$$

where $S_{t,0}^\Delta := \sum_{i=1}^t (\psi_i - (1-r)/2) - tr\frac{1/(1-\pi)}{1/\pi + 1/(1-\pi)}$ and $\beta > 0$. Then, $\widetilde{E}_t^\Delta$ forms an e-process and hence $\widetilde{p}_t^\Delta := 1/\widetilde{E}_t^\Delta$ forms an anytime p-value, and $\widetilde{\phi}_t^\Delta := \mathbb{1}(\widetilde{p}_t^\Delta \leqslant \alpha)$ forms a level-α sequential test for the weak null $\widetilde{\mathcal{H}}_0$.

Proof. In order to show that $\widetilde{E}_t^\Delta$ is an e-process, it suffices to find an NSM that almost surely upper bounds $\widetilde{E}_t^\Delta$ for each t under the weak null $\widetilde{\mathcal{H}}_0$: $\widetilde{\Delta}_t \leqslant 0$. As such, the proof proceeds in three steps. First, we justify why the one-sided NSM (75) given by Proposition 4 is a nonincreasing function of $\widetilde{\mu}_t$. Second, we adapt the aforementioned NSM to the A/B testing setup to obtain $M_t^\Delta(\widetilde{\Delta}_t)$ and note that it is a nonincreasing function of $\widetilde{\Delta}_t$. Third and finally, we observe that $\widetilde{E}_t^\Delta := M_t^\Delta(0)$ is upper bounded by $M_t^\Delta(\widetilde{\Delta}_t)$ under the weak null, thus proving the desired result.

Step 1: The one-sided NSM (75) is nonincreasing in $\widetilde{\mu}_t$. Recall the λ-indexed process from Step 1 of the proof of Proposition 4 given by

$$M_t(\lambda) := \prod_{i=1}^t \exp\left\{ \lambda (Z_i - \zeta(\mu_i)) - \lambda^2/8 \right\},$$

which can be rewritten as

$$M_t(\lambda) := \exp\left\{ S_t(\widetilde{\mu}_t) - \lambda^2/8 \right\},$$

where $S_t(\widetilde{\mu}_t) := \sum_{i=1}^t (Z_i - (1-r)/2) - tr\widetilde{\mu}_t$ and $\widetilde{\mu}_t := \frac{1}{t}\sum_{i=1}^t \mu_t$. In particular, notice that $M_t(\lambda)$ is a nonincreasing function of $\widetilde{\mu}_t$ for any $\lambda \geqslant 0$, and hence we also have that

$$M_t(\lambda) f_{\rho^2}^+(\lambda)$$

is a nonincreasing function of $\widetilde{\mu}_t$ where $f^+_{\rho^2}(\lambda)$ is the density of a folded Gaussian distribution given in (73), by virtue of $f^+_{\rho^2}(\lambda)$ being everywhere nonnegative, and 0 for all $\lambda < 0$. Finally, by Step 2 of the proof of Proposition 4, we have that

$$\int_\lambda M_t(\lambda) f^+_{\rho^2}(\lambda) d\lambda \equiv \frac{2}{\sqrt{t\rho^2/4+1}} \exp\left\{ \frac{\rho^2 S_t(\widetilde{\mu}_t)^2}{2(t\rho^2/4+1)} \right\} \Phi\left(\frac{\rho S_t(\widetilde{\mu}_t)}{\sqrt{t\rho^2/4+1}} \right)$$

is nonincreasing in $\widetilde{\mu}_t$, and forms an NSM when evaluated at the true means $(\widetilde{\mu}^\star_t)_{t=1}^\infty$.

Step 2: Applying Step 1 to the A/B testing setup to yield $M^\Delta_t(\widetilde{\delta}_t)$. Adapting Step 1 to the setup described in Proposition 5, let $\delta_1, \delta_2, \dots \in \mathbb{R}$ and let $\widetilde{\delta}_t := \sum_{i=1}^t \delta_i$. Define the partial sum process,

$$S^\Delta_t(\widetilde{\delta}_t) := \sum_{i=1}^t (\psi_i - (1-r)/2) - rt \frac{\widetilde{\delta}_t + \frac{1}{1-\pi}}{\frac{1}{\pi} + \frac{1}{1-\pi}}$$

and the associated process,

$$M^\Delta_t(\widetilde{\delta}_t) := \frac{2}{\sqrt{t\beta^2+1}} \exp\left\{ \frac{2\beta^2 S^\Delta_t(\widetilde{\delta}_t)^2}{t\beta^2+1} \right\} \Phi\left(\frac{2\beta S^\Delta_t(\widetilde{\delta}_t)}{\sqrt{t\beta^2+1}} \right),$$

where we have substituted $\rho := 2\beta > 0$. Notice that by construction, ψ_t is a [0, 1]-bounded random variable with mean $r\frac{\widetilde{\Delta}_t + 1/(1-\pi)}{1/\pi + 1/(1-\pi)} + (1-r)/2$, so $M^\Delta_t(\widetilde{\Delta}_t)$ forms an NSM. We are now ready to invoke the main part of the proof.

Step 3: The process $\widetilde{E}^\Delta_t$ is upper-bounded by the NSM $M^\Delta_t(\widetilde{\Delta}_t)$. Define the nonnegative process $(\widetilde{E}^\Delta_t)_{t=0}^\infty$ starting at one given by

$$\widetilde{E}^\Delta_t := M^\Delta_t(0) \equiv \frac{2}{\sqrt{t\beta^2+1}} \exp\left\{ \frac{2\beta^2 S^\Delta_t(0)^2}{t\beta^2+1} \right\} \Phi\left(\frac{2\beta S^\Delta_t(0)}{\sqrt{t\beta^2+1}} \right).$$

By Steps 1 and 2, we have that $\widetilde{E}^\Delta_t \leqslant M^\Delta_t(\widetilde{\Delta}_t)$ for any $\widetilde{\Delta}_t \leqslant 0$, and since $M^\Delta_t(\widetilde{\Delta}_t)$ is an NSM, we have that $(\widetilde{E}^\Delta_t)_{t=0}^\infty$ forms an e-process for $\mathcal{H}_0$: $\widetilde{\Delta}_t \leqslant 0$. This completes the proof.

□

C.7. Proof of Proposition 4

Proposition 4. *Given the same setup as Theorem 9, define*

$$\widetilde{B}_t := \sqrt{\frac{t\beta^2+1}{2(tr\beta)^2} \log\left(1 + \frac{\sqrt{t\beta^2+1}}{2\alpha}\right)}. \quad (46)$$

Then, $\widetilde{L}_t := \widehat{\mu}_t - \widetilde{B}_t$ forms a lower $(1-\alpha, \varepsilon)$-LPCS for $\widetilde{\mu}^\star_t := \frac{1}{t}\sum_{i=1}^t \mu^\star_i$, meaning

$$\mathbb{P}\left(\forall t,\ \widetilde{\mu}^\star_t \geqslant \widetilde{L}_t\right) \geqslant 1-\alpha. \quad (47)$$

Proof. The proof begins similar to that of Theorem 9 but with a slightly modified mixing distribution, and proceeds in four steps. First, we derive a sub-Gaussian NSM indexed by a parameter $\lambda \in \mathbb{R}$ identical to that of Theorem 9. Second, we mix this NSM over λ using a folded Gaussian density, and justify why the resulting process is also an NSM. Third, we derive an implicit lower CS for $(\widetilde{\mu}^\star_t)_{t=1}^\infty$. Fourth and finally, we compute a closed-form lower bound for the implicit CS.

Step 1: Constructing the λ-indexed NSM. This is exactly the same step as Step 1 in the proof of Theorem 9, found in Section A.5. In summary, we have that for any $\lambda \in \mathbb{R}$,

$$M_t(\lambda) := \prod_{i=1}^t \exp\left\{\lambda(Z_i - \zeta(\mu^\star_i)) - \lambda^2/8\right\}, \quad (72)$$

with $M_0(\lambda) \equiv 0$ forms an NSM with respect to the private filtration $\mathcal{Z}$.

Step 2: Mixing over $\lambda \in (0, \infty)$ to obtain a mixture NSM. Let us now construct a one-sided sub-Gaussian mixture NSM. First, note that the mixture of an NSM with respect to a probability density is itself an NSM (Robbins, 1970; Howard et al., 2020) and is a simple consequence of Fubini's theorem. For our purposes, we will consider the density of a *folded Gaussian* distribution with location zero and scale ρ^2. In particular, if $\Lambda \sim N(0, \rho^2)$, let $\Lambda_+ := |\Lambda|$ be the folded Gaussian. Then Λ_+ has a probability density function $f^+_{\rho^2}(\lambda)$ given by

$$f^+_{\rho^2}(\lambda) := \mathbb{1}(\lambda > 0) \frac{2}{\sqrt{2\pi\rho^2}} \exp\left\{ \frac{-\lambda^2}{2\rho^2} \right\}. \tag{73}$$

Note that $f^+_{\rho^2}$ is simply the density of a mean-zero Gaussian with variance ρ^2, but truncated from below by zero, and multiplied by two to ensure that $f^+_{\rho^2}(\lambda)$ integrates to one.

Then, since mixtures of NSMs are themselves NSMs, the process $(M_t)_{t=0}^\infty$ given by

$$M_t := \int_\lambda M_t(\lambda) f^+_{\rho^2}(\lambda) d\lambda \tag{74}$$

is an NSM. We will now find a closed-form expression for M_t. Many of the techniques used to derive the expression for M_t are identical to Step 2 of the proof of Theorem 9, but we repeat them here for completeness. To ease notation, define the partial sum $S_t^\star := \sum_{i=1}^t (Z_i - \zeta(\mu_i^\star))$. Writing out the definition of M_t, we have

$$\begin{aligned}
M_t &:= \int_\lambda \prod_{i=1}^t \exp\left\{ \lambda (Z_i - \zeta(\mu_i^\star)) - \lambda^2/8 \right\} f^+_{\rho^2}(\lambda) d\lambda \\
&= \int_\lambda \exp\left\{ \lambda \underbrace{\sum_{i=1}^t (Z_i - \zeta(\mu_i^\star))}_{S_t^\star} - t\lambda^2/8 \right\} f^+_{\rho^2}(\lambda) d\lambda \\
&= \int_\lambda \mathbb{1}(\lambda > 0) \exp\left\{ \lambda S_t^\star - t\lambda^2/8 \right\} \frac{2}{\sqrt{2\pi\rho^2}} \exp\left\{ \frac{-\lambda^2}{2\rho^2} \right\} d\lambda \\
&= \frac{2}{\sqrt{2\pi\rho^2}} \int_\lambda \mathbb{1}(\lambda > 0) \exp\left\{ \lambda S_t^\star - t\lambda^2/8 \right\} \exp\left\{ \frac{-\lambda^2}{2\rho^2} \right\} d\lambda \\
&= \frac{2}{\sqrt{2\pi\rho^2}} \int_\lambda \mathbb{1}(\lambda > 0) \exp\left\{ \lambda S_t^\star - \frac{\lambda^2(t\rho^2/4 + 1)}{2\rho^2} \right\} d\lambda \\
&= \frac{2}{\sqrt{2\pi\rho^2}} \int_\lambda \mathbb{1}(\lambda > 0) \exp\left\{ \frac{-\lambda^2(t\rho^2/4 + 1) + 2\lambda\rho^2 S_t^\star}{2\rho^2} \right\} d\lambda \\
&= \frac{2}{\sqrt{2\pi\rho^2}} \int_\lambda \mathbb{1}(\lambda > 0) \underbrace{\exp\left\{ \frac{-a(\lambda^2 - \frac{b}{a}2\lambda)}{2\rho^2} \right\}}_{(\star)} d\lambda,
\end{aligned}$$

where we have set $a := t\rho^2/4 + 1$ and $b := \rho^2 S_t^\star$. Completing the square in $(\star)$, we have that

$$\begin{aligned}
\exp\left\{ \frac{-a(\lambda^2 - \frac{b}{a}2\lambda)}{2\rho^2} \right\} &= \exp\left\{ \frac{-\lambda^2 + 2\lambda\frac{b}{a} + \left(\frac{b}{a}\right)^2 - \left(\frac{b}{a}\right)^2}{2\rho^2/a} \right\} \\
&= \exp\left\{ \frac{-(\lambda - b/a)^2}{2\rho^2/a} + \frac{a\,(b/a)^2}{2\rho^2} \right\} \\
&= \exp\left\{ \frac{-(\lambda - b/a)^2}{2\rho^2/a} \right\} \exp\left\{ \frac{b^2}{2a\rho^2} \right\}.
\end{aligned}$$

Plugging this back into our derivation of M_t and multiplying the entire quantity by $a^{-1/2}/a^{-1/2}$, we have

$$M_t = \frac{2}{\sqrt{2\pi\rho^2}} \int_\lambda \mathbb{1}(\lambda > 0) \underbrace{\exp\left\{ \frac{-a(\lambda^2 + \frac{b}{a}2\lambda)}{2\rho^2} \right\}}_{(\star)} d\lambda$$

$$= \frac{2}{\sqrt{2\pi\rho^2}} \int_\lambda \mathbb{1}(\lambda > 0) \exp\left\{ \frac{-(\lambda - b/a)^2}{2\rho^2/a} \right\} \exp\left\{ \frac{b^2}{2a\rho^2} \right\} d\lambda$$

$$= \frac{2}{\sqrt{a}} \exp\left\{ \frac{b^2}{2a\rho^2} \right\} \underbrace{\int_\lambda \mathbb{1}(\lambda > 0) \frac{1}{\sqrt{2\pi\rho^2/a}} \exp\left\{ \frac{-(\lambda - b/a)^2}{2\rho^2/a} \right\} d\lambda}_{(\star\star)}.$$

Now, notice that $(\star\star) = \mathbb{P}(N(b/a, \rho^2/a) \geqslant 0)$, which can be rewritten as $\Phi(b/\rho\sqrt{a})$, where Φ is the CDF of a standard Gaussian. Putting this all together and plugging in $a = t\rho^2/4 + 1$ and $b = \rho^2 S_t^\star$, we have the following expression for M_t,

$$M_t = \frac{2}{\sqrt{a}} \exp\left\{ \frac{b^2}{2a\rho^2} \right\} \Phi\left(\frac{b}{\rho\sqrt{a}} \right)$$

$$= \frac{2}{\sqrt{t\rho^2/4 + 1}} \exp\left\{ \frac{\rho^4 (S_t^\star)^2}{2(t\rho^2/4 + 1)\rho^2} \right\} \Phi\left(\frac{\rho^2 S_t^\star}{\rho\sqrt{t\rho^2/4 + 1}} \right)$$

$$= \frac{2}{\sqrt{t\rho^2/4 + 1}} \exp\left\{ \frac{\rho^2 (S_t^\star)^2}{2(t\rho^2/4 + 1)} \right\} \Phi\left(\frac{\rho S_t^\star}{\sqrt{t\rho^2/4 + 1}} \right). \tag{75}$$

Step 3: Deriving a $(1-\alpha)$-lower CS $(L_t')_{t=1}^\infty$ for $(\widetilde{\mu}_t^\star)_{t=1}^\infty$. Now that we have computed the mixture NSM $(M_t)_{t=0}^\infty$, we apply Ville's inequality to it and "invert" a family of processes — one of which is M_t — to obtain an *implicit* lower CS (we will further derive an *explicit* lower CS in Step 4).

First, let $(\mu_t)_{t=1}^\infty$ be an arbitrary real-valued process — i.e. not necessarily equal to $(\mu_t^\star)_{t=1}^\infty$ — and define their running average $\widetilde{\mu}_t := \frac{1}{t}\sum_{i=1}^t \mu_i$. Define the partial sum process in terms of $(\widetilde{\mu}_t)_{t=1}^\infty$,

$$S_t(\widetilde{\mu}_t) := \sum_{i=1}^t Z_i - tr\widetilde{\mu}_t - t(1-r)/2,$$

and the resulting nonnegative process,

$$M_t(\widetilde{\mu}_t) := \frac{2}{\sqrt{t\rho^2/4 + 1}} \exp\left\{ \frac{\rho^2 S_t(\widetilde{\mu}_t)^2}{2(t\rho^2/4 + 1)} \right\} \Phi\left(\frac{\rho S_t(\widetilde{\mu}_t)}{\sqrt{t\rho^2/4 + 1}} \right). \tag{76}$$

Notice that if $(\widetilde{\mu}_t)_{t=1}^\infty = (\widetilde{\mu}_t^\star)_{t=1}^\infty$, then $S_t(\widetilde{\mu}_t^\star) = S_t^\star$ and $M_t(\widetilde{\mu}_t^\star) = M_t$ from Step 2. Importantly, $(M_t(\widetilde{\mu}_t^\star))_{t=0}^\infty$ is an NSM. Indeed, by Ville's inequality, we have

$$\mathbb{P}(\exists t : M_t(\widetilde{\mu}_t^\star) \geqslant 1/\alpha) \leqslant \alpha. \tag{77}$$

We will now "invert" this family of processes to obtain an implicit lower boundary given by

$$L_t' := \inf\{\widetilde{\mu}_t : M_t(\widetilde{\mu}_t) < 1/\alpha\}, \tag{78}$$

and justify that $(L_t')_{t=1}^\infty$ is indeed a $(1-\alpha)$-lower CS for $\widetilde{\mu}_t^\star$. Writing out the probability of miscoverage at any time t, we have

$$\mathbb{P}(\exists t : \widetilde{\mu}_t^\star < L_t') \equiv \mathbb{P}\left(\exists t : \widetilde{\mu}_t^\star < \inf_{\widetilde{\mu}_t}\{M_t(\widetilde{\mu}_t) < 1/\alpha\} \right)$$

$$= \mathbb{P}\left(\exists t : M_t(\widetilde{\mu}_t^\star) \geqslant 1/\alpha \right)$$

$$\leqslant \alpha,$$

where the last line follows from Ville's inequality applied to $(M_t(\widetilde{\mu}_t^\star))_{t=0}^\infty$. In particular, L_t' forms a $(1-\alpha)$-lower CS, meaning

$$\mathbb{P}(\forall t,\ \widetilde{\mu}_t \geqslant L_t') \geqslant 1 - \alpha.$$

Step 4: Obtaining a closed-form lower bound $(\widetilde{L}_t)_{t=1}^{\infty}$ for $(L'_t)_{t=1}^{\infty}$. The lower CS of Step 3 is simple to evaluate via line- or grid-searching, but a closed-form expression may be desirable in practice, and for this we can compute a sharp lower bound on L'_t.

First, take notice of two key facts:

(a) When $\widetilde{\mu}_t = \frac{1}{tr}\sum_{i=1}^t Z_i - (1-r)/2r$, we have that $S_t(\widetilde{\mu}_t) = 0$ and hence $M_t(\widetilde{\mu}_t) < 1$, and

(b) $S_t(\widetilde{\mu}_t)$ is a strictly decreasing function of $\widetilde{\mu}_t \leqslant \frac{1}{tr}\sum_{i=1}^t Z_i - (1-r)/2r$, and hence so is $M_t(\widetilde{\mu}_t)$.

Property (a) follows from the fact that $\Phi(0) = 1/2$, and that $\sqrt{t\rho^2/4+1} > 1$ for any $\rho > 0$. Property (b) follows from property (a) combined with the definitions of $S_t(\cdot)$,

$$S_t(\widetilde{\mu}_t) := \sum_{i=1}^t Z_i - tr\widetilde{\mu}_t - t(1-r)/2,$$

and of $M_t(\cdot)$,

$$M_t(\widetilde{\mu}_t) := \frac{2}{\sqrt{t\rho^2/4+1}} \exp\left\{\frac{\rho^2 S_t(\widetilde{\mu}_t)^2}{2(t\rho^2/4+1)}\right\} \Phi\left(\frac{\rho S_t(\widetilde{\mu}_t)}{\sqrt{t\rho^2/4+1}}\right),$$

In particular, by facts (a) and (b), the infimum in (78) must be attained when $S_t(\cdot) \geqslant 0$. That is,

$$S_t(L'_t) \geqslant 0. \tag{79}$$

Using (79) combined with the inequality $1 - \Phi(x) \leqslant \exp\{-x^2/2\}$ (a straightforward consequence of the Cramér-Chernoff technique), we have the following lower bound on $M_t(L'_t)$:

$$\begin{aligned} M_t(L'_t) &= \frac{2}{\sqrt{t\rho^2/4+1}} \exp\left\{\frac{\rho^2 S_t(L'_t)^2}{2(t\rho^2/4+1)}\right\} \Phi\left(\frac{\rho S_t(L'_t)}{\sqrt{t\rho^2/4+1}}\right) \\ &\geqslant \frac{2}{\sqrt{t\rho^2/4+1}} \exp\left\{\frac{\rho^2 S_t(L'_t)^2}{2(t\rho^2/4+1)}\right\} \left(1 - \exp\left\{-\frac{\rho^2 S_t(L'_t)^2}{2(t\rho^2/4+1)}\right\}\right) \\ &= \frac{2}{\sqrt{t\rho^2/4+1}} \left(\exp\left\{\frac{\rho^2 S_t(L'_t)^2}{2(t\rho^2/4+1)}\right\} - 1\right) \\ &=: \breve{M}_t(L'_t). \end{aligned}$$

Finally, the above lower bound on $M_t(L'_t)$ implies that $1/\alpha \geqslant M_t(L'_t) \geqslant \breve{M}_t(L'_t)$ which yields the following lower bound on L_t':

$$\begin{aligned}
\breve{M}_t(L_t') \leqslant 1/\alpha \iff & \frac{2}{\sqrt{t\rho^2/4+1}}\left(\exp\left\{\frac{\rho^2 S_t(L_t')^2}{2(t\rho^2/4+1)}\right\} - 1\right) \leqslant 1/\alpha \\
\iff & \exp\left\{\frac{\rho^2 S_t(L_t')^2}{2(t\rho^2/4+1)}\right\} \leqslant 1 + \frac{\sqrt{t\rho^2/4+1}}{2\alpha} \\
\iff & \frac{\rho^2 S_t(L_t')^2}{2(t\rho^2/4+1)} \leqslant \log\left(1 + \frac{\sqrt{t\rho^2/4+1}}{2\alpha}\right) \\
\iff & S_t(L_t') \leqslant \sqrt{\frac{2(t\rho^2/4+1)}{\rho^2}\log\left(1 + \frac{\sqrt{t\rho^2/4+1}}{2\alpha}\right)} \\
\iff & \sum_{i=1}^t Z_i - trL_t' - t(1-r)/2 \leqslant \sqrt{\frac{2(t\rho^2/4+1)}{\rho^2}\log\left(1 + \frac{\sqrt{t\rho^2/4+1}}{2\alpha}\right)} \\
\iff & trL_t' \geqslant \sum_{i=1}^t Z_i - t(1-r)/2 - \sqrt{\frac{2(t\rho^2/4+1)}{\rho^2}\log\left(1 + \frac{\sqrt{t\rho^2/4+1}}{2\alpha}\right)} \\
\iff & L_t' \geqslant \frac{\sum_{i=1}^t (Z_i - (1-r)/2)}{tr} - \sqrt{\frac{2(t\rho^2/4+1)}{(tr\rho)^2}\log\left(1 + \frac{\sqrt{t\rho^2/4+1}}{2\alpha}\right)} \\
\iff & L_t' \geqslant \underbrace{\frac{\sum_{i=1}^t (Z_i - (1-r)/2)}{tr} - \sqrt{\frac{t\beta^2+1}{2(tr\beta)^2}\log\left(1 + \frac{\sqrt{t\beta^2+1}}{2\alpha}\right)}}_{\widetilde{L}_t},
\end{aligned}$$

where we set $\rho = 2\beta$ in the right-hand side of the final inequality. This precisely yields $\widetilde{L}_t$ as given in Proposition 4, completing the proof.

□

C.8. Proof of Theorem 12

Theorem 12 (Locally private adaptive A/B estimation). *Let $S_t(\widetilde{\Delta}_t') := (\sum_{i=1}^t \theta_i - t\widetilde{\Delta}_t')/2$ for any $\widetilde{\Delta}_t' \in [0,1]$ and define for any $\rho > 0$,*

$$\widetilde{M}_t^{\mathrm{EB}}(\widetilde{\Delta}_t') := \left(\frac{\rho^\rho e^{-\rho}}{\Gamma(\rho) - \Gamma(\rho,\rho)}\right)\left(\frac{1}{V_t+\rho}\right) F_t(\widetilde{\Delta}_t'), \tag{56}$$

where $F_t(\widetilde{\Delta}_t') := {}_1F_1(1, V_t+\rho+1, S_t(\widetilde{\Delta}_t') + V_t + \rho)$, and ${}_1F_1$ is Kummer's confluent hypergeometric function, and $\Gamma(\cdot,\cdot)$ is the upper incomplete gamma function. Then, when evaluated at the true $\widetilde{\Delta}_t$, we have that $\widetilde{M}_t^{\mathrm{EB}}(\widetilde{\Delta}_t)$ forms a nonnegative supermartingale. Consequently,

$$\widetilde{L}_t^\Delta := \inf\left\{\widetilde{\Delta}_t \in [0,1] : \widetilde{M}_t^{\mathrm{EB}}(\widetilde{\Delta}_t) < 1/\alpha\right\} \tag{57}$$

forms a lower $(1-\alpha)$-CS for the running ATE $\widetilde{\Delta}_t$.

Proof. The proof proceeds in three steps and follows a similar form to the proof of Waudby-Smith et al. (2022, Theorem 2). First, we show that a collection of processes (indexed by $\lambda \in (0,1)$) each form $\mathcal{Q}_{\mu^\star}^\infty$-NSMs with respect to the private filtration $\mathcal{Z}$. Second, we mix over $\lambda \in (0,1)$ using the truncated gamma density to obtain the NSM obtained in Theorem 12. Third and finally, we "invert" the aforementioned NSM to obtain the LPCS of Theorem 12.

Step 1: Showing that M_t^λ forms an NSM. Consider the process $(M_t^\lambda)_{t=1}^\infty$ given by

$$M_t^\lambda := \prod_{i=1}^t \exp\left\{\lambda(\theta_i - \Delta_i) - \psi_E(\lambda)(\theta_i - \widehat{\theta}_{i-1})^2/4\right\}. \tag{80}$$

We will show that $(M_t^\lambda)_{t=1}^\infty$ forms an NSM. First, note that $M_0^\lambda \equiv 1$ by construction, and M_t^λ is always positive. It remains to show that M_t^λ forms a supermartingale. Writing out the conditional expectation of M_t^λ given $\mathcal{Z}_{t-1}$, we have that

$$\mathbb{E}(M_t^\lambda \mid \mathcal{Z}_{t-1}) = M_{t-1}^\lambda \underbrace{\mathbb{E}\left[\exp\left\{\lambda(\theta_t - \Delta_t) - \psi_E(\lambda)(\theta_t - \widehat{\theta}_{t-1})^2/4\right\} \mid \mathcal{Z}_{t-1}\right]}_{(\dagger)}, \tag{81}$$

and hence it suffices to prove that $(\dagger) \leqslant 1$. Denote for the sake of succinctness,

$$\xi_t := \theta_t - \Delta_t \quad \text{and} \quad \eta_t := \widehat{\theta}_{t-1} - \Delta_t,$$

and note that $\mathbb{E}(\xi_t \mid \mathcal{Z}_{t-1}) = 0$. Using the proof of Fan et al. (2015, Proposition 4.1), we have that $\exp\{b\lambda - b^2\psi_E(\lambda)\} \leqslant 1 + b\lambda$ for any $\lambda \in [0, 1)$ and $b \geqslant -1$. Noticing that $(\theta_t - \widehat{\theta}_{t-1})/2 \geqslant -1$ and setting $b := (\xi_t - \eta_t)/2 = (\theta_t - \widehat{\theta}_{t-1})/2$, we have that

$$\begin{aligned}
&\mathbb{E}\left[\exp\left\{\lambda\xi_t - \psi_E(\lambda)(\xi_t - \eta_t)^2/4\right\} \,\middle|\, \mathcal{Z}_{t-1}\right] \\
&= \mathbb{E}\left[\exp\left\{\lambda(\xi_t - \eta_t) - \psi_E(\lambda)(\xi_t - \eta_t)^2/4\right\} \mid \mathcal{Z}_{t-1}\right] \exp(\lambda\eta_t) \\
&\leqslant \mathbb{E}\left[1 + (\xi - \eta_t)\lambda \mid \mathcal{Z}_{t-1}\right] \exp(\lambda\eta_t) \\
&= \mathbb{E}\left[1 - \eta_t\lambda \mid \mathcal{Z}_{t-1}\right] \exp(\lambda\eta_t) \leqslant 1,
\end{aligned}$$

where the last line follows from the fact that ξ_t is conditionally mean zero and the inequality $1 - x \leqslant \exp(-x)$ for all $x \in \mathbb{R}$. This completes Step 1 of the proof.

Step 2: Mixing over λ using the truncated gamma density. For any distribution F on $(0, 1)$,

$$\widetilde{M}_t^{\mathrm{EB}} := \int_{\lambda \in (0,1)} M_t^\lambda \mathrm{d}F(\lambda) \tag{82}$$

forms a test supermartingale by Fubini's theorem. In particular, we will use the truncated gamma density $f(\lambda)$ given by

$$f(\lambda) = \frac{\rho^\rho e^{-\rho(1-\lambda)} (1-\lambda)^{\rho-1}}{\Gamma(\rho) - \Gamma(\rho, \rho)}, \tag{83}$$

as the mixing density. Writing out $\widetilde{M}_t^{\mathrm{EB}} \equiv \widetilde{M}_t^{\mathrm{EB}}(\widetilde{\Delta}_t)$ using $dF(\lambda) := f(\lambda)d\lambda$, and using the shorthand $S_t \equiv S_t(\widetilde{\Delta}_t)$, we have

$$\begin{aligned}
\widetilde{M}_t^{\mathrm{EB}} &:= \int_0^1 \exp\left\{\lambda S_t - V_t\psi_E(\lambda)\right\} f(\lambda)\mathrm{d}\lambda \\
&= \int_0^1 \exp\left\{\lambda S_t - V_t\psi_E(\lambda)\right\} \frac{\rho^\rho e^{-\rho(1-\lambda)} (1-\lambda)^{\rho-1}}{\Gamma(\rho) - \Gamma(\rho, \rho)} \mathrm{d}\lambda \\
&= \frac{\rho^\rho e^{-\rho}}{\Gamma(\rho) - \Gamma(\rho, \rho)} \int_0^1 \exp\{\lambda\left(\rho + S_t + V_t\right)\} (1-\lambda)^{V_t+\rho-1} \,\mathrm{d}\lambda \\
&= \left(\frac{\rho^\rho e^{-\rho}}{\Gamma(\rho) - \Gamma(\rho, \rho)}\right) \left(\frac{1}{V_t + \rho}\right) \left(\frac{\Gamma(b)}{\Gamma(a)\Gamma(b-a)} \int_0^1 e^{zu} u^{a-1} (1-u)^{b-a-1} \mathrm{d}u\right)\Bigg|_{\substack{a=1 \\ b=V_t+\rho+1 \\ z=S_t+V_t+\rho}} \\
&= \left(\frac{\rho^\rho e^{-\rho}}{\Gamma(\rho) - \Gamma(\rho, \rho)}\right) \left(\frac{1}{V_t + \rho}\right) {}_1F_1(1, V_t + \rho + 1, S_t + V_t + \rho),
\end{aligned}$$

which completes this step.

Step 3: Applying Ville's inequality and inverting the mixture NSM. Notice that $\widetilde{\Delta}_t < \widetilde{L}_t^\Delta$ if and only if $\widetilde{M}_t(\widetilde{\Delta}_t) \geqslant 1/\alpha$, and hence by Ville's inequality for nonnegative supermartingales (Ville, 1939), we have that

$$\mathbb{P}(\exists t : \widetilde{\Delta}_t < \widetilde{L}_t^\Delta) = \mathbb{P}(\exists t : \widetilde{M}_t^{\mathrm{EB}} \geqslant 1/\alpha) \leqslant \alpha,$$

and hence $\widetilde{L}_t^\Delta$ forms a lower $(1 - \alpha, \varepsilon)$-LPCS for $\widetilde{\Delta}_t$. This completes the proof. □

D. A more detailed survey of related work

There is a rich literature exploring the intersection of statistics and differential privacy. Wasserman & Zhou (2010) studied DP estimation rates under various metrics for several privacy mechanisms. Duchi et al. (2013a;b; 2018) articulated a new "locally private minimax rate" — the fastest worst-case estimation rate with respect to any estimator *and* LDP mechanism together — and studied them in several estimation problems. To accomplish this they provide locally private analogues of the famous Le Cam, Fano, and Assouad bounds that are common in the nonparametric minimax estimation literature. As an example application, Duchi et al. (2013a;b; 2018) derived minimax rates for nonparametric density estimation in Sobolev spaces, and showed that a naive application of the Laplace mechanism cannot achieve said rates, but a different carefully designed mechanism can. This study of density estimation was extended to Besov spaces by Butucea et al. (2020). Butucea & ISSARTEL (2021) employed this minimax framework to study the fundamental limits of private estimation of nonlinear functionals. Acharya et al. (2021b) extended the locally private Le Cam, Fano, and Assouad bounds to central DP setting. Duchi & Ruan (2018) developed a framework akin to Duchi et al. (2013a;b; 2018) but from the *local* minimax point of view (here, "local" refers to the type of minimax rate considered, not "local DP"). Barnes et al. (2020) studied the locally private Fisher information for parametric models. All of the aforementioned works are focused on estimation rates, rather than inference — i.e. confidence sets, p-values, and so on (though some asymptotic inference is implicitly possible in (Duchi & Ruan, 2018; Barnes et al., 2020)).

The first work to explicitly study inference under DP constraints was Vu & Slavkovic (2009), who developed asymptotically valid private hypothesis tests for some parametric problems, including Bernoulli proportion estimation, and independence testing. Several works have furthered the study of differentially private goodness-of-fit and independence testing (Wang et al., 2015; Gaboardi et al., 2016; Berrett & Butucea, 2020; Amin et al., 2020; Acharya et al., 2020a;b; 2021a; 2022). Couch et al. (2019) develop nonparametric tests of independence between categorical and continuous variables. Awan & Slavković (2018) derive private uniformly most powerful nonasymptotic hypothesis tests in the binomial case. Karwa & Vadhan (2018), Gaboardi et al. (2019), and Joseph et al. (2019) study nonasymptotic CIs for the mean of Gaussian random variables. Canonne et al. (2019) study optimal private tests for simple nulls against simple alternatives. Covington et al. (2021) derive nonasymptotic CIs for parameters of location-scale families. Ferrando et al. (2022) introduces a parametric bootstrap method for deriving asymptotically valid CIs.

All of the previously mentioned works either consider goodness-of-fit testing, independence testing, or parametric problems where distributions are known up to some finite-dimensional parameter. Drechsler et al. (2021) study nonparametric CIs for medians. To the best of our knowledge, no prior work derives private nonasymptotic CIs (nor CSs) for means of bounded random variables.

Moreover, like most of the statistics literature, the prior work on private statistical inference is non-sequential, with the exception of Wang et al. (2022) who study private analogues of Wald's sequential probability ratio test (Wald, 1945) for simple hypotheses, and Berrett & Yu (2021) who study locally private online changepoint detection. Another interesting paper is that of Jun & Orabona (2019, Sections 7.1 & 7.2) — the authors study online convex optimization with sub-exponential noise, but also consider applications to martingale concentration inequalities (and thus CSs) as well as locally private stochastic subgradient descent.