Private Sequential Hypothesis Testing for Statisticians: Privacy, Error Rates, and Sample Size

Wanrong Zhang
Harvard University

Yajun Mei
Georgia Institute of Technology

Rachel Cummings
Columbia University

Abstract

The *sequential hypothesis testing* problem is a class of statistical analyses where the sample size is not fixed in advance. Instead, the decision-process takes in new observations sequentially to make real-time decisions for testing an alternative hypothesis against a null hypothesis until some stopping criterion is satisfied. In many common applications of sequential hypothesis testing, the data can be highly sensitive and may require privacy protection; for example, sequential hypothesis testing is used in clinical trials, where doctors sequentially collect data from patients and must determine when to stop recruiting patients and whether the treatment is effective. The field of *differential privacy* has been developed to offer data analysis tools with strong privacy guarantees, and has been commonly applied to machine learning and statistical tasks. In this work, we study the sequential hypothesis testing problem under a slight variant of differential privacy, known as *Renyi differential privacy*. We present a new private algorithm based on Wald's Sequential Probability Ratio Test (SPRT) that also gives strong theoretical privacy guarantees. We provide theoretical analysis on statistical performance measured by Type I and Type II error as well as the expected sample size. We also empirically validate our theoretical results on several synthetic databases, showing that our algorithms also perform well in practice. Unlike previous work in private hypothesis testing that focused only on the classical fixed sample setting, our results in the sequential setting allow a conclusion to be reached much earlier, and thus saving the cost of collecting additional samples.

Proceedings of the 25[th] International Conference on Artificial Intelligence and Statistics (AISTATS) 2022, Valencia, Spain. PMLR: Volume 151. Copyright 2022 by the author(s).

1 Introduction

Hypothesis testing is a fundamental task in statistics and machine learning, and involves testing a null hypothesis H_0 against an alternative hypothesis H_1, given observed data. For the usual statistical hypothesis tests, the sample size is fixed before the data are collected, but for a sequential test we observe *streaming data*, where the total sample size depends on the data and is thus a random variable. Sequential hypothesis testing is valuable because it may enable a decision to be reached earlier than with a fixed sample size test, which is critical when waiting for additional samples is costly.

The most prominent algorithm for sequential hypothesis testing is the Sequential Probability Ratio Test (SPRT) initially developed by [Wald, 1945] for efficient testing of anti-aircraft gunnery during World War II, and later used in the design of fully sequential clinical trials [Armitage, 1950, Armitage, 1954]. This algorithm continuously monitors the log-likelihood ratio of the observed data under the alternative and under the null hypotheses, and halts as soon as this ratio takes a value that is either very large or very small, reflecting that one hypothesis is overwhelmingly more likely than the other, given the observed data. The analyst running SPRT can choose these thresholds to trade-off her desired confidence in her final decision with making decisions quickly (with respect to the number of samples). In modern day, SPRT and other techniques for sequential hypothesis testing are widely used for many real-world applications, including clinical trials and quality control [Wald, 2004, Siegmund, 2013, Whitehead, 1997, Ghosh and Sen, 1991].

Performance of a sequential testing procedure is evaluated using four main criteria: two operating characteristic (OC) functions to describe the accuracy of final decisions, and two average sample number (ASN) functions to describe how quickly a decision was reached.

The two OC criteria are the probability of Type I error, Pr[reject H_0 $|H_0$], and the probability of Type II error, Pr[accept H_0 $|H_1$]. Since the number of observations T is a random variable, the two ASN functions are the expected sample size under the null and alternative hypotheses, $\mathbb{E}_{H_0}[T]$ and $\mathbb{E}_{H_1}[T]$, respectively. [Wald and Wolfowitz, 1948] showed that the sequential probability ratio test (SPRT) is the optimal test of testing a simple null H_0 against a simple alternative H_1 when observations are assumed to be sampled i.i.d., where optimality is defined as simultaneously minimizing both $\mathbb{E}_{H_0}[T]$ and $\mathbb{E}_{H_1}[T]$ subject to constraints on Type I and Type II error probabilities.

In modern applications of sequential hypothesis testing — for example to medical clinical trials — privacy also becomes another crucial performance criterion, as the data and decisions can be highly sensitive. The field of *differential privacy* [Dwork et al., 2006] has emerged as the gold standard in private data analysis by providing algorithms with strong worst-case privacy guarantees. It is a parameterized privacy notion, where the privacy parameter ϵ allows for a smooth tradeoff between accuracy of the analysis and privacy to the individuals in the database. Informally, an algorithm is ϵ-differentially private if it ensures that any particular output of the algorithm is at most e^{ϵ} more likely when a single user's data are changed. In recent years, tools for differentially private data analysis have been deployed in practice by major organizations such as Google [Erlingsson et al., 2014], Apple [Differential Privacy Team, Apple, 2017], Microsoft [Ding et al., 2017], and the U.S. Census Bureau [Dajani et al., 2017].

In this work, we provide the first differentially private algorithm for the sequential hypothesis testing problem with theoretical guarantees on the Type I and Type II error, and the expected sample size. By focusing on the metrics most relevant to the field of statistics and its practitioners, our work may be more readily deployed in practice. One real-world application of our results is the design of statistically valid sequential experiments and clinical trials before data are collected or observed. Typically when designing sequential experiments, a scientist must develop and pre-register a well-justified protocol for making final decisions under all possible data outcomes, and no further adjustments to the protocol can be made once data collection has begun. Fully sequential design of clinical trials, as suggested by [Armitage, 1950, Armitage, 1954], where evaluation occurs after each new patient outcome was not always possible for statistical or practical reasons – e.g., it is difficult to convene a data and safety monitoring committee after each observation. With recent advancements in statistics and computing, it has become feasible to continuously monitor and evaluate every patient [Whitehead, 1997]. Modern examples of fully sequential trials include the "MADIT" clinical trial to evaluate the effect of an implanted defibrillator [DeMets, 1998] and a COVID-19 therapeutics trial intended to speed up the decision process [Harrell, 2020]. Fully sequential trials risk leaking patient's sensitive information, especially for patients with data collected shortly before the trial is halted. Our proposed private sequential test can be used for monitoring trials where privacy protection is necessary, such as those with irreversible clinical outcomes like death or severe infectious disease. It can also balance the tradeoff between small expected sample sizes for rapid decision, controlled Type I and Type II error properties, and formal privacy protections.

1.1 Our contribution

In this work, we combine tools from differential privacy with classical statistical methods for sequential hypothesis testing to develop a private version of Wald's SPRT, which we call PrivSPRT.

The most natural existing tool for privatizing Wald's SPRT is a private subroutine called AboveThresh [Dwork et al., 2009, Dwork and Roth, 2014] (also known as SparseVector). This algorithm takes in a database X and a stream of queries $q_1, q_2, \ldots$, and sequentially privately tests whether the numerical value of each query q_i evaluated on the database $q_i(X)$ is above or below a pre-specified threshold. A natural first attempt at a private version of SPRT would be to instantiate AboveThresh using the SPRT test statistic as the query and using the SPRT stopping criteria as the threshold (see Section 2.1 for more details). However, as we show in Section 4, the random noise internal to AboveThresh that is used to guarantee privacy causes extremely poor performance in terms of the relevant OC and ASN metrics. In particular, we note that while AboveThresh was designed to provide good performance with respect to high-probability finite-sample performance guarantees that are commonly used in the computer science literature, it fails to provide good performance on the metrics that are most relevant to the statistics community, such as Type I and Type II error.

We instead build our algorithm PrivSPRT using a generalized version of AboveThresh from [Zhu and Wang, 2020] instantiated with Gaussian noise (rather than Laplace as in [Dwork et al., 2009]), and we show that this modification results in good performance in terms of the OC and ASN metrics of interest. Specifically, we give bounds on the expected sample size of PrivSPRT (Theorem 8) and the Type I and Type II error (Theorem 9). We ana-

lyze the privacy of PrivSPRT through a generalization of DP known as *Renyi differential privacy* (RDP) [Mironov, 2017], which is often preferred in practice due to its tighter composition properties with Gaussian noise [Wang et al., 2019]. We show that PrivSPRT satisfies RDP (Theorem 7), which also implies that is satisfies DP (Theorem 4). Finally, we perform experiments to empirically validate our theoretical findings (Section 4).

1.2 Related work

Background on non-private SPRT was presented earlier in this section, so we focus our attention here on private hypothesis testing. Private (fixed-sample-size) hypothesis testing has previously been considered in the static setting, where the analyst wishes to test a hypothesis (or family of hypothesis) at a single point in time for a fixed database [Gaboardi et al., 2016, Gaboardi and Rogers, 2018, Sheffet, 2018, Couch et al., 2019, Canonne et al., 2019]. Dynamic or online private sequential decision making has recently gained traction in various settings, including recent work on private sequential change-point detection [Cummings et al., 2018, Cummings et al., 2020, Zhang et al., 2021]. These works all rely on the AboveThresh/SparseVector technique to achieve privacy in sequential change-point problems, where the focus is on the privacy of parameter estimation of change-point. Our work deals with the sequential hypothesis testing problem which is essentially a classification problem, and our aim is to provide a unifying approach by showing that a generalization of this technique can be applied to solve general private sequential hypothesis testing problems for a more general class of accuracy objectives. [Wang et al., 2020] considers privatization of SPRT. Their algorithm is to add Laplace noise to the thresholds to generate a noisy stopping time, and then use exponential mechanism to output the binary decision. They show that the algorithm can provide a weaker notion of privacy that is data dependent, and it will only converge to DP when the stopping time goes to ∞. In contrast, our results aim to minimize stopping time, and therefore, a direct comparison would not be applicable.

2 Preliminaries

This section provides the background on sequential hypothesis testing (Section 2.1) and the differentially private tools (Section 2.2) that will be brought to bear in our PrivSPRT algorithm.

2.1 Sequential hypothesis testing

A sequence X of data points, $x_1, x_2, \cdots$, are observed sequentially, i.e., arriving one at a time. Let $f_t(x_1, \ldots, x_t)$ denote the true joint probability density function (pdf) of the first t observations, $(x_1, x_2, \ldots, x_t)$. Under the simplest model where the data points are sampled i.i.d. from some distribution f, then $f_t(x_1, \ldots, x_t) = \prod_{i=1}^{t} f(x_i)$. In more general dependence models, $f_t(x_1, \ldots, x_t) = \prod_{i=1}^{t} f(x_i|x_1, \cdots, x_{i-1})$.

In *sequential hypothesis testing* problems, the analyst has two possible hypotheses on the pdfs – f_{0t} and f_{1t} – and her goal is to quickly (i.e., with as few samples as possible) and correctly test the null hypotheses $H_0 : f_t = f_{0t}$ against the alternative $H_1 : f_t = f_{1t}$.[1] At each time t, the analyst must make one of the following three decisions: (1) halt collecting observations and *accept* the null hypothesis H_0, (2) halt collecting observations and *reject* the null hypothesis H_0, or (3) continue collecting observations to provide additional information.

There are four main criteria to assess the performance of sequential tests, including two operating characteristic (OC) functions and two average sample number (ASN) functions [Wald, 2004, Siegmund, 2013]. The two OC functions are Type I error, $\Pr_0[\text{reject } H_0]$ (i.e., rejecting H_0 when H_0 is true), and Type II error, $\Pr_1[\text{accept } H_0]$ (i.e., accepting H_0 when H_1 is true), which address correct decision-making, and are well-studied in the standard classification or hypothesis testing contexts. The two ASN functions are the expected sample size under both the null and alternative hypotheses, i.e., $\mathbb{E}_0[T]$ and $\mathbb{E}_1[T]$, which ensure that decisions are made efficiently and that unnecessary costs are not incurred by collecting too many samples. In sequential hypothesis testing problems, the objective is to simultaneously minimize $\mathbb{E}_0[T]$ and $\mathbb{E}_1[T]$ subject to the constraints that Type I and Type II error probabilities are both small.

Wald's sequential probability ratio test (SPRT) [Wald, 1945] is a celebrated optimal solution when testing a simple null H_0 , where the joint distribution is completely specified, against a simple alternative H_1 under the simplest i.i.d. model, where the data are independent and identically distributed. The idea behind SPRT is straightforward: the analyst continues to collect observations until she has enough evidence to confidently decide whether H_0 or H_1 is true, as measured by the cumulative log-

[1] For simplicity in the remainder of this paper, we will abuse notation to use the subscripts 0 and 1 to indicate probability with respect to the distributions given in H_0 and H_1, respectively.

likelihood ratio statistic being either too large or too small. Mathematically, at each time t, the analyst calculates the cumulative log-likelihood ratio statistic: $\ell_t = \log \frac{f_{1t}(x_1,\ldots,x_t)}{f_{0t}(x_1,\ldots,x_t)}$. Under the i.i.d. model, this test statistic becomes: $\ell_t = \sum_{i=1}^{t} \log \frac{f_1(x_i)}{f_0(x_i)}$. Moreover, the analyst chooses two positive constants a, b, and runs the SPRT test until the following stopping time is reached: $T = \min\{t \geq 1 : \ell_t \notin (-a, b)\}$. After reaching the stopping criterion, a statistical decision is made based on the following rule:

$$\begin{aligned} \text{Reject } H_0 \quad & \text{if } \ell_T \geq b, \\ \text{Accept } H_0 \quad & \text{if } \ell_T \leq -a. \end{aligned}$$

Intuitively, the set $(-a, b)$ is the range of test statistics where the analyst is uncertain between H_0 and H_1. If the test statistic ever falls outside of this range, then the analyst can have high confidence about one of the hypotheses being true. Under the i.i.d. model, the SPRT is exactly optimal in the sense of minimizing both expected sample sizes, $\mathbb{E}_0[T]$ and $\mathbb{E}_1[T]$, simultaneously, among all other (sequential or fixed-sample size) tests whose Type I and Type II error probabilities are same as (or smaller than) those of the SPRT [Wald and Wolfowitz, 1948]. Below we denote $x(a) \approxeq y(a)$ if $x(a)/y(a) \to 1$ if $a \to \infty$ (or if $a \to 0$).

Theorem 1 (Error Rates [Wald, 1945]). *The approximation of Type I error of SPRT is* $\Pr_0[\ell_T \geq b] \approxeq \frac{1-exp(-a)}{\exp(b)-\exp(-a)}$, *and the approximation of the Type II error of SPRT is* $\Pr_1[\ell_T \leq -a] \approxeq \exp(-a)\frac{\exp(b)-1}{\exp(b)-\exp(-a)}$.

The additional assumption that the observations x_t are independent and identically distributed is required to give the expected sample size.

Theorem 2 (Expected Sample Size [Wald, 1945]). *When $x_1, x_2, \ldots$ are sampled i.i.d., SPRT has expected samples sizes:*

$$\mathbb{E}_1[T] \approxeq \frac{-a\exp(-a)(\exp(b)-1) + b\exp(b)(1-\exp(-a))}{D_{KL}(f_1||f_0)(\exp(b)-\exp(-a))}, \tag{1}$$

$$\mathbb{E}_0[T] \approxeq \frac{-a(\exp(b)-1) + b(1-\exp(-a))}{-D_{KL}(f_0||f_1)(\exp(b)-\exp(-a))}. \tag{2}$$

2.2 Differential privacy

Differential privacy is a statistical notion of database privacy, which ensures that the output of an algorithm will still have approximately the same distribution is a single data entry were to be changed. Differential privacy considers a general database space $\mathcal{D}$. If databases are real-valued and contain a fixed number n of entries, then $\mathcal{D} = \mathbb{R}^n$; in our sequential hypothesis testing setting, our database will be of a random size so $\mathcal{D} = \mathbb{R}^*$. Two databases $X, X' \in \mathcal{D}$ are said to be *neighboring* if they differ in at most one entry.

Definition 1 (Differential Privacy [Dwork et al., 2006]). *A randomized algorithm $\mathcal{M} : \mathcal{D} \to \mathcal{R}$ is (ϵ, δ)-differentially private if for every pair of neighboring databases $X, X' \in \mathcal{D}$, and for every subset of possible outputs $\mathcal{S} \subseteq \mathcal{R}$,* $\Pr[\mathcal{M}(X) \in \mathcal{S}] \leq \exp(\epsilon)\Pr[\mathcal{M}(X') \in \mathcal{S}] + \delta$.

Renyi differential privacy (RDP) is a relaxation of differential privacy based on the Renyi divergence, defined as $D_\alpha(P||Q) = \frac{1}{\alpha-1}\log \mathbb{E}_Q\left(\frac{P(x)}{Q(x)}\right)^\alpha$. This privacy notion requires that the distribution over outputs on two neighboring databases is close in Renyi divergence.

Definition 2 (Renyi Differential Privacy [Mironov, 2017]). *A randomized algorithm $\mathcal{M} : \mathcal{D} \to \mathcal{R}$ is (α, ϵ)-RDP with order $\alpha \geq 1$, if for neighboring datasets $X, X' \in \mathcal{D}$ it holds that* $D_\alpha(\mathcal{M}(X)||\mathcal{M}(X')) \leq \epsilon$.

Renyi differential privacy is desirable for its straightforward *composition*, meaning that the privacy parameters degrade gracefully as additional computations are performed on the data, even when the private mechanisms are chosen adaptively. This allows us to design RDP mechanisms using simple private building blocks.

Theorem 3 (Basic RDP Composition [Mironov, 2017]). *Let $\mathcal{M}_1 : \mathcal{D} \to \mathcal{R}$ is (α, ϵ_1)-RDP and $\mathcal{M}_2 : \mathcal{D} \to \mathcal{R}$ is (α, ϵ_2)-RDP, then the mechanism defined as $(\mathcal{M}_1, \mathcal{M}_2)$ satisfies $(\alpha, \epsilon_1 + \epsilon_2)$-RDP.*

While DP also satisfies its own variant of composition, RDP is especially amenable to composition of Gaussian noise mechanisms. We can also easily translate between the notions of RDP and DP because any (α, ϵ)-RDP mechanism is also $(\epsilon_\delta, \delta)$-differential privacy for $\delta > 0$, as shown below in Theorem 4. Thus when running multiple RDP mechanisms, a common approach is to first perform RDP composition across the mechanisms and then translate the RDP guarantee into one of differential privacy.

Theorem 4 (From RDP to DP [Mironov, 2017]). *If $\mathcal{M}$ is (α, ϵ)-RDP, then it is also $(\epsilon + \frac{\log 1/\delta}{\alpha-1}, \delta)$-differential privacy for any $0 < \delta < 1$.*

Mechanisms for achieving both privacy notions typically add noise that scales with the *sensitivity* of the function being evaluated, which is the maximum change in the function's value between two neighboring databases. For a real-valued function q, this is formally defined as: $\Delta q = \max_{X,X' \text{ neighbors}} |q(X) - q(X')|$.

The Gaussian mechanism with parameters $(\epsilon, \delta, \sigma)$ takes in a function q, database X, and outputs $q(X) + \mathcal{N}(0, \sigma^2)$. The scale of the noise is fully specified as $\sigma = \sqrt{2\log(1.25/\delta)}\Delta q/\epsilon$, given the privacy parameters ϵ and δ and the query sensitivity Δq.

Theorem 5 (Privacy of Gaussian Mechanism [Dwork and Roth, 2014]). *The Gaussian Mechanism with parameter* $\sigma = \sqrt{2\log(1.25/\delta)}\Delta q/\epsilon$ *is* (ϵ, δ)*-differentially private.*

The AboveThresh algorithm [Dwork et al., 2009, Dwork and Roth, 2014] is a DP mechanism for handling a sequence of queries arriving online. It takes in a potentially unbounded stream of queries, compares the answer of each query to a fixed noisy threshold, and halts when it finds a noisy answer that exceeds the noisy threshold (denoted as $\top$, and otherwise $\perp$), where the added noise follows the Laplace distribution. In many cases, more concentrated noise (e.g., Gaussian) is preferred, and [Zhu and Wang, 2020] gives the generalized version of GenAboveThresh (presented in Algorithm 1), using general noise-adding mechanisms $\mathcal{M}_1$ and $\mathcal{M}_2$. These mechanisms can be any RDP algorithms that take in a real-valued input and produce a noisy estimate of the value. Our algorithm PrivSPRT will rely on an instantiation of GenAboveThresh using Gaussian mechanisms for differential privacy.

Algorithm 1 Generalized Above Noisy Threshold: GenAboveThresh$(X, \Delta, \{q_1, q_2, \ldots\}, H, \mathcal{M}_1, \mathcal{M}_2)$

Input: database X, stream of queries $\{q_1, q_2, \ldots\}$ each with sensitivity Δ, threshold H, noise-adding mechanisms $\mathcal{M}_1, \mathcal{M}_2$ that each add noise to their real-valued input.
Let $\hat{H} \sim \mathcal{M}_1(H)$
for each query i **do**
  Let $\hat{q}_i \sim \mathcal{M}_2(q_i(X))$
  if $\hat{q}_i > \hat{H}$ **then**
    Output $a_i = \top$
    Halt
  else
    Output $a_i = \perp$
  end if
end for

Theorem 6 (Privacy of GenAboveThresh [Zhu and Wang, 2020].). *Let* $\mathcal{M}_1$ *be any private mechanism that satisfies* $\epsilon_1(\alpha)$*-RDP for queries with sensitivity* Δ*, and* $\mathcal{M}_2$ *be any private mechanism that satisfies* $\epsilon_2(\alpha)$*-RDP for queries with sensitivity* 2Δ*. Let* T *be a random variable indicating the stopping time of Algorithm 1 instantiated with* $(X, \Delta, \{q_1, q_2, \ldots\}, H, \mathcal{M}_1, \mathcal{M}_2)$*. Then Algorithm 1 (denotes by* $\mathcal{M}$*) satisfies*

$$D_\alpha(\mathcal{M}(X)||\mathcal{M}(X')) \leq \epsilon_1(\alpha) + \epsilon_2(\alpha) + \frac{\log \sup \mathbb{E}[T|Z_1]}{\alpha - 1}, \tag{3}$$

and

$$D_\alpha(\mathcal{M}(X)||\mathcal{M}(X')) \leq \frac{\alpha - (\gamma - 1/\gamma)}{\alpha - 1}\epsilon_1(\frac{\gamma}{\gamma - 1}\alpha) + \epsilon_2(\alpha) + \frac{\log \mathbb{E}_{Z_1}(\mathbb{E}[T|Z_1]^\gamma)}{\gamma(\alpha - 1)}, \tag{4}$$

for all $\gamma > 1$ *and* $1 < \alpha < \infty$*, where* Z_1 *is the added noise from* $\mathcal{M}_1$*.*

In the case where the expected length is bounded by $t_{\max}$, Theorem 6 implies an RDP bound of the form $\epsilon_1(\alpha) + \epsilon_2(\alpha) + \log(1 + t_{\max})/(\alpha - 1)$.

3 Private Sequential Hypothesis Testing

In this section, we present our main result, which is a differentially private algorithm for the sequential hypothesis testing problem that also has small expected sample size and low Type I and Type II errors. We present our PrivSPRT algorithm in Section 3.1 and the theoretical results on privacy, error rates, and sample size in Section 3.2.

3.1 PrivSPRT algorithm

We present our algorithm for private sequential hypothesis testing, PrivSPRT, given formally in Algorithm 2. The algorithm is a private version of SPRT, and it uses two parallel instantiations of GenAboveThresh to ensure privacy of the statistical decision. It instantiates two Gaussian mechanims with parameters σ_1 and σ_2 as the noise-adding mechanisms, $\mathcal{M}_1$ and $\mathcal{M}_2$, respectively. At each time t, the algorithm computes the log-likelihood ratio ℓ_t for $x_1, x_2, \ldots, x_t$, and uses the Gaussian mechanism to add noise to the log-likelihood ratio. It then compares this noisy statistic against two pre-fixed noisy thresholds that depend on the SPRT decision thresholds a and b, and the other Gaussian mechanism with parameter σ_1. The stopping condition of PrivSPRT is similar to that of SPRT, only using noisy versions of the thresholds. Once the stopping condition is reached, the algorithm stops collecting additional samples and outputs its statistical decision.

It is useful to highlight that we add noises to the cumulative log-likelihood ratio statistics, will allow us to maintain the first-order statistical optimality of our proposed algorithms. Here, the first-order optimality means the expected sample sizes of our algorithms subject to the privacy constraints converge to the classical optimal non-private expected sample size results up to $O(1)$. Meanwhile, we should mention that one could also add noises individual log-likelihood ratio statistics to satisfy the privacy constraints, but doing so will

severely affect the expected sample sizes, and thus yield to algorithms that are suboptimal from the statistical efficiency viewpoint.

The sensitivity of the log-likelihood ratios is defined as: $\Delta(\ell) = \max_x \log \frac{f_1(x)}{f_0(x)} - \min_{x'} \log \frac{f_1(x')}{f_0(x')}$. For certain distributions, including Gaussians, the sensitivity $\Delta(\ell)$ is unbounded and therefore would require infinite noise to preserve privacy. We instead use a truncation parameter $A > 0$ to control the sensitivity of the log-likelihood ratio calculation, and add noise proportional to the post-truncation range. We note that the idea of truncating the likelihood for privacy also appears in [Canonne et al., 2019] for private simple hypotheses testing and [Zhang et al., 2021] for private sequential change-point detection. The A-truncated log-likelihood ratio is

$$\ell_t(A) = \sum_{i=1}^t [\log \tfrac{f_1(x_i)}{f_0(x_i)}]_{-A}^{A},$$

where the truncation operation is defined as $[x]_{-A}^{A} = -A, \text{if } x < -A; A, \text{if } x > A; x, \text{otherwise.}$

Algorithm 2 Private Sequential Probability Ratio Test: PrivSPRT($X, f_1, f_2, -a, b, \sigma_1, \sigma_2, A$)

Input: database X, distributions f_0, f_1, SPRT thresholds $-a, b$, Gaussian mechanisms $\mathcal{M}_1$ with parameter $(\epsilon'/2, \delta, \sigma_1)$ and $\mathcal{M}_2$ with parameters $(\epsilon'/2, \delta, \sigma_2)$, truncation parameter A
Let $\hat{-a} \sim \mathcal{M}_1(-a)$ and $\hat{b} \sim \mathcal{M}_1(b)$
for each time t **do**
  Compute $\ell_t(A) = \sum_{i=1}^t \left[\log \frac{f_1(x_i)}{f_0(x_i)}\right]_{-A}^{A}$
  Let $\hat{\ell}_t^a \sim \mathcal{M}_2(\ell_t(A))$ and $\hat{\ell}_t^b \sim \mathcal{M}_2(\ell_t(A))$
  if $\hat{\ell}_t^b > \hat{b}$ **then**
    Halt and output $d = 1$ (reject H_0)
  else if $\hat{\ell}_t^a < \hat{-a}$ **then**
    Halt and output $d = 0$ (accept H_0)
  else
    Proceed to the next iteration
  end if
end for

Comparing to standard AboveThresh. One may wonder why GenAboveThresh is needed, and whether the original AboveThresh algorithm of [Dwork et al., 2009, Dwork and Roth, 2014] with Laplace noise (as referred to as SparseVector) would be sufficient, perhaps with some loss in accuracy. In fact, this change to Laplace noise would break the desirable statistical properties of (non-private) SPRT. The properties of the SPRT depends on the overshoot of $\ell_T - b$ or $\ell_T - (-a)$, and to maintain the first-order optimality on the expected sample size, controlling the second moments of the noisy statistics is necessary. Adding Laplace noise will make the variance too large, and thus the desirable properties will break down. Empirically, we show in Section 4 that using Laplace noise instead of Gaussian noise results in undesirable performance. On the theoretical side, statistical analysis of the SPRT is traditionally based on renewal theory and overshoot analysis in applied probability, which both rely heavily on the central limit theorem (CLT), and thus the standard techniques are still applicable when adding Gaussian noise for privacy. On the other hand, if we add Laplace noise, the standard statistical techniques are inapplicable to characterize the overshoots; it remains an open problem to develop new tools to analyze the corresponding statistical properties.

3.2 Theoretical results on privacy, sample size, and error rates

In this subsection, we provide formal results on the privacy guarantees and statistical properties of PrivSPRT. For analyzing the expected sample size, we will relate $\mathbb{E}_0[T]$ and $\mathbb{E}_1[T]$ to the input parameters a, b. Similarly for analyzing the error rates, we will relate the Type I and Type II error to a and b. Recall that these errors respectively correspond to the false positive and false negative rates of the algorithm, which can be respectively defined as $\alpha = \Pr_0[\ell_t(A) + Z_t^b \geq b + Z_b]$ and $\beta = \Pr_1[\ell_t(A) + Z_t^a \leq a + Z_a]$ from PrivSPRT.

While our statistical properties of sample size and error rate are analyzed under the assumption that $x_1, x_2, \ldots$ follow either H_0 or H_1, as is standard in the statistics literature, our privacy guarantees hold unconditionally, regardless of the actual data distribution.

Privacy. Privacy of PrivSPRT follows by composition of two parallel instantiation of Algorithm 1, one each for the upper and lower bounds on ℓ_t. Theorem 3 gives Renyi divergence bounds for the outputs on two neighboring databases for GenAboveThresh, but it only implies Renyi differential privacy when the conditional expectation of the stopping time or the moments of conditional expectation of the stopping time are bounded. [Zhu and Wang, 2020] shows that the stopping time of GenAboveThresh instantiated with Gaussian noise and non-negative queries has bounded moments of the conditional expectation of the stopping time, and thus it satisfies RDP. However, in our case, the log-likelihood ratio queries can be negative, and this result cannot be immediately applied in our setting. Therefore, to prove that PrivSPRT is private, we must show that the expectation of the stopping time T is bounded. We remark that we can alternatively halt Algorithm 2 when t reaches an upper bound, and then make a decision using hypothesis testing methods when the sample size is fixed. However, this approach requires new analysis for the sample size and error

rates [Siegmund, 2013]. The full proof of Theorem 7 appears in Appendix A.1.

Theorem 7 (Privacy). *Let* $T_A = E_{Z_A}[1 + \rho_1^{-1} + \frac{5(a+Z_A)+3\sqrt{2}\sigma_2}{2\mu_0}]^\gamma$ *and* $T_B = E_{Z_B}[1 + \rho_0^{-1} + \frac{5(b+Z_B)+3\sqrt{2}\sigma_2}{2\mu_1}]^\gamma$, *where* $\rho_0 = 1 - \exp(-\frac{(1-c)\mu_1^2}{2A^2})$ *and* $\rho_1 = 1 - \exp(-\frac{(1-c)\mu_0^2}{2A^2})$. *Then algorithm 2 satisfies* $(\alpha, \frac{\alpha\gamma/(\gamma-1)-1}{\alpha-1}\frac{2\alpha A^2}{\sigma_1^2} + \frac{4\alpha A^2}{\sigma_2^2} + \frac{2\log\max\{T_A,T_B\}}{\gamma(\alpha-1)})$*-RDP, for any* $1 < \alpha < \infty$.

Corollary 1. *For* σ_1 *and* σ_2 *are chosen to be the parameters specified in the Gaussian mechanisms that satisfy* $(\epsilon'/2, \delta)$*-differential privacy,* PrivSPRT$(X, f_1, f_2, -a, b, \sigma_1, \sigma_2, A)$ *in Algorithm 2 satisfies* $(\alpha, (\frac{\alpha\gamma/(\gamma-1)-1}{\alpha-1} + 1)\epsilon' + \frac{2\log\max\{T_A,T_B\}}{\alpha-1})$*-RDP, for any* $1 < \alpha < \infty$.

Because we are using the Gaussian mechanism as the noise-adding mechanism, the dependence of the stopping time in the privacy guarantee is unavoidable. T_A and T_B in Theorem 7 are the moments of the conditional expectation of the stopping time, which depending on the true underlying distribution that generated the data; T_A is roughly $O((\frac{a+\sigma_2}{\mu_0})^\gamma)$, and similarly T_B is roughly $O((\frac{b+\sigma_2}{\mu_1})^\gamma)$. Theorem 7 and Corollary 1 further imply an (ϵ, δ)-differential privacy bound for PrivSPRT by Theorem 4. For $\delta < 1/2\log\max\{T_A, T_B\}$, and σ_1 and σ_2 are chosen to be the parameters specified in the Gaussian mechanisms that satisfy $(\epsilon'/2, \delta)$-differential privacy, PrivSPRT is (ϵ, δ)-differentially private, with $\epsilon = (\frac{\alpha\gamma/(\gamma-1)-1}{\alpha-1} + 1)\epsilon' + 4\log(1/\delta)/(\alpha-1)$.

Sample Size. When analyzing statistical properties of PrivSPRT, an important quantity is the expectation of the truncated individual log-likelihood ratios:

$$\mu_0 = -\mathbb{E}_0[\log\tfrac{f_1(x)}{f_0(x)}]_{-A}^{A}, \quad \text{and} \quad \mu_1 = \mathbb{E}_1[\log\tfrac{f_1(x)}{f_0(x)}]_{-A}^{A}.$$

When A goes to ∞, the above expectations converge to the KL-divergence between f_0 and f_1.

A technical challenge that arises in bounding the expected sample size is that the noisy log-likelihood ratio at time t cannot be decomposed into a summation of t i.i.d. random variables because of the noise terms. This preludes the use of Wald's identity [Wald, 1944], which is used in the proof of bounded sample size for non-private SPRT, and relates the expectation of a sum of randomly-many finite-mean, i.i.d. random variables to the expected number of terms in the sum and the expectation of the random variables.

Instead, we leverage a critical fact that $\mathbb{E}_i[T] = \sum_{t=0}^{\infty}\Pr_i[T > t]$ for $i \in \{0, 1\}$, and thus relate the expected sample size to the probability of the noisy truncated log-likelihood ratio being within the noisy thresholds at each time t. Since the event is less probable for a large t, we partition the range $[0, \infty)$ into several sub-intervals, and bound the probability in each sub-interval seperately. This results in our $O(\frac{b}{\mu_1})$ bound on the expected sample size in Theorem 8 when the noise parameters σ_1 and σ_2 go to 0. This result is consistent with the non-private sample size result $O(\frac{b}{D_{KL}})$, and it is first-order optimal. We note that a similar idea of partitioning the whole range into sub-intervals also appears in [Liu and Mei, 2020], where it was applied only for handling Gaussian data.

The last term in the bound of Theorem 8 is the additional cost that comes from adding Gaussian noise, which quantifies the cost of privacy. In the proof, we permit large values of the difference between the Gaussian noise Z_t to Z_b (or Z_a) for a large t, which reduces the additional expected sample size required for privacy. The analysis relies on partitioning the range into k intervals and a time-specific threshold depending on a constant c, and the results are under the optimal choice of k and c. The proof is given in Appendix A.2.

Theorem 8 (Sample Size). *The expected sample size of* PrivSPRT$(X, f_1, f_2, -a, b, \sigma_1, \sigma_2, A)$ *under* H_1 *satisfies* $\mathbb{E}_1[T] \le 1 + \min_{k\in\mathbb{N}}\min_{c\in(0,1)}(\frac{b}{(1-c)\mu_1} + \frac{1}{2(k+1)}\frac{b}{(1-c)\mu_1} + (k+1)\rho_1^{-1} + \frac{3\sqrt{2(\sigma_1^2+\sigma_2^2)}}{4(1-c)\mu_1})$, *where* $\rho_0 = 1 - \exp(-\frac{(1-c)\mu_1^2}{2A^2})$. *Similarly, the expected sample size under* H_0 *satisfies* $\mathbb{E}_0[T] \le 1 + \min_{k\in\mathbb{N}}\min_{c\in(0,1)}(\frac{a}{(1-c)\mu_0} + \frac{1}{2(k+1)}\frac{a}{(1-c)\mu_0} + (k+1)\rho_0^{-1} + \frac{3\sqrt{2(\sigma_1^2+\sigma_2^2)}}{4(1-c)\mu_0})$, *where* $\rho_1 = 1 - \exp(-\frac{(1-c)\mu_0^2}{2A^2})$.

To interpret the results in Theorem 8, we choose a specific (potentially suboptimal) values of k and c. Choosing $k = 1$ and $c = \frac{1}{2}$ gives $\mathbb{E}[T] \le 1 + \rho_1^{-1} + \frac{5b}{2\mu_1} + \frac{3\sqrt{2(\sigma_1^2+\sigma_2^2)}}{2\mu_1}$, which is $O(\frac{b}{\mu_1}) + O(\frac{\sqrt{(\sigma_1^2+\sigma_2^2)}}{\mu_1})$. The first term is the same as in the classical non-private results, and the second term is the additional cost for privacy. Since σ_1 and σ_2 will be chosen to scale with $\frac{A}{\epsilon'}$, the additional cost for privacy is $O(\frac{A}{\mu_1\epsilon})$, where μ_1 is the expectation of the truncated log-likelihood ratios, which serves as a distance measure similar to the KL divergence. Our expected sample size and error rate results converge to the classical non-private results up to $O(1)$, ignoring the dependence on ϵ. The asymptotic dependence on ϵ is $O(1/\epsilon)$, which matches the sample complexity dependence on ϵ in the simpler problem of private simple hypothesis testing [Canonne et al., 2019].

Error rates. We now move to provide guarantees for the Type I and Type II error rates of PrivSPRT. In the classical sequential hypothesis testing literature for non-private SPRT, the standard technique to characterize the error rates is based on the change of measure

method that heavily utilizes the likelihood ratio statistics. Unfortunately, the test statistics of PrivSPRT are no longer the likelihood ratio, since the algorithm add Gaussian noise and truncates the log-likelihood for privacy. As a result, the standard change-of-measure technique is no longer applicable.

To characterize the error rates of PrivSPRT, we apply an alternative method based on the brute force estimation of the error probabilities, which was first proposed in [Sahu and Kar, 2015] in the context of distributed hypothesis testing in sensor networks. It turns out that this alternative method is also applicable to the setting of PrivSPRT. The main idea is as follows: Type I error, $\Pr_0[d=1]$, can be written as a sum of probabilities of the noisy log-likelihood ratio being above the noisy threshold at time t and the event that the stopping time is t for all $t>0$: $\sum_{t=1}^{\infty}\Pr_0[\ell_t(A)+Z_t > b+Z_b \wedge T=t]$. We then partition the range of time $[1,\infty)$ into several sub-intervals and analyze them separately as before with the expected sample size. Although the high-level approach is similar to analyzing the expected sample size, the sub-intervals need to be carefully chosen here to give a meaningful bound for the error rates. The detailed proof is deferred to Appendix A.3.

Theorem 9 (Error Rate)**.** *Let $d\in\{0,1\}$ be the decision output by* PrivSPRT$(X, f_1, f_2, -a, b, \sigma_1, \sigma_2, A)$. *Then the Type I error is bounded by:*

$$Pr_0[d=1] \le \min_{k\in\mathbb{N}}\min_{c\in(0,1)}\{Q_1+Q_2+Q_3\}, \quad (5)$$

where $Q_1 = 2\rho_0^{-1}\exp(-\frac{2b(1-c)\mu_0}{A^2})(1+k\exp(\frac{1}{8k}))$, $Q_2 = k\exp(\frac{1}{4k+3}))$ *and* $Q_3 = \frac{\sqrt{2(\sigma_1^2+\sigma_2^2)}}{4(1-c)\mu_0}$, *and* $\rho_0 = 1-\exp(-\frac{(1-c)\mu_0^2}{2A^2})$. *The Type II error is bounded by:*

$$Pr_1[d=0] \le \min_{k\in\mathbb{N}}\min_{c\in(0,1)}\{W_1+W_2+W_3\}, \quad (6)$$

where $W_1 = 2\rho_1^{-1}\exp(-\frac{2a(1-c)\mu_1}{A^2})(1+k\exp(\frac{1}{8k}))$, $W_2 = k\exp(\frac{1}{4k+3}))$, *and* $W_3 = \frac{\sqrt{2(\sigma_1^2+\sigma_2^2)}}{4(1-c)\mu_1}$, *and* $\rho_1 = 1-\exp(-\frac{(1-c)\mu_1^2}{2A^2})$.

To interpret the results, in Theorem 9, choosing $k=1$ and $c = 1-\frac{A^2}{2\mu_0}$ gives $\Pr_0[d=1] \le 2\rho_0^{-1}(1+\exp(\frac{1}{8}))\exp(-b)+\frac{\sqrt{2(\sigma_1^2+\sigma_2^2)}}{2A^2}+\exp(\frac{1}{7})$. Again, the first term is the same as the non-private result $O(\exp(-b))$. The additional $O(\frac{\sqrt{\sigma_1^2+\sigma_2^2}}{A^2})$ term quantifies the cost of privacy. Since we are instantiating the Gaussian mechanisms with noise parameters σ_1 and σ_2 proportional to the sensitivity $2A$ and the privacy parameter ϵ, the additional error term is reduced to $O(\frac{1}{\epsilon A})$. This implies the algorithm will incur a larger error rate for stronger privacy guarantees.

4 Numerical Results

In this section, we present results from Monte Carlo experiments designed to validate the theoretical results of PrivSPRT. We only need to validate the statistical properties of PrivSPRT— sample size and error rates — since the privacy guarantee holds even in the worst-case over databases and hypotheses. In Section 4.1, we focus on sequentially testing means of Bernoulli distributions; in Appendix B.1, we provide additional empirical results on testing means of Gaussian distributions. In Appendix B.2, we demonstrate empirically that the classic AboveThresh mechanism does not provide satisfactory performance in terms of sample size and error rates, thus justifying our algorithmic modifications made in PrivSPRT.

4.1 Testing on Bernoulli Data

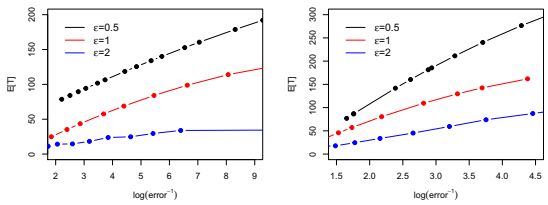
In this section, our experiment focus on Bernoulli data, where $x_1, x_2, \cdots \sim Ber(\theta)$ are sampled i.i.d. from a Bernoulli distribution with parameter θ. Monitoring Bernoulli data is one of the early research in the fully sequential design in clinical trials, see [Armitage, 1950]. For instance, one want to evaluate the effect of a new drug or treatment on the mortality rate of an unknown infectious disease such as COVID-19 in a sub-population of groups.

Here we consider two different scenarios that are simple yet useful to shed new lights on real-world applications. One is when the distance between the null hypothesis and the alternative hypothesis on θ is large, say, $H_0:\theta=0.7$ against $H_1:\theta=0.3$, e.g., the effect of a new treatment is expected to be significant to reduce the mortality rate among people whose age is 65 years or older in a developing country. The other is when the distance between the null hypothesis and the alternative hypothesis on θ is small, say, $H_0:\theta=0.6$ against $H_1:\theta=0.4$, e.g., the effect of a drug to certain age group with certain diseases in a developed country. Since $\mu_0=\mu_1$ under this setting, the expected sample sizes under H_1 and H_0 are identical, and similarly, the Type I error and Type II error are also identical. For simplicity, we will use $E[T]$ and error to denote the expected sample size and the error, respectively.

To obtain an accurate estimate of Type I and Type II errors, we use the importance sampling technique for the Monte Carlo simulations. This is because the estimate of the Type I error based on n independent trials is $n^{-1}\sum_{k=1}^{n}\frac{f_0(X[1:T])}{f_1(X[1:T])}I(\ell_T(A)+Z_T \ge b+Z_b)$ where the sample X is generated from f_1 has much smaller variance compared to the naive estimate $n^{-1}\sum_{k=1}^{n}I(\ell_T(A)+Z_T\ge b+Z_b)$ where the sample X is generated from f_0.

We use two $(\epsilon'/2, \delta = 1e-05)$-differentially private Gaussian mechanisms as the noise-adding mechanisms in PrivSPRT, corresponding to $\sigma_1^2 = 32\log(1.25/\delta)A^2/\epsilon^2$ and $\sigma_2^2 = 128\log(1.25/\delta)A^2/\epsilon^2$. Although the log-likelihood ratio is uniformly bounded for Bernoulli data, we invoke the truncation with parameter A because μ_0 and μ_1 are linear with respect to A for Bernoulli data, which makes the validation easier. For each simulation, we repeat the process for 10^5 times. The results are presented in Figure 1, which plots the expected sample size $E[T]$ against the log scale of 1/error, with varying the privacy parameter $\epsilon' = 0.5, 1, 2$. From this figure, when we want to provide a stronger privacy, i.e., when ϵ becomes smaller, then we will have larger expected sample sizes for given Type I and Type II error probabilities constraints. This is consistent with our intuition on the tradeoff between privacy and statistical efficiency.

We also conduct experiments for testing $H_0 : \theta = 0.7$ against $H_1 : \theta = 0.2$, when $E_1[T]$ and $E_2[T]$ are not symmetric. We vary this truncation parameter $A = 0.05, 0.2, 0.5, 0.7$ in our experiments. For each fixed A and ϵ, we choose thresholds a and b through Monte Carlo simulation to control the Type I error and Type II error at the same level (10^{-6}).The results of these simulations are presented in Table 1.



(a) Large distance (b) Small distance

Figure 1: Three-way trade-off between privacy, expected sample size, and error rate. For large distance (left), we are testing $H_0 : \theta = 0.7$ against $H_1 : \theta = 0.3$; for small distance (right), we are testing $H_0 : \theta = 0.6$ against $H_1 : \theta = 0.4$.

Table 1 shows three positive results. First, for each fixed privacy parameter ϵ', the expected sample sizes are almost the same across varying A, and the thresholds are almost linear with respect to A. This suggests that the expected sample size $\mathbb{E}_0[T]$ (resp. $\mathbb{E}_1[T]$) is proportional to a/A or A/a (resp. b/A or A/b). The parameter A controls a trade-off between how much information is lost from truncation in the log-likelihood ratios and how much noise is added for privacy. Thus expected sample sizes are larger for a larger $A = 0.7$ with $\epsilon' = 0.5, 1$, as the additional noise starts to dominate the information provided by the log-likelihood ratios. Second, in our setting, the expectation of the truncated log-likehood ratio $\mu_0 = 0.6A$ and $\mu_1 = 0.4A$. We see from Table 1 that $\mathbb{E}_0[T]/\mathbb{E}_1[T]$ is roughly 2/3

Table 1: Numerical values of expected sample size under H_0 and H_1, Type I error and Type II error for testing the Bernoulli parameter.

A	ϵ'	a, b	error rates	$\mathbb{E}_0[T]$	$\mathbb{E}_1[T]$
0.05	0.5	8,7.5	10^{-6}	139.662	172.89
	1	4.3, 4.3		86.12	122.144
	2	2.5, 2.5		61.683	88.307
0.2	0.5	32,32		139.456	195.504
	1	16.8, 16.8		85.336	123.542
	2	9.5,9.5		56.645	83.127
0.5	0.5	80,80		139.252	199.718
	1	43, 43		88.137	127.986
	2	25, 25		61.494	88.182
0.7	0.5	125,120		173.305	227.387
	1	63,63		95.304	136.336
	2	35,35		61.944	87.363
	∞	16, 16		29.607	28.318

for all the cases, which further validates Theorem 8 that $\mathbb{E}_0[T]$ (resp. $\mathbb{E}_1[T]$) is $O(a/A)$ (resp. $O(b/A)$). Third, for each fixed A, $a/\mathbb{E}_0[T]$ (resp. $b/\mathbb{E}_1[T]$) decreases as ϵ' increases for weaker privacy, which is consistent with Theorem 8, because the additional cost does not involve the threshold, and it decreases for weaker privacy.

4.2 Comparing with the Standard AboveThresh

We conduct experiments for testing means of Gaussian data using both the PrivSPRT with Gaussian noise and the original AboveThresh algorithm with Laplace noise. The results are presented in Table 2. For PrivSPRT we use two $(\epsilon'/2, \delta = 1e-05)$-differentially private Gaussian mechanisms as the noise-adding mechanisms. In AboveThresh we use two $\epsilon/2$-differentially private Laplace mechanisms, and the total privacy loss is 2ϵ. We vary the truncation parameter $A = 0.5, 1, 2, 5$, and choose the thresholds a, b through Monte Carlo simulation with importance sampling to control Type I and Type II errors at the 0.05 level. Table 2 shows that using the original AboveThresh algorithm with Laplace noise results in much larger expected sample sizes, given that the Type I and Type II errors are fixed. We note that although the overall privacy cost for PrivSPRT is slightly larger, PrivSPRT provides a better trade-off between privacy and accuracy. The detailed setup and discussion are deferred to Appendix B.

Acknowledgements

W.Z. was supported by a Computing Innovation Fellowship from the Computing Research Association

(CRA) and the Computing Community Consortium (CCC). R.C. was supported in part by NSF grants CNS-1850187 and CNS-1942772 (CAREER), and a JPMorgan Chase Faculty Research Award. Y.M. was supported in part by NSF-DMS grant 2015405. Part of this work was completed while W.Z. and R.C. were at Georgia Institute of Technology.

References

[Armitage, 1950] Armitage, P. (1950). Sequential analysis with more than two alternative hypotheses, and its relation to discriminant function analysis. *Journal of the Royal Statistical Society. Series B (Methodological)*, 12(1):137–144.

[Armitage, 1954] Armitage, P. (1954). Sequential tests in prophylactic and therapeutic trials. *Quarterly Journal of Medicine*, 23(91):255–274.

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

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

[Cummings et al., 2020] Cummings, R., Krehbiel, S., Lut, Y., and Zhang, W. (2020). Privately detecting changes in unknown distributions. In *Proceedings of the 37th International Conference on Machine Learning*, pages 958–968.

[Cummings et al., 2018] Cummings, R., Krehbiel, S., Mei, Y., Tuo, R., and Zhang, W. (2018). Differentially private change-point detection. In *Advances in Neural Information Processing Systems*, pages 10825–10834.

[Dajani et al., 2017] Dajani, A. N., Lauger, A. D., Singer, P. E., Kifer, D., Reiter, J. P., Machanavajjhala, A., Garfinkel, S. L., Dahl, S. A., Graham, M., Karwa, V., Kim, H., Lelerc, P., Schmutte, I. M., Sexton, W. N., Vilhuber, L., and Abowd, J. M. (2017). The modernization of statistical disclosure limitation at the U.S. census bureau. Presented at the September 2017 meeting of the Census Scientific Advisory Committee.

[DeMets, 1998] DeMets, D. L. (1998). Sequential designs in clinical trials. *Cardiac Electrophysiology Review*, 2(1):57–60.

[Differential Privacy Team, Apple, 2017] Differential Privacy Team, Apple (2017). Learning with privacy at scale. https://machinelearning.apple.com/docs/learning-with-privacy-at-scale/appledifferentialprivacysystem.pdf.

[Ding et al., 2017] Ding, B., Kulkarni, J., and Yekhanin, S. (2017). Collecting telemetry data privately. In *Advances in Neural Information Processing Systems 30*, NIPS '17, pages 3571–3580. Curran Associates, Inc.

[Dwork et al., 2006] Dwork, C., McSherry, F., Nissim, K., and Smith, A. (2006). Calibrating noise to sensitivity in private data analysis. In *Proceedings of the 3rd Conference on Theory of Cryptography*, TCC '06, pages 265–284.

[Dwork et al., 2009] Dwork, C., Naor, M., Reingold, O., Rothblum, G. N., and Vadhan, S. P. (2009). On the complexity of differentially private data release: efficient algorithms and hardness results. In *Proceedings of the 41st ACM Symposium on Theory of Computing*, STOC '09, pages 381–390.

[Dwork and Roth, 2014] Dwork, C. and Roth, A. (2014). The algorithmic foundations of differential privacy. *Foundations and Trends in Theoretical Computer Science*, 9(3–4):211–407.

[Erlingsson et al., 2014] Erlingsson, Ú., Pihur, V., and Korolova, A. (2014). RAPPOR: Randomized aggregatable privacy-preserving ordinal response. In *Proceedings of the 2014 ACM Conference on Computer and Communications Security*, CCS '14, pages 1054–1067, New York, NY, USA. ACM.

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

[Gaboardi and Rogers, 2018] Gaboardi, M. and Rogers, R. (2018). Local private hypothesis testing: Chi-square tests. In *International Conference on Machine Learning*, pages 1626–1635. PMLR.

[Ghosh and Sen, 1991] Ghosh, B. K. and Sen, P. K. (1991). *Handbook of sequential analysis*. CRC Press.

[Harrell, 2020] Harrell, F. E. (2020). Sequential bayesian designs for rapid learning in COVID-19 clinical trials. URL https://www.fharrell.com/talk/seqbayes/ Last checked 10/15/21.

[Liu and Mei, 2020] Liu, K. and Mei, Y. (2020). Improved performance properties of the cisprt algorithm for distributed sequential detection. *Signal Processing*, 172:107573.

[Mironov, 2017] Mironov, I. (2017). Rényi differential privacy. In *2017 IEEE 30th Computer Security Foundations Symposium (CSF)*, pages 263–275. IEEE.

[Sahu and Kar, 2015] Sahu, A. K. and Kar, S. (2015). Distributed sequential detection for gaussian shift-in-mean hypothesis testing. *IEEE Transactions on Signal Processing*, 64(1):89–103.

[Sheffet, 2018] Sheffet, O. (2018). Locally private hypothesis testing. In *International Conference on Machine Learning*, pages 4605–4614. PMLR.

[Siegmund, 2013] Siegmund, D. (2013). *Sequential analysis: tests and confidence intervals*. Springer Science & Business Media.

[Wald, 1944] Wald, A. (1944). On cumulative sums of random variables. *The Annals of Mathematical Statistics*, 15(3):283–296.

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

[Wald, 2004] Wald, A. (2004). *Sequential analysis*. Courier Corporation.

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

[Wang et al., 2020] Wang, Y., Sibai, H., Mitra, S., and Dullerud, G. E. (2020). Differential privacy for sequential algorithms. *arXiv preprint arXiv:2004.00275*.

[Wang et al., 2019] Wang, Y.-X., Balle, B., and Kasiviswanathan, S. P. (2019). Subsampled renyi differential privacy and analytical moments accountant. In Chaudhuri, K. and Sugiyama, M., editors, *Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics*, volume 89 of *Proceedings of Machine Learning Research*, pages 1226–1235. PMLR.

[Whitehead, 1997] Whitehead, J. (1997). *The Design and Analysis of Sequential Clinical Trials*. Wiley, New York, revised second edition edition.

[Zhang et al., 2021] Zhang, W., Krehbiel, S., Tuo, R., Mei, Y., and Cummings, R. (2021). Single and multiple change-point detection with differential privacy. *Journal of Machine Learning Research*, 22(29):1–36.

[Zhu and Wang, 2020] Zhu, Y. and Wang, Y.-X. (2020). Improving sparse vector technique with renyi differential privacy. *Advances in Neural Information Processing Systems*, 33.

Supplementary Material: Private Sequential Hypothesis Testing for Statisticians: Privacy, Error Rates, and Sample Size

A Omitted Proofs

In this appendix, we provide proofs for our main theorems, which were omitted in the body of the paper due to space reasons. We restate the theorems here for convenience.

A.1 Proof of privacy

Theorem 7 (Privacy). *Let* $T_A = E_{Z_A}[1 + \rho_1^{-1} + \frac{5(a+Z_A)+3\sqrt{2}\sigma_2}{2\mu_0}]^\gamma$ *and* $T_B = E_{Z_B}[1 + \rho_0^{-1} + \frac{5(b+Z_B)+3\sqrt{2}\sigma_2}{2\mu_1}]^\gamma$, *where* $\rho_0 = 1 - \exp(-\frac{(1-c)\mu_1^2}{2A^2})$ *and* $\rho_1 = 1 - \exp(-\frac{(1-c)\mu_0^2}{2A^2})$. *Then algorithm 2 satisfies* $(\alpha, \frac{\alpha\gamma/(\gamma-1)-1}{\alpha-1}\frac{2\alpha A^2}{\sigma_1^2} + \frac{4\alpha A^2}{\sigma_2^2} + \frac{2\log\max\{T_A,T_B\}}{\gamma(\alpha-1)})$*-RDP, for any* $1 < \alpha < \infty$.

Proof. We first show that the expectation of the stopping time T is bounded given Z_A and Z_B. We instead show the equivalent fact that $P_i(T = \infty) = 0$ for $i = 0, 1$. Define a constant $d = a + b$. If $T = \infty$, then for any positive integer r, the following inequalities must hold:

$$(\sum_{i=kr+1}^{(k+1)r} [\log \frac{f_1(x_i)}{f_0(x_i)}]_{-A}^{A} + Z)^2 < d^2 \quad k = 0, 1, 2, \ldots, \tag{7}$$

where $Z \sim N(0, \sigma_2^2)$. We can further express Z as a summation of r independent Gaussians $\sum_{i=1}^{r} Z_i$, and then (7) is equivalent to

$$(\sum_{i=kr+1}^{(k+1)r} ([\log \frac{f_1(x_i)}{f_0(x_i)}]_{-A}^{A} + Z_i))^2 < d^2 \quad k = 0, 1, 2, \ldots. \tag{8}$$

To prove $P_i(T = \infty) = 0$ for $i = 0, 1$, it is sufficient to show that the probability is zero that (8) holds for all integer values of k. Since the variance of $[\log \frac{f_1(x_i)}{f_0(x_i)}]_{-A}^{A} + Z_i$ is not zero, and it is bounded below by the variance of $[\log \frac{f_1(x_i)}{f_0(x_i)}]_{-A}^{A}$, the expected value of $(\sum_{i=1}^{j}([\log \frac{f_1(x_i)}{f_0(x_i)}]_{-A}^{A} + Z_i))^2$ converges to ∞ as j goes to ∞. Therefore, there exists a positive integer r such that

$$P[(\sum_{i=1}^{j}([\log \frac{f_1(x_i)}{f_0(x_i)}]_{-A}^{A} + Z_i))^2 < r^2] < 1. \tag{9}$$

From (9) it follows that the probability that (8) is fulfilled for all values of k up to ∞ is equal to zero (using a union bound over all k), and thus $P_i(T = \infty) = 0$ for $i = 0, 1$. Hence, $E_i[T|Z_A, Z_b]$ is bounded, and then $E_i[T|Z_A, Z_b]^\gamma$ is bounded. We use the same method to compute the upper bound of $E_i[T|Z_A, Z_b]$ as in the proof in A.2 with $\sigma_1 = 0$, and a and b replaced by $a + Z_A$ and $b + Z_B$, respectively. For $i = 0, 1$, we denote $E_i[T|Z_A, Z_b]^\gamma$ as T_A and T_B, respectively. Since we consider the worst case for privacy, we take the maximum over T_A and T_B in the final bound. For Gaussian mechanisms, $\epsilon_1(\frac{\gamma}{\gamma-1}\alpha) = \frac{\gamma\alpha A^2}{(\gamma-1)\sigma_1^2}$ and $\epsilon_2(\alpha) = \frac{2\alpha A^2}{\sigma_2^2}$. From inequality (4) in Theorem 6, it follows that using GenAboveThreshfor truncated log-likelihood ratio queries satisfies $(\frac{\alpha\gamma/(\gamma-1)-1}{\alpha-1}\frac{\alpha A^2}{\sigma_1^2} + \frac{2\alpha A^2}{\sigma_2^2} + \frac{\log\max\{T_A,T_B\}}{\gamma(\alpha-1)})$-RDP, for any $1 < \alpha < \infty$. Then privacy of PrivSPRT follows from composition of two parallel instantiations of Algorithm 1. □

A.2 Proof of sample size

Theorem 8 (Sample Size). *The expected sample size of* PrivSPRT$(X, f_1, f_2, -a, b, \sigma_1, \sigma_2, A)$ *under* H_1 *satisfies* $\mathbb{E}_1[T] \leq 1 + \min_{k\in\mathbb{N}} \min_{c\in(0,1)}(\frac{b}{(1-c)\mu_1} + \frac{1}{2(k+1)}\frac{b}{(1-c)\mu_1} + (k+1)\rho_1^{-1} + \frac{3\sqrt{2(\sigma_1^2+\sigma_2^2)}}{4(1-c)\mu_1})$, *where* $\rho_0 = 1 - \exp(-\frac{(1-c)\mu_1^2}{2A^2})$. *Similarly, the expected sample size under* H_0 *satisfies* $\mathbb{E}_0[T] \leq 1 + \min_{k\in\mathbb{N}} \min_{c\in(0,1)}(\frac{a}{(1-c)\mu_0} + \frac{1}{2(k+1)}\frac{a}{(1-c)\mu_0} + (k+1)\rho_0^{-1} + \frac{3\sqrt{2(\sigma_1^2+\sigma_2^2)}}{4(1-c)\mu_0})$, *where* $\rho_1 = 1 - \exp(-\frac{(1-c)\mu_0^2}{2A^2})$.

Proof. In the proof, we leverage a critical fact that $\mathbb{E}_i[T] = \sum_{t=0}^{\infty} \Pr_i(T > t)$ for $i \in \{0, 1\}$, and thus relate the expected sample size to the probability of the noisy truncated log-likelihood ratio within the noisy thresholds for each time t. Since the event is less probable for a large t, we partition the range $[0, \infty)$ into several sub-intervals, and bound the probability in each sub-interval seperately. We provide the detailed proof for $E_1[T]$, the proof is the same for $E_0[T]$ with b replaced by a and μ_1 replaced by μ_0.

$$\begin{aligned} E_1[T] &= \sum_{t=0}^{\infty} P_1(T > t) \\ &\leq \sum_{t=0}^{\infty} P_1(\ell_t + Z_t \leq b + Z_b) \\ &\leq \sum_{t=0}^{\infty} P_1(\ell_t - t\mu_1 \leq b - t\mu_1 + \delta_t) + P_1(Z_t \leq Z_b - \delta_t) \end{aligned} \tag{10}$$

We will bound the first term in (10) as follows. Let $\delta_t = ct\mu_1$, where c is a constant within $(0, 1)$.

$$\begin{aligned} &\sum_{t=0}^{\infty} P_1(\ell_t - t\mu_1 \leq b - t\mu_1 + \delta_t) \\ =&\sum_{t=0}^{\infty} P_1(\ell_t - t\mu_1 \leq b - (1-c)t\mu_1) \end{aligned} \tag{11}$$

Let γ denote $\frac{b}{(1-c)\mu_1}$, and m denote $(1-c)\mu_1$. We bound the infinite sum in (11) by partitioning $[0, \infty]$ into four sub-intervals:

$$[0, \gamma], \quad (\gamma, \frac{3}{2}\gamma], \quad (\frac{3}{2}\gamma, 2\gamma], \quad (2\gamma, \infty).$$

Let S_1, S_2, S_3, S_4 respectively denote the summation value as the index t ranges over these sub-intervals. When $t \in [0, \gamma]$, we have $b - (1-c)t\mu_1 > 0$. Since $\ell_t - t\mu_1$ is a mean-zero random variable, we bound S_1 by

$$\begin{aligned} S_1 &= \sum_{t=1}^{[\gamma]} P_1(\ell_t - t\mu_1 \leq b - (1-c)t\mu_1) \\ &\leq \sum_{t=1}^{[\gamma]} 1 \leq \gamma + 1. \end{aligned} \tag{12}$$

When $t > \gamma$, following Hoeffding inequality, we have $P_1(\ell_t - t\mu_1 \leq b - (1-c)t\mu_1) \leq \exp(-\frac{(b-mt)^2}{2tA^2})$. We will use

the following observation as the main tool. For any i and j with $i < j$, we have

$$\begin{aligned}
&\sum_{i}^{j} \exp(-\frac{(b-mt)^2}{2tA^2}) \\
\leq &\sum_{i}^{j} \exp(-\frac{b^2}{2jA^2} + \frac{bm}{A^2} - \frac{m^2 t}{2A^2}) \\
= &\exp(-\frac{b^2}{2jA^2} + \frac{bm}{A^2}) \sum_{i}^{j} \exp(-\frac{m^2 t}{2A^2}) \\
= &\exp(-\frac{b^2}{2jA^2} + \frac{bm}{A^2}) \frac{\exp(-\frac{m^2 i}{2A^2}) - \exp(-\frac{m^2 (j+1)}{2A^2})}{1 - \exp(-\frac{m^2}{2A^2})} \\
\leq &\rho^{-1} \exp(-\frac{b^2}{2jA^2} + \frac{bm}{A^2}) \exp(-\frac{m^2 i}{2A^2}),
\end{aligned} \tag{13}$$

where $\rho = 1 - \exp(-\frac{m^2}{2A^2})$. By applying (13) to the case where $i = \frac{3}{2}\gamma$ and $j = 2\gamma$, we obtain a bound on

$$S_3 \leq \rho^{-1}. \tag{14}$$

Similarly, by applying (13) to the case where $i = 2\gamma$ and $j = \infty$, we obtain a bound on

$$S_4 \leq \rho^{-1}. \tag{15}$$

To bound S_2, we further partition the sub-interval $(\gamma, \frac{3}{2}\gamma]$ into k intervals:

$$(\gamma, \frac{k+2}{k+1}\gamma], \quad \text{and} \quad (\frac{j+2}{j+1}\gamma, \frac{j+1}{j}\gamma], \text{ for } j = 2, \ldots, k.$$

For the first interval $(\gamma, \frac{k+2}{k+1}\gamma]$, since $b - mt < 0$ and $\ell_t - t\mu_1$ is a mean-zero random variable, we have the simple fact that $P_1(\ell_t - t\mu_1 \leq b - (1-c)t\mu_1) \leq \frac{1}{2}$. Then the summation over the first interval is bounded by $\frac{1}{2(k+1)}\gamma$. By applying (13) to the remaining $k-1$ intervals with $i = \frac{j+2}{j+1}\gamma$ and $j = \frac{j+1}{j}\gamma$, we obtain

$$\begin{aligned}
S_2 &\leq \frac{1}{2(k+1)}\gamma + \sum_{j=2}^{k} \sum_{t=[\frac{j+2}{j+1}\gamma]}^{[\frac{j+1}{j}\gamma]} \exp(-\frac{(b-mt)^2}{2tA^2}) \\
&\leq \frac{1}{2(k+1)}\gamma + (k-1)\rho^{-1},
\end{aligned} \tag{16}$$

for any k. Combining (12), (16), (14) and (15), we can bound the first term in (10) by $1 + \gamma + \min_k\{\frac{1}{2(k+1)}\gamma + (k+1)\rho^{-1}\}$.

Next we bound the second term in (10). We will use the fact that $\Pr(N(0, \sigma^2) > x) \leq \frac{1}{2}\exp(-\frac{x^2}{2\sigma^2})$ for a Gaussian distribution.

$$\begin{aligned}
\Pr(Z_b - Z_t \geq \delta_t) &\leq \Pr(N(0, \sigma_1^2 + \sigma_2^2) \geq \delta_t) \\
&\leq \frac{1}{2}\exp(-\frac{\delta_t^2}{2(\sigma_1^2 + \sigma_2^2)})
\end{aligned} \tag{17}$$

We now consider the sum of these terms over all t:

$$\sum_{t=0}^{\infty} \Pr(Z_b - Z_t \geq \delta_t) \leq \frac{1}{2}\sum_{t=0}^{\infty} \exp(-\frac{(ct\mu_1)^2}{2(\sigma_1^2 + \sigma_2^2)}) \tag{18}$$

$$= \frac{\sqrt{2(\sigma_1^2 + \sigma_2^2)}}{2(1-c)\mu_1} \sum_{t=0}^{\infty} \exp(-t^2) \tag{19}$$

$$\leq \frac{3\sqrt{2(\sigma_1^2 + \sigma_2^2)}}{4(1-c)\mu_1}. \tag{20}$$

We combine these to derive the final bound as desired: $E_1[T] \leq 1 + \frac{b}{(1-c)\mu_1} + \min_k \min_c \{\frac{1}{2(k+1)} \frac{b}{(1-c)\mu_1} + (k+1)\rho^{-1} + \frac{3\sqrt{2(\sigma_1^2+\sigma_2^2)}}{4(1-c)\mu_1}\}$. The bound on $E_0[T]$ follows by symmetry, with b replaced by a, and μ_1 replaced by μ_0.

□

A.3 Proof of error rate

Theorem 9 (Error Rate). *Let $d \in \{0,1\}$ be the decision output by* PrivSPRT$(X, f_1, f_2, -a, b, \sigma_1, \sigma_2, A)$. *Then the Type I error is bounded by:*

$$Pr_0[d=1] \leq \min_{k\in\mathbb{N}} \min_{c\in(0,1)} \{Q_1 + Q_2 + Q_3\}, \tag{5}$$

where $Q_1 = 2\rho_0^{-1} \exp(-\frac{2b(1-c)\mu_0}{A^2})(1 + k\exp(\frac{1}{8k})$, $Q_2 = k\exp(\frac{1}{4k+3}))$ *and* $Q_3 = \frac{\sqrt{2(\sigma_1^2+\sigma_2^2)}}{4(1-c)\mu_0}$, *and* $\rho_0 = 1 - \exp(-\frac{(1-c)\mu_0^2}{2A^2})$. *The Type II error is bounded by:*

$$Pr_1[d=0] \leq \min_{k\in\mathbb{N}} \min_{c\in(0,1)} \{W_1 + W_2 + W_3\}, \tag{6}$$

where $W_1 = 2\rho_1^{-1} \exp(-\frac{2a(1-c)\mu_1}{A^2})(1 + k\exp(\frac{1}{8k})$, $W_2 = k\exp(\frac{1}{4k+3}))$, *and* $W_3 = \frac{\sqrt{2(\sigma_1^2+\sigma_2^2)}}{4(1-c)\mu_1}$, *and* $\rho_1 = 1 - \exp(-\frac{(1-c)\mu_1^2}{2A^2})$.

Proof. The proof of error rates is based on a brute force estimation of the error probabilities: We can write the Type I error $\Pr_0[d=1]$ as a sum of probabilities of the noisy log-likelihood ratio being above the noisy threshold at time t for all $t > 0$: $\sum_{t=1}^{\infty} \Pr_0[\ell_t(A) + Z_t > b + Z_b \ \wedge \ T = t]$. We then partition the range $[1, \infty)$ into several sub-intervals and analyze them separately. We provide the detailed proof for $P_0(d=1)$. The proof is the same for $P_0(d=1)$ with b replaced by a, and μ_0 replaced by μ_1. To start, we have the following brute force estimation:

$$\begin{aligned}
P_0(d=1) &= P_0(\ell(T) + Z_T > b + Z_b) \\
&= \sum_{t=1}^{\infty} P_0(T = t, \ell(t) + Z_t > b + Z_b) \\
&\leq \sum_{t=1}^{\infty} P_0(\ell(t) + Z_t > b + Z_b) \\
&\leq \sum_{t=1}^{\infty} P_0(\ell(t) + t\mu_0 > b + t\mu_0 - \delta_t) + \Pr(Z_t - Z_b > \delta_t).
\end{aligned} \tag{21}$$

We choose $\delta_t = ct\mu_0$, where c is a constant within $(0,1)$. To simply the notation, we let γ denote $\frac{b}{(1-c)\mu_0}$, and m denote $(1-c)\mu_0$. We bound the first term in (21) using similar technique as in the proof of Theorem 8. We partition $[1, \infty)$ into four sub-intervals:

$$[0, \frac{1}{2}\gamma], \quad (\frac{1}{2}\gamma, \gamma], \quad (\gamma, 2\gamma], \quad (2\gamma, \infty).$$

We will use the following observation as the main tool. For any i and j with $i < j$, we have

$$\begin{aligned}
&\sum_{i}^{j} \exp(-\frac{(b+mt)^2}{2tA^2}) \\
\leq &\sum_{i}^{j} \exp(-\frac{b^2}{2jA^2} - \frac{bm}{A^2} - \frac{m^2 t}{2A^2}) \\
= &\exp(-\frac{b^2}{2jA^2} - \frac{bm}{A^2}) \sum_{i}^{j} \exp(-\frac{m^2 t}{2A^2}) \\
\leq &\rho^{-1} \exp(-\frac{b^2}{2jA^2} - \frac{bm}{A^2}) \exp(-\frac{m^2 i}{2A^2}),
\end{aligned} \tag{22}$$

where $\rho = 1 - \exp(-\frac{m^2}{2A^2})$. By applying (22) to the case that $i = 1$ and $j = \frac{1}{2}\gamma$, we obtain

$$\begin{aligned} S_1 &\leq \rho^{-1} \exp(-\frac{2bm}{A^2}) \exp(-\frac{m^2}{2A^2}) \\ &\leq \rho^{-1} \exp(-\frac{2bm}{A^2}), \end{aligned} \tag{23}$$

where (23) follows from the fact that $\exp(-\frac{m^2}{2A^2}) < 1$. Similarly, we have

$$S_4 \leq \rho^{-1} \exp(-\frac{2bm}{A^2}). \tag{24}$$

We further partition $(\frac{1}{2}\gamma, \gamma]$ into k sub-intervals $(\frac{k+j-1}{2k}\gamma, \frac{k+j}{2k}\gamma]$ for $j = 1, 2, \ldots, k$. We have

$$\begin{aligned} S_2 &\leq \sum_{j=1}^{k} \rho^{-1} \exp(-\frac{bm}{A^2}\frac{k}{k+j} - \frac{bm}{A^2} - \frac{bm}{2A^2}\frac{k+j-1}{2k}) \\ &= \sum_{j=1}^{k} \rho^{-1} \exp(-\frac{bm}{2A^2}(2 + (\frac{2k}{k+j} + \frac{k+j-1}{2k}))) \\ &\leq \sum_{j=1}^{k} \rho^{-1} \exp(-\frac{bm}{2A^2}(2 + 2 - \frac{1}{2k})) \qquad (25) \\ &= k\rho^{-1} \exp(-\frac{2bm}{A^2}\frac{8k-1}{8k}). \qquad (26) \end{aligned}$$

Similarly, we have

$$S_3 \leq k\rho^{-1} \exp(-\frac{2bm}{A^2}\frac{4k+3}{4k+3}). \tag{27}$$

Next we bound the error from the added Gaussian noise.

$$\begin{aligned} \sum_{t=1}^{\infty} \Pr(Z_t - Z_b \geq \delta_t) &\leq \frac{1}{2}\sum_{t=0}^{\infty} \exp(-\frac{(ct\mu_0)^2}{2(\sigma_1^2 + \sigma_2^2)}) \qquad (28) \\ &= \frac{\sqrt{2(\sigma_1^2 + \sigma_2^2)}}{2(1-c)\mu_0} \sum_{t=1}^{\infty} \exp(-t^2) \qquad (29) \\ &\leq \frac{\sqrt{2(\sigma_1^2 + \sigma_2^2)}}{4(1-c)\mu_0} \qquad (30) \end{aligned}$$

Combining (23), (26), (27), (24) and (30), we obtain

$$P_0(d=1) \leq \min_k \min_c \{2\rho^{-1} \exp(-\frac{2b(1-c)\mu_0}{A^2})(1 + k\exp(\frac{1}{8k}) + k\exp(\frac{1}{4k+3})) + \frac{\sqrt{2(\sigma_1^2 + \sigma_2^2)}}{4(1-c)\mu_0}\}$$

□

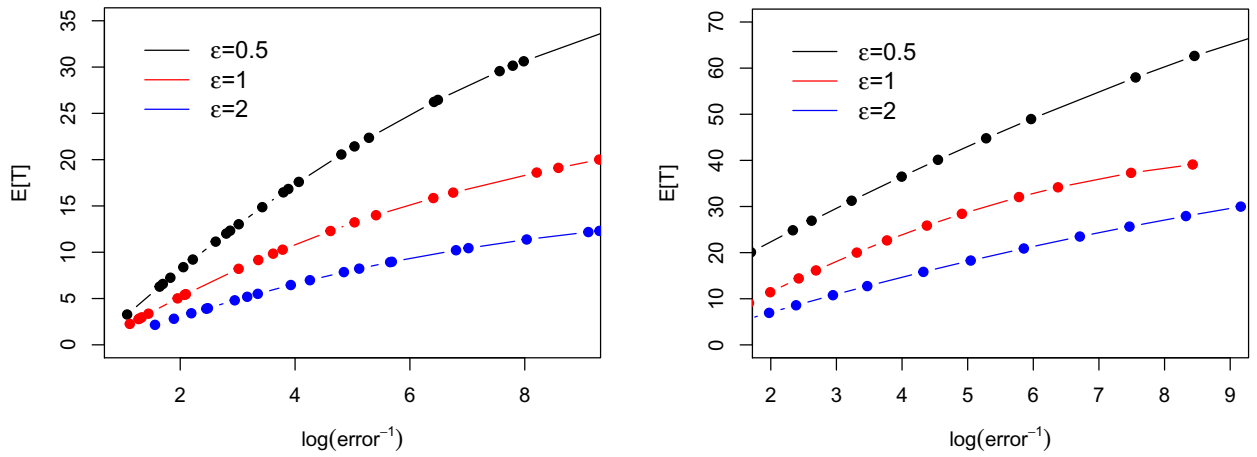
B Additional Experiments

B.1 Testing on Gaussian Data

In this section, our experiments focus on testing means of Gaussian data, where $x_1, x_2, \ldots \sim N(\mu, 1)$ are sampled i.i.d. from a Gaussian distribution with mean μ. We again consider two different scenarios: large distance between the null and alternative hypotheses on μ corresponding to $H_0 : \mu = 0$ against $H_1 : \mu = 2$, and a small distance between the null and alternative hypotheses on μ corresponding to $H_0 : \mu = 0$ against $H_1 : \mu = 1$. We will denote the expected sample size as $\mathbb{E}[T]$ since $\mathbb{E}_0[T]$ and $\mathbb{E}_1[T]$ are identical for Gaussian data, and similarly, we denote

the Type I error and Type II error as errors. We use two $(2, \epsilon'/2)$-RDP Gaussian mechanisms as our private mechanisms in PrivSPRT, corresponding to $\sigma_1 = 2\sqrt{2}A/\epsilon'$ and $\sigma_2 = 4A/\epsilon'$. We note that this setting offers an $(\epsilon' + 2\log(1/\delta), \delta)$-differential privacy guarantee, where the additional term $2\log(1/\delta)$ is small.

In Figure 2, we plot the expected sample size $\mathbb{E}[T]$ against the log scale of 1/error, and we vary the privacy parameter $\epsilon' = 0.5, 1, 2$. This experiment is conducted under the setting where we fix the truncation threshold $A = 0.5$ and vary the decision threshold a, b. For each simulation, we repeat the process 10^5 times and report average performance. As in the case with Bernoulli data, we see that we experience a larger expected sample size for a given Type I and Type II error constraint as ϵ decreases. Additionally, we need fewer samples to distinguish μ for a large distance regime (Left vs. Right, note the different scales on the y-axes).



(a) Large distance

(b) Small distance

Figure 2: Three-way trade-off between privacy, expected sample size, and error rate. For large distance (left), we are testing $H_0 : \mu = 0$ against $H_1 : \mu = 2$; for small distance (right), we are testing $H_0 : \mu = 0$ against $H_1 : \mu = 1$.

We again conduct experiments to further validate our theoretical results empirically in the Gaussian setting. In Table 2 (left), we vary the truncation parameter $A = 0.5, 1, 2, 5$ and the privacy parameter $\epsilon' = 0.5, 1, 2, \infty$. For each fixed A and ϵ', we choose thresholds a, b through Monte Carlo simulation with importance sampling to control Type I and Type II errors at the 0.05 level. Similar to the results for testing the Bernoulli parameter, the thresholds are almost linear with respect to A. The expected sample sizes $\mathbb{E}_0[T]$ and $\mathbb{E}_1[T]$ are almost the same for all $A = 0.5, 1, 2$, and $\mathbb{E}_0[T]$ and $\mathbb{E}_1[T]$ increase for a larger $A = 5$, because the noise added for privacy dominates the information provided by the log-likelihood ratios. This suggests that a relatively small A is preferred, and as long as A is not too large, it has little impact on the performance. Moreover, we observe that $a/\mathbb{E}_0[T]$ (resp. $b/\mathbb{E}_1[T]$) decreases as ϵ' increases for weaker privacy, as the additional cost term in Theorem 8 decreases for less noise.

B.2 Using the standard AboveThresh.

To compare against the performance of our PrivSPRT, we also conduct experiments for testing means of Gaussian data using the original AboveThresh algorithm with Laplace noise that satisfies $\epsilon/2$-differential privacy. We now vary the truncation parameter $A = 0.5, 1, 2, 5$, and choose the thresholds such that the Type I and Type II error are below 0.05. The results are presented in Table 2 (right). Table 2 shows that using the original AboveThresh algorithm with Laplace noise results in much larger expected sample sizes, given that the Type I and Type II errors are fixed at the 0.05 level. We note that although the overall privacy cost for PrivSPRT is slightly larger, PrivSPRT provides a better trade-off between privacy and accuracy.

We also empirically study the overshoot property when adding Laplace noise. We again consider testing $H_0 : \theta = 0.7$

against $H_1 : \theta = 0.2$ for Bernoulli data. We choose this setting because $\mu_0 \neq \mu_1$, to have a comprehensive view of $\mathbb{E}_0[T]$ and $\mathbb{E}_1[T]$, and the Type I and Type II errors. We now fix the truncation parameter $A = 0.5$, and vary the privacy parameter $\epsilon = 0.5, 1, 2$ and the thresholds $a, b = 10, 20, 40$. The results are presented in Table 3.

Table 2: Numerical values of expected sample size $\mathbb{E}[T]$, error rates for testing the Gaussian mean using PrivSPRT(left) and the original AboveThresh with Laplace noise (right). The thresholds a, b are chosen to control Type I and Type II error at 0.05 (within the Monte Carlo simulation errors).

PrivSPRTwith Gaussian noise					AboveThresh with Laplace noise				
A	ϵ'	$a = b$	error rates	$\mathbb{E}[T]$	A	ϵ	$a = b$	error rates	$\mathbb{E}[T]$
0.5	0.5	9	0.05	12.547	0.5	0.5	28	0.05	22.821
	1	4		7.298		1	12		12.621
	2	2.1		4.890		2	6.5		10.482
1	0.5	18		12.485	1	0.5	59		26.032
	1	8.2		7.367		1	29		17.165
	2	4		4.792		2	15		13.291
2	0.5	36		13.333	2	0.5	112		23.581
	1	16		7.460		1	52		15.438
	2	8		5.000		2	28		12.564
5	0.5	90		16.943	5	0.5	270		24.426
	1	40		10.156		1	140		21.067
	2	98		6.190		2	70		15.872
	∞	2		1.793					

On the theoretical side, we should expect the expected sample size to be $O(b/\mu_1)$ for non-private SPRT. However, we see from Table 3 that the expected sample sizes are nonlinear with respect to the thresholds for strong privacy ($\epsilon = 0.5, 1$), which is no longer consistent with the CLT theorem for non-private SPRT. In contrast, we observe from Table 1 that $\mathbb{E}_0[T]$ (resp. $\mathbb{E}_1[T]$) is $O(a/\mu_0)$ (resp. $O(b/\mu_1)$) in Section 4 when adding Gaussian noise in PrivSPRT. Intuitively, it appears that the overshoot analysis when adding Laplace noise relies heavily on the additional noise, rather than the statistical information provided by log-likelihood ratios. Characterizing the relevant statistical properties when adding Laplace noise requires new tools, which we leave as future work for the privacy and statistics communities.

Table 3: Numerical values of expected sample sizes $\mathbb{E}_0[T]$ and $\mathbb{E}_1[T]$, Type I error and Type II error for testing Bernoulli parameter using the original AboveThresh algorithm with Laplace noise.

$a = b$	ϵ	Type I	Type II	$\mathbb{E}_0[T]$	$\mathbb{E}_1[T]$
10	0.5	0.3634	0.3562	3.537	3.762
	1	0.2184	0.3132	9.246	10.399
	2	0.0181	0.2185	22.317	29.035
20	0.5	0.2577	0.1750	11.824	11.62
	1	0.0235	0.0140	35.353	43.121
	2	1.03e-05	3.53e-05	55.136	77.450
40	0.5	0.0164	0.0266	51.257	66.529
	1	8.11e-08	2.4e-04	99.026	144.114
	2	1.04e-20	3.79e-19	121.27	179.172