The Structure of Optimal Private Tests for Simple Hypotheses*

Clément L. Canonne
Stanford University
USA
ccanonne@cs.stanford.edu

Gautam Kamath
Simons Institute for the Theory of Computing
USA
g@csail.mit.edu

Audra McMillan
Boston University and Northeastern University
USA
audram@bu.edu

Adam Smith
Boston University
USA
ads22@bu.edu

Jonathan Ullman
Northeastern University
USA
jullman@ccs.neu.edu

ABSTRACT

Hypothesis testing plays a central role in statistical inference, and is used in many settings where privacy concerns are paramount. This work answers a basic question about privately testing simple hypotheses: given two distributions P and Q, and a privacy level ε, how many i.i.d. samples are needed to distinguish P from Q subject to ε-differential privacy, and what sort of tests have optimal sample complexity? Specifically, we characterize this sample complexity up to constant factors in terms of the structure of P and Q and the privacy level ε, and show that this sample complexity is achieved by a certain randomized and clamped variant of the log-likelihood ratio test. Our result is an analogue of the classical Neyman–Pearson lemma in the setting of private hypothesis testing. We also give an application of our result to the private change-point detection. Our characterization applies more generally to hypothesis tests satisfying essentially any notion of algorithmic stability, which is known to imply strong generalization bounds in adaptive data analysis, and thus our results have applications even when privacy is not a primary concern.

CCS CONCEPTS

• **Mathematics of computing → Hypothesis testing and confidence interval computation**; • **Security and privacy**; • **Theory of computation → Design and analysis of algorithms**;

KEYWORDS

differential privacy, hypothesis testing

ACM Reference Format:
Clément L. Canonne, Gautam Kamath, Audra McMillan, Adam Smith, and Jonathan Ullman. 2019. The Structure of Optimal Private Tests for Simple Hypotheses. In *Proceedings of the 51st Annual ACM SIGACT Symposium on the Theory of Computing (STOC '19), June 23–26, 2019, Phoenix, AZ, USA.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3313276.3316336

*In memory of Stephen E. Fienberg (1942–2016). A full version of this paper is available as [19].

STOC '19, June 23–26, 2019, Phoenix, AZ, USA

ACM ISBN 978-1-4503-6705-9/19/06…$15.00
https://doi.org/10.1145/3313276.3316336

1 INTRODUCTION

Hypothesis testing plays a central role in statistical inference, analogous to that of decision or promise problems in computability and complexity theory. A hypothesis testing problem is specified by two disjoint sets of probability distributions over the same set, called hypotheses, $\mathcal{H}_0$ and $\mathcal{H}_1$. An algorithm T for this problem, called a *hypothesis test*, is given a sample x from an unknown distribution P, with the requirement that $T(x)$ should, with high probability, output "0" if $P \in \mathcal{H}_0$, and "1" if $P \in \mathcal{H}_1$. There is no requirement for distributions outside of $\mathcal{H}_0 \cup \mathcal{H}_1$. In computer science, such problems sometimes go by the name *distribution property testing*.

Hypothesis testing problems are important in their own right, as they formalize yes-or-no questions about an underlying population based on a randomly drawn sample, such as whether education strongly influences life expectancy, or whether a particular medical treatment is effective. Successful hypothesis tests with high degrees of confidence remain the gold standard for publication in top journals in the physical and social sciences. Hypothesis testing problems are also important in the theory of statistics and machine learning, as many lower bounds for estimation and optimization problems are obtained by reducing from hypothesis testing.

This paper aims to understand the structure and sample complexity of optimal hypothesis tests subject to strong privacy guarantees. Large collections of personal information are now ubiquitous, but their use for effective scientific discovery remains limited by concerns about privacy. In addition to the well-understood settings of data collected during scientific studies, such as clinical experiments and surveys, many other data sources where privacy concerns are paramount are now being tapped for socially beneficial analysis, such as Social Science One [70], which aims to allow access to data collected by Facebook and similar companies.

We study algorithms that satisfy *differential privacy (DP)* [32], a restriction on the algorithm that ensures meaningful privacy guarantees against an adversary with arbitrary side information [47]. Differential privacy has come to be the de facto standard for the analysis of private data, used as a measure of privacy for data analysis systems at Google [36], Apple [25], and the U.S. Census Bureau [24]. Differential privacy and related distributional notions of *algorithmic stability* can be crucial for statistical validity even

when confidentiality is not a direct concern, as they provide generalization guarantees in an adaptive setting [29].

Consider an algorithm that takes a set of data points from a set $\mathcal{X}$—where each point belongs to some individual—and produces some public output. We say the algorithm is differentially private if no single data point can significantly impact the distribution on outputs. Formally, we say two data sets $x, x' \in \mathcal{X}^n$ of the same size are *neighbors* if they differ in at most one entry.

Definition 1.1 ([32]). A randomized algorithm T taking inputs in $\mathcal{X}^*$ and returning random outputs in a space with event set $\mathcal{S}$ is *ε-differentially private* if for all $n \geq 1$, for all neighboring data sets $x, x' \in \mathcal{X}^n$, and for all events $S \in \mathcal{S}$, $\mathbb{P}[T(x) \in S] \leq e^{\varepsilon}\mathbb{P}[T(x') \in S]$. For the special case of tests returning output in $\{0, 1\}$, the output distribution is characterized by the probability of returning "1". Letting $g(x) = \mathbb{P}[T(x) = 1]$, we can equivalently require that

$$\max\left(\frac{g(x)}{g(x')}, \frac{1-g(x)}{1-g(x')}\right) \leq e^{\varepsilon}.$$

For algorithms with binary outputs, this definition is essentially equivalent to all other commonly studied notions of privacy and distributional algorithmic stability (see "Connections to Algorithmic Stability", below).

Contribution: The Sample Complexity of Private Tests for Simple Hypotheses. We focus on the setting of i.i.d. data and singleton hypotheses $\mathcal{H}_0, \mathcal{H}_1$, which are called *simple hypotheses*. The algorithm is given a sample of n points $x_1, \ldots, x_n$ drawn i.i.d. from one of two distributions, P or Q, and attempts to determine which one generated the input. That is, $\mathcal{H}_0 = \{P^n\}$ and $\mathcal{H}_1 = \{Q^n\}$. We investigate the following question.

> *Given two distributions P and Q and a privacy parameter $\varepsilon > 0$, what is the minimum number of samples (denoted $SC_{\varepsilon}^{P,Q}$) needed for an ε-differentially private test to reliably distinguish P from Q, and what are optimal private tests?*

These questions are well understood in the classical, nonprivate setting. The number of samples needed to distinguish P from Q is $\Theta(1/H^2(P, Q))$, where H^2 denotes the squared Hellinger distance (3).[1] Furthermore, by the Neyman–Pearson lemma, the exactly optimal test consists of computing the likelihood ratio $P^n(x)/Q^n(x)$ and comparing it to some threshold.

We give analogous results in the private setting. First, we give a closed-form expression that characterizes the sample complexity up to universal multiplicative constants, and highlights the range of ε in which private tests use a similar amount of data to the best nonprivate ones. We also give a specific, simple test that achieves that sample complexity. Roughly, the test makes a noisy decision based on a "clamped" log likelihood ratio in which the influence of each data point is limited. The sample complexity has the form $\Theta(1/\mathrm{adv}_1)$, where adv_1 is the advantage of the test over random guessing on a sample of size $n = 1$. The optimal test and its sample complexity are described in Theorem 1.2.

Our result provides the first instance-specific characterization of a statistical problem's complexity for differentially private algorithms. Understanding the private sample complexity of statistical problems is delicate. We know there are regimes where statistical problems can be solved privately "for free" asymptotically (e.g. [20, 32, 45, 69]) and others where there is a significant cost, even for relaxed definitions of privacy (e.g. [15, 35]), and we remain far from a general characterization of the statistical cost of privacy. Duchi, Jordan, and Wainwright [26] give a characterization for the special case of simple tests by *local* differentially private algorithms, a more restricted setting where samples are randomized individually, and the test makes a decision based on these randomized samples. Our characterization in the general case is more involved, as it exhibits several distinct regimes for the parameter ε.

Our analysis relies on a number of tools of independent interest: a characterization of private hypothesis testing in terms of couplings between distributions on $\mathcal{X}^n$, and a novel interpretation of Hellinger distance as the advantage over random guessing of a specific, randomized likelihood ratio test.

The Importance of Simple Hypotheses. Many of the hypotheses that arise in application are not simple, but are so-called *composite hypotheses*. For example, deciding if two features are independent or far from it involves sets $\mathcal{H}_0$ and $\mathcal{H}_1$ each containing many distributions. Yet many of those tests can be reduced to simple ones. For example, deciding if the mean of a Gaussian is less than 0 or greater than 1 can be reduced to testing if the mean is either 0 or 1. Furthermore, simple tests arise in lower bounds for estimation—the well-known characterization of parametric estimation in terms of Fisher information is obtained by showing that the Fisher information measures variability in the Hellinger distance and then employing the Hellinger-based characterization of nonprivate simple tests (e.g. [12, Chap. II.31.2, p.180]).

Our characterization of private testing implies similar lower bounds for estimation (along the lines of lower bounds of Duchi and Ruan [27] in the local model of differential privacy).

Connection to Algorithmic Stability. For hypothesis tests with constant error probabilities, sample complexity bounds for differential privacy are equivalent, up to constant factors, to sample complexity bounds for other notions of distributional algorithmic stability, such as (ε, δ)-DP [31], concentrated DP [14, 34], KL- and TV-stability [9, 77] (see [3, Lemma 5]). (Briefly: if we ensure that $\Pr(T(x) = 1) \in [0.01, 0.99]$ for all x, then an additive change of ε corresponds to an multiplicative change of $1 \pm O(\varepsilon)$, and vice-versa.) Consequently, our results imply optimal tests for use in conjunction with stability-based generalization bounds for adaptive data analysis, which has generated significant interest in recent years [9, 28–30, 37, 38, 64, 65, 78].

1.1 Hypothesis Testing

To put our result in context, we review classical results about nonprivate hypothesis testing. Let P and Q be two probability distributions over an arbitrary domain $\mathcal{X}$. A *hypothesis test* $K \colon \mathcal{X}^* \to \{\text{"}P\text{"}, \text{"}Q\text{"}\}$ is an algorithm that takes a set of samples $x \in \mathcal{X}^*$ and attempts to determine if it was drawn from P or Q. Define the *advantage* of a test K given n samples as

$$\mathrm{adv}_n(K) = \mathop{\mathbb{P}}_{x \sim P^n}[K(x) = \text{"}P\text{"}] - \mathop{\mathbb{P}}_{x \sim Q^n}[K(x) = \text{"}P\text{"}]. \qquad (1)$$

[1] This statement is folklore, but see, e.g., [8] for the lower bound, [18] or Corollary 2.2 for the upper bound.

We say that K *distinguishes* P *from* Q *with sample complexity* $SC^{P,Q}(K)$ if for every $n \geq SC^{P,Q}(K)$, $\mathrm{adv}_n(K) \geq 2/3$. We say $SC^{P,Q} = \min_K SC^{P,Q}(K)$ is the *sample complexity of distinguishing* P *from* Q.

Most hypothesis tests are based on some real-valued *test statistic* $S\colon \mathcal{X}^* \to \mathbb{R}$ where

$$K_S(x) = \begin{cases} \text{“}P\text{”} & \text{if } S(X) \geq \kappa \\ \text{“}Q\text{”} & \text{otherwise} \end{cases}$$

for some threshold κ. We will sometimes abuse notation and use the test statistic S and the implied hypothesis test K_S interchangeably.

The classical Neyman–Pearson Lemma says that the exact optimal test[2] for distinguishing P, Q is the *log-likelihood ratio test* given by the test statistic

$$\mathrm{LLR}(x_1, \ldots, x_n) = \sum_{i=1}^{n} \log \frac{P(x_i)}{Q(x_i)}\,. \tag{2}$$

Another classical result says that the optimal sample complexity is characterized by the *squared Hellinger distance* between P, Q, which is defined as

$$H^2(P,Q) = \frac{1}{2}\int_{\mathcal{X}} \left(\sqrt{P(x)} - \sqrt{Q(x)}\right)^2 dx. \tag{3}$$

Specifically, $SC^{P,Q} = SC^{P,Q}(\mathrm{LLR}) = \Theta(1/H^2(P,Q))$. Note that the same metric provides upper and lower bounds on the sample complexity.

1.2 Our Results

Our main result is an approximate characterization of the sample complexity of ε-differentially private tests for distinguishing P and Q. Analogous to the non-private case, we will write $SC_\varepsilon^{P,Q} = \min_{\varepsilon\text{-DP } K} SC^{P,Q}(K)$ to denote the *sample complexity of* ε*-differentially privately (*ε*-DP) distinguishing* P *from* Q, and we characterize this quantity up to constant factors in terms of the structure of P, Q and the privacy parameter ε. Specifically, we show that a *privatized clamped log-likelihood ratio test* is optimal up to constant factors. Privacy may be achieved through either the Laplace or Exponential mechanism, and we will prove optimality of both methods.

For $b \geq a$, we define the *clamped log-likelihood ratio statistic*,

$$\mathrm{cLLR}_{a,b}(x) = \sum_i \left[\log \frac{P(x_i)}{Q(x_i)}\right]_a^b,$$

where $[\cdot]_a^b$ denotes the projection onto the interval $[a, b]$ (that is, $[z]_a^b = \max(a, \min(z, b))$).

Define the *soft clamped log-likelihood test*:

$$\mathrm{scLLR}_{a,b}(x) = \begin{cases} P & \text{with probability } \propto \exp(\frac{1}{2}\,\mathrm{cLLR}_{a,b}(x)) \\ Q & \text{with probability } \propto 1 \end{cases}$$

The test scLLR is an instance of the *exponential mechanism* [54], and thus $\mathrm{scLLR}_{a,b}$ satisfies ε-differential privacy for $\varepsilon = \frac{b-a}{2}$.

[2] More precisely, given any test K, there is a setting of the threshold κ for the log-likelihood ratio test that weakly dominates K, meaning that $\mathbb{P}_{x\sim Q}[\mathrm{LLR}(x) = P] \leq \mathbb{P}_{x\sim Q}[K(x) = P]$ and $\mathbb{P}_{x\sim P}[\mathrm{LLR}(x) = Q] \leq \mathbb{P}_{x\sim P}[K(x) = Q]$ (keeping the true positive rates $\mathbb{P}_{x\sim P}[K(x) = P], \mathbb{P}_{x\sim Q}[K(x) = Q]$ fixed). One may need to randomize the decision when $S(X) = \kappa$ to achieve some tradeoffs between false negative and positive rates.

Similarly, define the *noisy clamped log-likelihood ratio test*:

$$\mathrm{ncLLR}_{a,b}(X) = \begin{cases} P & \text{if } \mathrm{cLLR}_{a,b}(x) + \mathrm{Lap}\left(\frac{1}{\varepsilon(b-a)}\right) > 0 \\ Q & \text{otherwise} \end{cases}$$

The test ncLLR is an instance of postprocessing the Laplace mechanism [32], and satisfies ε-differential privacy.

Our main result is that, for every P, Q, and every ε, the tests $\mathrm{scLLR}_{-\varepsilon',\varepsilon}$ and $\mathrm{ncLLR}_{-\varepsilon',\varepsilon}$ are optimal up to constant factors, for some appropriate $0 \leq \varepsilon' \leq \varepsilon$. To state the result more precisely, we introduce some additional notation. First define

$$\tau = \tau(P,Q) \triangleq \max\left\{\int_{\mathcal{X}} \max\{P(x) - e^{\varepsilon}Q(x), 0\}\,dx, \int_{\mathcal{X}} \max\{Q(x) - e^{\varepsilon}P(x), 0\}\,dx\right\}, \tag{4}$$

and assume without loss of generality that $\tau = \int_{\mathcal{X}} \max\{P(x) - e^{\varepsilon}Q(x), 0\}\,dx$, which we assume for the remainder of this work.[3] Next, let $0 \leq \varepsilon' \leq \varepsilon$ be the largest value such that

$$\int_{\mathcal{X}} \max\{P(x) - e^{\varepsilon}Q(x), 0\}\,dx = \int_{\mathcal{X}} \max\{Q(x) - e^{\varepsilon'}P(x), 0\}\,dx = \tau,$$

whose existence is guaranteed by a continuity argument (a formal argument is in the full version). We give an illustration of the definition of τ and ε' in Figure 1. Finally, define $\tilde{P} = \min\{e^{\varepsilon}Q, P\}$ and $\tilde{Q} = \min\{e^{\varepsilon'}P, Q\}$ and normalize by $(1-\tau)$ to obtain distributions

$$P' = \tilde{P}/(1-\tau) \quad \text{and} \quad Q' = \tilde{Q}/(1-\tau). \tag{5}$$

The distributions P', Q' are such that

$$-\varepsilon' \leq \log \frac{P'(x)}{Q'(x)} \leq \varepsilon\,,$$

and

$$P = (1-\tau)P' + \tau P'' \quad \text{and} \quad Q = (1-\tau)Q' + \tau Q'',$$

where P'' and Q'' are distributions with disjoint support. The quantity τ is the smallest possible number for which such a representation is possible. With these definitions in hand, we can now state our main result.

Theorem 1.2. *For every pair of distributions* P, Q*, and every* $\varepsilon > 0$*, the optimal sample complexity for* ε*-differentially private tests is achieved by either the soft or noisy clamped log-likelihood test, and satisfies*

$$SC_\varepsilon^{P,Q} = \Theta(SC^{P,Q}(\mathrm{ncLLR}_{-\varepsilon',\varepsilon})) = \Theta(SC^{P,Q}(\mathrm{scLLR}_{-\varepsilon',\varepsilon})) = \Theta\left(\frac{1}{\varepsilon\tau(P,Q) + (1-\tau)H^2(P',Q')}\right) = \Theta\left(\frac{1}{\mathrm{adv}_1(\mathrm{scLLR}_{-\varepsilon',\varepsilon})}\right).$$

When $\varepsilon \geq \max_x |\log P(x)/Q(x)|$, Theorem 1.2 reduces to $SC_\varepsilon^{P,Q} = \Theta\left(\frac{1}{H^2(P,Q)}\right)$, which is the sample complexity for distinguishing between P and Q in the non-private setting. This implies that we get privacy for free asymptotically in this parameter regime. We will

[3] For $\alpha \geq 0$, the quantity $D_\alpha(P\|Q) = \int \max(P(x) - \alpha Q(x), 0)\,dx$ is an f-divergence and has appeared in the literature before under the names *α-divergence*, *hockey-stick divergence*, or *elementary divergence* [6, 52] (for $\alpha = 1$, one obtains the usual total variation distance). Thus, τ is the maximum of the divergences $D_{e^\varepsilon}(P\|Q)$ and $D_{e^\varepsilon}(Q\|P)$. It can also be described as the smallest value δ such that P and Q are (ε, δ)-indistinguishable [33].

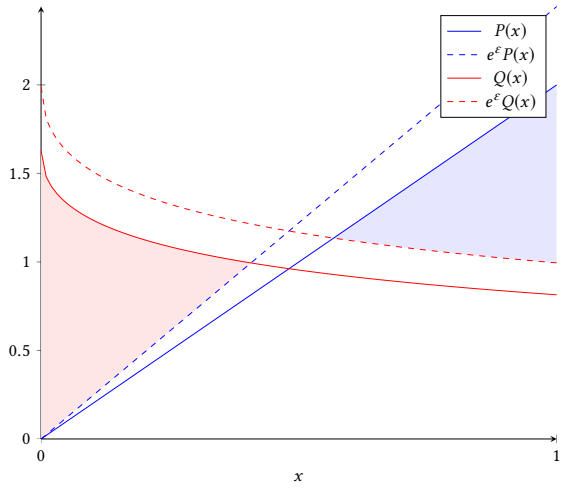

Figure 1: An illustration of the definition of τ, for $\varepsilon = 0.2$ and for two densities P, Q over $\mathcal{X} = [0,1]$. The blue shaded area represents $\int \max\{P(x) - e^\varepsilon Q(x), 0\}\, dx$, while the red corresponds to $\int \max\{Q(x) - e^\varepsilon P(x), 0\}\, dx$. The larger of these two is $\tau(P,Q)$. If the blue area is larger than the red area, the definition of ε' corresponds to lowering the dotted blue curve until the two are the same size.

focus on proving the first equality, a proof of the second appears in the full version.

Comparison to Known Bounds. For $\varepsilon < 1$, the bounds

$$\frac{1}{H^2(P,Q)} \le SC_\varepsilon^{P,Q} \le O\left(\frac{1}{\varepsilon H^2(P,Q)}\right)$$

follow directly from the non-private sample complexity. Namely, the lower bound is the non-private sample complexity and the upper bound is obtain by applying the *sample-and-aggregate* technique [59] to the optimal non-private test. They can be recovered from Theorem 1.2 by noting that

$$\begin{aligned}
\varepsilon H^2(P,Q) &= \frac{\varepsilon}{2}\|\sqrt{P} - \sqrt{Q}\|_2^2 \\
&= O\left(\varepsilon \left\|\sqrt{P} - \sqrt{\tilde{P}}\right\|_2^2 + \varepsilon \left\|\sqrt{\tilde{P}} - \sqrt{\tilde{Q}}\right\|_2^2 + \varepsilon \left\|\sqrt{\tilde{Q}} - \sqrt{Q}\right\|_2^2\right) \\
&= O(\varepsilon\tau + \varepsilon(1-\tau)H^2(P',Q') + \varepsilon\tau) \\
&= O(\varepsilon\tau + (1-\tau)H^2(P',Q'))
\end{aligned}$$

and

$$\begin{aligned}
&\varepsilon\tau + (1-\tau)H^2(P',Q') \\
&= \varepsilon \cdot \int_{\mathcal{S}} |P(x) - e^\varepsilon Q(x)|\, dx + \|\sqrt{\tilde{P}} - \sqrt{\tilde{Q}}\|_2^2 \\
&\le \varepsilon \cdot \int_{\mathcal{S}} |P(x) - Q(x)|\, dx + \|\sqrt{P} - \sqrt{Q}\|_2^2 \\
&\le \varepsilon \cdot \frac{1 + e^{\varepsilon/2}}{e^{\varepsilon/2} - 1} \cdot \int_{\mathcal{S}} (\sqrt{P(x)} - \sqrt{Q(x)})^2\, dx + H^2(P,Q) \\
&= O(H^2(P,Q)),
\end{aligned}$$

where $\mathcal{S} = \{\, x : \; P(x) - e^\varepsilon Q(x) > 0 \,\}$.

1.2.1 Application: Private Change-Point Detection. As an application of our result, we obtain optimal private algorithms for *change-point detection*. Given distributions P and Q, an algorithm solving *offline change-point detection for P and Q* takes a stream $x = (x_1, x_2, \ldots, x_n) \in \mathcal{X}^n$ with the guarantee that the there is an index k^* such that first k^* elements are sampled i.i.d. from P and the latter elements are sampled i.i.d. from Q, and attempts to output $\hat{k} \approx k^*$. We can also consider an online variant where elements x_i arrive one at a time.

Change-point detection has a long history in statistics and information theory (e.g. [50, 51, 53, 55, 56, 60–63, 67, 68, 73]). Cummings *et al.* [22] recently gave the first private algorithms for change-point detection. Their algorithms are based on a private version of the log-likelihood ratio, and in cases where the log-likelihood ratio is not strictly bounded, they relax to a weaker distributional variant of differential privacy. Using Theorem 1.2, we can achieve the standard worst-case notion of differential privacy, and to achieve optimal error bounds for every P, Q.

Theorem 1.3 (Informal). *For every pair of distributions P and Q, and every $\varepsilon > 0$, there is an ε-differentially private algorithm that solves offline change-point detection for P and Q such that, with probability at least $9/10$, $|\hat{k} - k^*| = O(SC_\varepsilon^{P,Q})$.*

The expected error in this result is optimal up to constant factors for every pair P, Q, as one can easily show that the error must be at least $\Omega(SC_\varepsilon^{P,Q})$. Theorem 1.3 can be extended to give an arbitrarily small probability β of failure, and can be extended to the online change-point detection problem, although with more complex accuracy guarantees. Our algorithm introduces a general reduction from private change-point detection for families of distributions $\mathcal{H}_0$ and $\mathcal{H}_1$ to private hypothesis testing for the same family, which we believe to be of independent interest.

1.3 Techniques

First Attempts. A folklore result based on the sample-and-aggregate paradigm [59] shows that for every P, Q and every $\varepsilon > 0$, $SC_\varepsilon^{P,Q} = O(\frac{1}{\varepsilon} SC^{P,Q})$, meaning privacy comes at a cost of *at most* $O(\frac{1}{\varepsilon})$.[4] However, there are many examples where $SC_\varepsilon^{P,Q} = O(SC^{P,Q})$ even when $\varepsilon = o(1)$, and understanding this phenomenon is crucial.

A few illustrative pairs of distributions P, Q will serve to demonstrate the difficulties that go into formulating and proving Theorem 1.2. First, consider the domain $\mathcal{X} = \{0,1\}$ of size two (i.e. Bernoulli distributions). To distinguish $P = \mathrm{Ber}(\frac{1+\alpha}{2})$ from $Q = \mathrm{Ber}(\frac{1-\alpha}{2})$, the optimal non-private test statistic is $S(x) = \sum_i x_i$, which requires $\Theta(\frac{1}{\alpha^2}) = \Theta(\frac{1}{H^2(P,Q)})$ samples.

To make the test differentially private, one can use a soft version of the test that outputs "P" with probability proportional to $\exp(\varepsilon \sum_i x_i)$ and "Q" with probability proportional to $\exp(\varepsilon \sum_i 1 - x_i)$. This private test has sample complexity $\Theta(\frac{1}{\alpha^2} + \frac{1}{\alpha\varepsilon})$, which is optimal. In contrast, if $P = \mathrm{Ber}(0)$ and $Q = \mathrm{Ber}(\alpha)$, then the non-private sample complexity becomes $\Theta(\frac{1}{\alpha})$, and the optimal private sample complexity becomes $\Theta(\frac{1}{\alpha\varepsilon})$. In general, for $\mathcal{X} = \{0,1\}$, one

[4] See, e.g., [16] for a proof.

can show that

$$SC_\varepsilon^{P,Q} = \Theta\left(\frac{1}{H^2(P,Q)} + \frac{1}{\varepsilon TV(P,Q)}\right). \tag{6}$$

The sample complexity of testing Bernoulli distributions (6) already demonstrates an important phenomenon in private hypothesis testing—for many distributions P, Q, there is a "phase transition" where the sample complexity takes one form when ε is sufficiently large and another when ε is sufficiently small, and often the sample complexity in the "large ε regime" is equal to the non-private complexity up to lower order terms. A key challenge in obtaining Theorem 1.2 is to understand these transitions, and to understand the sample complexity in each regime.

Since each of the terms in (6) is a straightforward lower bound on the sample complexity of private testing, one might conjecture that (6) holds for every pair of distributions. However, our next illustrative example shows that this conjecture is false even for the domain $\mathcal{X} = \{0, 1, 2\}$. Consider the two densities

$$P = (0, 0.5, 0.5) \quad \text{and} \quad Q = (2\alpha^{3/2}, 0.5 + \alpha - \alpha^{3/2}, 0.5 - \alpha - \alpha^{3/2}).$$

For these distributions, the log-likelihood ratio statistic is roughly equivalent to the statistic that counts the number of occurrences of "0," $S_{\{0\}}(x) = \sum_i \mathbb{1}\{x_i = 0\}$, and has sample complexity $\Theta(\frac{1}{\alpha^{3/2}})$. For this pair of distributions, the optimal private test depends on the relationship between α and ε. One such test is to simply use the soft version of $S_{\{0\}}$, and the other is to use the soft version of the test $S_{\{0,1\}} = \sum_i \mathbb{1}\{x_i \in \{0, 1\}\}$ that counts occurrences of the first two elements. One can prove that the better of these two tests is optimal up to constant factors, giving

$$SC_\varepsilon^{P,Q} = \Theta\left(\min\left\{\frac{1}{\alpha^{3/2}\varepsilon}, \frac{1}{\alpha^2} + \frac{1}{\alpha\varepsilon}\right\}\right).$$

For these distributions, (6) reduces to $\Theta(\frac{1}{\alpha^{3/2}} + \frac{1}{\alpha\varepsilon})$, so these distributions show that the optimal sample complexity can be much larger than (6). Moreover, these distributions exhibit that the optimal sample complexity can vary with ε in complex ways, making several transitions and never matching the non-private complexity unless α or ε is constant.

Key Ingredients. The second example above demonstrates that the optimal test itself can vary with ε in an intricate fashion, which makes it difficult to construct a single test for which we can prove matching upper and lower bounds on the sample complexity. The use of the clamped log-likelihood test arose out of an attempt to find a single test that is optimal for the second pair of distributions P and Q, and relies on a few crucial technical ingredients.

First, our upper bound on sample complexity relies on a key observation that the Hellinger distance between two distributions is exactly the advantage, adv_1, of a soft log-likelihood ratio test sLLR on a sample of size 1. The sLLR test is a randomized test that outputs P with probability proportional to $\sqrt{P^n(x)/Q^n(x)} = e^{LLR(x)/2}$, where $\mathrm{LLR}(x) = \sum_i \log\frac{P(x_i)}{Q(x_i)}$, and Q with probability proportional to 1. This characterization of $H^2(P, Q)$ as the advantage of the sLLR tester may be of independent interest.

This observation is crucial for our work, because it implies that sLLR is ε-DP if $\sup_{x\in\mathcal{X}}\left|\log\frac{P(x)}{Q(x)}\right| \le \varepsilon$. That is, in the case that $\sup_{x\in\mathcal{X}}\left|\log\frac{P(x)}{Q(x)}\right| \le \varepsilon$, we get ε-DP *for free* (since Hellinger distance, and thus sLLR, characterizes the optimal asymptotic sample complexity). Thus, we use the clamped log-likelihood ratio test, which forces the log-likelihood ratio to be bounded. Our lower bound in a sense shows that any loss of power in the test due to clamping is *necessary* for differentially private tests.

The lower bound proof proceeds by finding a coupling ρ of P^n and Q^n with low expected Hamming distance $\mathbb{E}_{(X,Y)\sim\rho}[d_H(X, Y)]$, which in turn implies that the sample complexity is large (see e.g. [3]). The coupling we use essentially splits the support of P and Q into two subsets, those elements with $\log\frac{P(x)}{Q(x)} \ge \varepsilon$ and the remaining elements. To construct the coupling, given a sample $X \sim P^n$, the high-ratio elements for which $\log\frac{P(x)}{Q(x)} \ge \varepsilon$ are resampled with probability related to the ratio. The contribution of this step to the Hamming distance gives us the $\varepsilon\tau$ part of the lower bound. The data set consisting of the $m \le n$ elements that have not yet been resampled is then coupled using the total variation distance coupling between P^m and Q^m. Therefore, with probability $1 - \mathrm{TV}(P^m, Q^m)$ this part of the data set remains unchanged. This part of the coupling results in the $H^2(P', Q')$ part of the lower bound.

The proof of the upper bound also splits into two parts, roughly corresponding to the same aspects of the distributions P and Q as above. That is, we view our tester as either counting the number of high-ratio elements or computing the log-likelihood ratio on low-ratio elements. A useful observation is that this duality between the upper and lower bounds is inevitable. In Section 3, we characterize the advantage of the optimal tester in terms of Wasserstein distance between P and Q with metric $\min\{\varepsilon d_H(X, Y), 1\}$. That is, the advantage of the optimal tester must be matched by some coupling of P^n and Q^n.

1.4 Related Work

Early work on differentially private hypothesis testing began in the Statistics community with [72, 74]. More recently, there has been a significant number of works on differentially private hypothesis testing. One line of work [17, 21, 40, 44, 49, 71, 76] designs differentially private versions of popular test statistics for testing goodness-of-fit, closeness, and independence, as well as private ANOVA, focusing on the performance at small sample sizes. Work by Wang et al. [75] focuses on generating statistical approximating distributions for differentially private statistics, which they apply to hypothesis testing problems. A recent work by Awan and Slavkovic [5] gives a universally optimal test when the domain size is two, however Brenner and Nissim [13] shows that such universally optimal tests cannot exist when the domain has more than two elements. A complementary research direction, initiated by Cai *et al.* [16], studies the minimax sample complexity of private hypothesis testing. [3] and [4] have given worst-case nearly optimal algorithms for goodness-of-fit and closeness testing of arbitrary discrete distributions. That is, there exists some worst-case distribution P such that their algorithm has optimal sample complexity for testing goodness-of-fit to P. Recently, [2] designed nearly optimal algorithms for estimating properties like support size and entropy.

Another related area [1, 41, 66] studies hypothesis testing in the *local model* of differential privacy. In particular, Duchi, Jordan, and Wainwright [26] proved an analogue of our result for the restricted

case of locally differentially private algorithms. Their characterization shows that, the optimal sample complexity for ε-DP local algorithms is $\Theta(1/(\varepsilon^2 TV(P,Q)^2))$. This characterization does not exhibit the same phenomena that we demonstrate in the central model—privacy never comes "for free" if $\varepsilon = o(1)$, and the sample complexity does not exhibit different regimes depending on ε. More generally, local-model tests are considerably simpler, and simpler to reason about, than central-model tests.

There are also several rich lines of work attempting to give tight instance-specific characterizations of the sample complexities of various differentially private computations, most notably *linear query release* [11, 43, 48, 57, 58] and *PAC and agnostic learning* [10, 39, 46]. The problems considered in these works are arguably more complex than the hypothesis testing problems we consider here, the characterizations are considerably looser, and are only optimal up to polynomial factors.

There has been a recent line of work [9, 23, 28–30, 37, 38, 64, 65, 78] on *adaptive data analysis*, in which the same dataset is used repeatedly across multiple statistics analyses, the choice of each analysis depends on the outcomes of previous analyses. The key theme in these works is to show that various strong notions of algorithmic stability, including differential privacy imply generalization bounds in the adaptive setting. Our characterization applies to all notions of stability considered in these works.

As an application of our private hypothesis testing results, we provide algorithms for private change-point detection. As discussed in Section 1.2.1, change-point detection has enjoyed a significant amount of study in information theory and statistics. Our results are in the private minimax setting, as recently introduced by Cummings et al. [22]. We improve on their results by improving the detection accuracy and providing strong privacy guarantees for all pairs of hypotheses.

2 UPPER BOUND ON SAMPLE COMPLEXITY OF PRIVATE TESTING

In this section, we establish the upper bound part of Theorem 1.2, establishing that the soft clamped log-likelihood ratio test scLLR and the noisy clamped log-likelihood ratio test ncLLR achieve the stated sample complexity. In more detail, we begin in Section 2.1 by characterizing the Hellinger distance between two distributions as the advantage, adv_1, of a specific randomized test, sLLR. Besides being of independent interest, this reformulation also yields some insight on its privatized variant, scLLR. In Section 2.2, we introduce the noisy clamped log-likelihood ratio test (ncLLR), and show that its sample complexity is at least that of scLLR. We then proceed in Section 2.3 to upper-bound the sample complexity of ncLLR, which also implies that same bound on that of scLLR.

2.1 Hellinger Distance Characterizes the Soft Log-Likelihood Test

Recall that the advantage of a test, adv_n, was defined in Equation (1) and the squared Hellinger distance, $H^2(P,Q)$, between two distributions P and Q is defined as $H^2(P,Q) = \frac{1}{2}\int_{\mathcal{X}}(\sqrt{P(x)} - \sqrt{Q(x)})^2\,dx$.

It has long been known that the Hellinger distance characterizes the asymptotic sample complexity of non-private testing (see, e.g., [12]). In this section we show that the Hellinger distance exactly characterizes the advantage of the following randomized test given a single data point:

$$\mathrm{sLLR}(x) = \begin{cases} P & \text{with probability } g(x) \\ Q & \text{with probability } 1 - g(x) \end{cases}$$

where

$$g(x) = \frac{\exp\left(\frac{1}{2}\log\frac{P(x)}{Q(x)}\right)}{1 + \exp\left(\frac{1}{2}\log\frac{P(x)}{Q(x)}\right)} \in [0,1].$$

Considering the advantage of sLLR might seem puzzling at first glance, since the classic likelihood ratio test LLR enjoys a better advantage. More specifically, the value of adv_1 for these two tests is $H^2(P,Q)$ (Theorem 2.1) and $TV(P,Q)$ (by the definition of total variation distance), respectively, and $H^2(P,Q) \le TV(P,Q)$ (see, e.g., [42]), so it would appear that LLR is the better test. There are two relevant features of sLLR which will be useful. First, as mentioned before, sLLR is naturally private if the likelihood ratio is bounded. Second, a tensorization property of Hellinger distance allows us to easily relate the advantage of the n-sample test to the advantage of the 1-sample test.

Theorem 2.1. *For any two distributions P, Q, the advantage,* adv_1, *of* sLLR *is $H^2(P,Q)$.*

Proof. Note that we can rewrite

$$\begin{aligned} H^2(P,Q) &= \frac{1}{2}\int_{\mathcal{X}}(\sqrt{P(x)} - \sqrt{Q(x)})^2\,dx \\ &= \frac{1}{2}\int_{\mathcal{X}}(P(x) - Q(x))\frac{\sqrt{P(x)} - \sqrt{Q(x)}}{\sqrt{P(x)} + \sqrt{Q(x)}} \\ &= \frac{1}{2}\int_{\mathcal{X}}(P(x) - Q(x))\frac{\sqrt{P(x)/Q(x)} - 1}{\sqrt{P(x)/Q(x)} + 1}\,dx. \end{aligned}$$

Now, $g(x) = \sqrt{\frac{P(x)}{Q(x)}}/\left(\sqrt{\frac{P(x)}{Q(x)}}+1\right) = \frac{1}{2}\left(\sqrt{\frac{P(x)}{Q(x)}}-1\right)/\left(\sqrt{\frac{P(x)}{Q(x)}}+1+1\right)$, and therefore $H^2(P,Q) = \int_{\mathcal{X}}(P(x) - Q(x))g(x)\,dx = \mathbb{E}_{x\sim P}[g(x)] - \mathbb{E}_{x\sim Q}[g(x)] = \mathbb{P}_{x\sim P}[\mathrm{sLLR}(x) = P] - \mathbb{P}_{x\sim Q}[\mathrm{sLLR}(x) = P]$. Thus, the advantage of sLLR is $H^2(P,Q)$, as claimed. □

This tells us the advantage of the test which takes only one sample. As a corollary, we can derive the sample complexity of distinguishing P and Q using sLLR.

Corollary 2.2. $SC^{P,Q}(\mathrm{sLLR}) = O\left(\frac{1}{H^2(P,Q)}\right)$.

Proof. Our analysis is similar to that of [18]. Observe that the test sLLR which gets n samples from either P or Q is equivalent to the test sLLR which gets 1 sample from either P^n or Q^n. By Theorem 2.1, we have that the advantage of either (equivalent) test is $\mathrm{adv} = H^2(P^n, Q^n)$. We will require the following tensorization property of the squared Hellinger distance: $H^2(P_1 \times \cdots \times P_n, Q_1 \times \cdots \times Q_n) = 1 - \prod_{i=1}^n (1 - H^2(P_i, Q_i))$. With this in hand, $\mathrm{adv} = H^2(P^n, Q^n) = 1 - (1 - H^2(P,Q))^n = 1 - \exp\left(n\log(1 - H^2(P,Q))\right) \ge 1 - \exp(-nH^2(P,Q))$. Setting $n = \Omega(1/H^2(P,Q))$, we get $\mathrm{adv} \ge 2/3$, as desired. □

2.2 The Noisy Log-Likelihood Ratio Test

We now consider the noisy log-likelihood ratio test, which, similar to $\text{scLLR}_{-\varepsilon',\varepsilon}$, is also ε-differentially private.

$$\text{ncLLR}_{-\varepsilon',\varepsilon}(X) = \sum_i \left[\log \frac{P(x_i)}{Q(x_i)}\right]_{-\varepsilon'}^{\varepsilon} + \text{Lap}(2)$$

Here Lap(2) denotes a *Laplace random variable*, which has density proportional to $\exp(-|z|/2)$. Readers familiar with differential privacy will note that scLLR corresponds to the exponential mechanism, while ncLLR responds to the *report noisy max* mechanism [33], and thus the two should behave quite similarly. In particular, we have the following lemma.

Lemma 2.3. *For any P and Q:*

(1) $SC^{P,Q}(\text{ncLLR}_{-\varepsilon',\varepsilon}) = \Omega(SC^{P,Q}(\text{scLLR}_{-\varepsilon',\varepsilon}))$.
(2) *Furthermore, if $-\varepsilon' \le \log \frac{P}{Q} \le \varepsilon$ then $SC^{P,Q}(\text{ncLLR}_{-\varepsilon',\varepsilon}) = \Theta(SC^{P,Q}(\text{scLLR}_{-\varepsilon',\varepsilon}))$.*

Proof. Recall that

$$\text{scLLR}_{-\varepsilon',\varepsilon}(X) = \begin{cases} P & \text{with probability } g_\varepsilon(X) \\ Q & \text{with probability } 1 - g_\varepsilon(X) \end{cases}$$

where $g_\varepsilon(X) = \frac{e^{\frac{1}{2}\,\text{cLLR}_{-\varepsilon',\varepsilon}(X)}}{1+e^{\frac{1}{2}\,\text{cLLR}_{-\varepsilon',\varepsilon}(X)}}$. If we let the threshold $\kappa = 0$ then the test based on the test statistic $\text{ncLLR}_{-\varepsilon',\varepsilon}$ is

$$\text{ncLLR}_{-\varepsilon',\varepsilon}(X) = \begin{cases} P & \text{with probability } h_\varepsilon(X) \\ Q & \text{with probability } 1 - h_\varepsilon(X) \end{cases}$$

where

$$h_\varepsilon(X) = \begin{cases} 1 - \frac{1}{2}e^{-\frac{\text{cLLR}_{-\varepsilon',\varepsilon}(X)}{2}} & \text{if } \text{cLLR}_{-\varepsilon',\varepsilon}(X) > 0 \\ \frac{1}{2}e^{\frac{\text{cLLR}_{-\varepsilon',\varepsilon}(X)}{2}} & \text{if } \text{cLLR}_{-\varepsilon',\varepsilon}(X) < 0 \end{cases}.$$

Now, we will use the following two inequalities: if $x < 0$, then $\frac{1}{2}e^{\frac{x}{2}} \le \frac{e^{\frac{1}{2}x}}{1+e^{\frac{1}{2}x}} \le 2\frac{1}{2}e^{\frac{x}{2}}$, and if $x > 0$, $\frac{8}{9}(1 - \frac{1}{2}e^{-\frac{x}{2}}) \le \frac{e^{\frac{1}{2}x}}{1+e^{\frac{1}{2}x}} \le 1 - \frac{1}{2}e^{-\frac{x}{2}}$. Therefore, noting that $\mathbb{P}_{X\sim P^n}\left[\text{scLLR}_{-\varepsilon',\varepsilon}(X) = P\right] = \mathbb{E}_{P^n}[g_\varepsilon(X)]$ and $\mathbb{P}_{X\sim P^n}\left[\text{ncLLR}_{-\varepsilon',\varepsilon}(X) = P\right] = \mathbb{E}_{P^n}[h_\varepsilon(X)]$ are the probabilities of success, we have $\frac{8}{9}\mathbb{E}_{P^n}[h_\varepsilon(X)] \le \mathbb{E}_{P^n}[g_\varepsilon(X)] \le 2\mathbb{E}_{P^n}[h_\varepsilon(X)]$ and $\frac{8}{9}\mathbb{E}_{Q^n}[h_\varepsilon(X)] \le \mathbb{E}_{Q^n}[g_\varepsilon(X)] \le 2\mathbb{E}_{Q^n}[h_\varepsilon(X)]$. Therefore, if ncLLR has a probability of success of 5/6 then scLLR has a probability of success of 2/3. This implies that $SC^{P,Q}(\text{ncLLR}) \ge SC^{P,Q}(\text{scLLR})$.

If $\log \frac{P}{Q} \in [-\varepsilon', \varepsilon]$ then $\text{cLLR}_{-\varepsilon',\varepsilon}(X) = \text{LLR}(X)$ so we have $P^n(X) > Q^n(X)$ iff $\text{LLR}(X) > 0$ iff $g_\varepsilon(X) \le h_\varepsilon(X)$ and therefore

$$\mathbb{E}_{P^n}[g_\varepsilon] - \mathbb{E}_{Q^n}[g_\varepsilon] = \int (P^n(X) - Q^n(X))g_\varepsilon(X)\,dX$$
$$\le \int (P^n(X) - Q^n(X))h_\varepsilon(X)\,dX = \mathbb{E}_{P^n}[h_\varepsilon] - \mathbb{E}_{Q^n}[h_\varepsilon],$$

which completes the proof. □

Corollary 2.4. *If* ncLLR *has asymptotically optimal sample complexity then* scLLR *has asymptotically optimal sample complexity.*

2.3 The Sample Complexity of ncLLR

In this section we prove the upper bound in Theorem 1.2 for the case where $\text{TV}(P,Q) < 1$ (i.e., the supports of P and Q have non-empty intersection). This assumption ensures that $\tilde{P}, \tilde{Q} \neq 0$, so that P', Q' are well defined. A proof of the case where $\text{TV}(P,Q) = 1$ is in the full version. In order to prove the upper bound in Theorem 1.2, we restate it as follows.

Theorem 2.5. *The $\text{ncLLR}_{-\varepsilon',\varepsilon}$ test is ε-DP and $SC^{P,Q}(\text{ncLLR}_{-\varepsilon',\varepsilon}) \le O\left(\frac{1}{\varepsilon\tau + (1-\tau)H^2(P',Q')}\right) = O\left(\min\left\{\frac{1}{\varepsilon\tau}, \frac{1}{(1-\tau)H^2(P',Q')}\right\}\right)$.*

Theorem 2.5 combined with a matching lower bound (given later in Theorem 3.4) imply that $\text{ncLLR}^{\varepsilon}_{-\varepsilon'}$ has asymptotically optimal sample complexity. Thus, by Corollary 2.4, $\text{scLLR}^{\varepsilon}_{-\varepsilon'}$ has asymptotically optimal sample complexity.

Before proving the bound, we pause to provide some intuition for its form. As discussed in the introduction, we can write P and Q as mixtures $P = (1-\tau)P' + \tau P''$ and $Q = (1-\tau)Q' + \tau Q''$ where P'', Q'' have disjoint support. Now consider a thought experiment, in which the test that must distinguish P from Q using a sample of size n is given, along with the sample x, a list of binary labels $b_1, b_2, \ldots, b_n$ that indicate for each record whether it was sampled from the first component of the mixture (either P' or Q'), or the second component (either P'' or Q''). Of course this can only be a though experiment—these labels are not available to a real test.

Because the mixture weights are the same for both P and Q, the number of labels of each type would be distributed the same under P and under Q, and so the tester would be faced with two independent testing problems: distinguishing P'' from Q'' using a sample of size about τn, and distinguishing P' from Q' using a sample of size about $(1-\tau)n$. It would suffice for the tester to solve either of these problems.

Theorem 2.5 shows that the real tests (scLLR and ncLLR) do as well as the hypothetical tester that has access to the labels. The two arguments to the minimum in the theorem statement correspond directly to the ε-DP sample complexity of distinguishing P'' from Q'' (which requires $n\tau \ge 1/\varepsilon$) or distinguishing P' from Q' (which requires $n(1-\tau) \ge 1/H^2(P',Q')$). The proof proceeds by breaking the clamped log-likelihood ratio into two pieces, each corresponding to one of the two mixture components (again, this decomposition is not known to the algorithm). These two pieces correspond to the test statistics of the optimal testers for the two separate sub-problems in the thought experiment. We show that the test does well at distinguishing P from Q as long as either of these pieces is sufficiently informative.

On a more mechanical level, our proof of Theorem 2.5 bounds the expectation and standard deviation of the two pieces of the test statistic. We use the following simple lemma, which states that a test statistic S performs well if the distribution of the test statistic on P and Q, $S(P)$ and $S(Q)$, must not overlap too much. The proof is a simple application of Chebyshev's inequality.

Lemma 2.6. *Given a function $f : \mathcal{X} \to \mathbb{R}$, constant $c > 0$ and $n > 0$, if the test statistic $S(X) = \sum_i f(x_i)$ satisfies $\max\left\{\sqrt{\text{Var}_{P^n}[S(X)]}, \sqrt{\text{Var}_{Q^n}[S(X)]}\right\} \le c|\mathbb{E}_{P^n}[S(X)] - \mathbb{E}_{Q^n}[S(X)]|$ then S can be used to distinguish between P and Q with probability of success 2/3 and sample complexity at most $n' = 12c^2n$.*

The definitions of $\tilde{P}$ and $\tilde{Q}$ lend naturally to consider a partition of the space $\mathcal{X}$, depending on which quantities achieve the minimum in $\min(e^{\varepsilon}Q, P)$ and $\min(e^{\varepsilon'}P, Q)$. This partition will itself play a crucial role in both the proof of the theorem, and later in our lower bound: accordingly, define

$$\begin{aligned} \mathcal{S} &= \left\{ x : P(x) - e^{\varepsilon}Q(x) > 0 \right\} \\ \mathcal{T} &= \left\{ x : Q(x) - e^{\varepsilon'}P(x) > 0 \right\} \end{aligned} \tag{7}$$

and set $\mathcal{A} = \mathcal{X} \setminus (\mathcal{S} \cup \mathcal{T})$.

Proof of Theorem 2.5. Observe that for all $x \in \mathcal{A}$, $\tilde{P}(x) = P(x)$ and $\tilde{Q}(x) = Q(x)$ (so that $P'(x)/Q'(x) = P(x)/Q(x)$), that for all $x \in \mathcal{S}$, $\log(P'(x)/Q'(x)) = \varepsilon$ and that for all $x \in \mathcal{T}$, $\log(P'(x)/Q'(x)) = -\varepsilon'$. To show that the test works, we first show that the clamped log likelihood ratio (without noise) is a useful test statistic. In order to apply Lemma 2.6, we first calculate the difference Δ_{gap} in the expectations of $\text{cLLR}_{-\varepsilon',\varepsilon}$ under P and Q. For the remainder of the proof, we omit the $-\varepsilon', \varepsilon$ subscript (since clamping always occurs to the same interval).

$$\begin{aligned} \frac{1}{n}\Delta_{\text{gap}} &= \frac{1}{n}\mathop{\mathbb{E}}_{P^n}[\text{cLLR}(X)] - \mathop{\mathbb{E}}_{Q^n}[\text{cLLR}(X)] \\ &= \int_{\mathcal{X}} (P(x) - Q(x)) \log \frac{P'(x)}{Q'(x)}\, dx \\ &= (P(\mathcal{S}) - Q(\mathcal{S}))\varepsilon + \int_{\mathcal{A}} (\tilde{P}(x) - \tilde{Q}(x)) \log \frac{P'(x)}{Q'(x)}\, dx \\ &\quad + (Q(\mathcal{T}) - P(\mathcal{T}))\varepsilon' \\ &= (\tilde{P}(\mathcal{S}) - \tilde{Q}(\mathcal{S}) + \tau)\varepsilon + \int_{\mathcal{A}} (\tilde{P}(x) - \tilde{Q}(x)) \log \frac{P'(x)}{Q'(x)}\, dx \\ &\quad + (\tilde{Q}(\mathcal{T}) - \tilde{P}(\mathcal{T}) + \tau)\varepsilon' \end{aligned}$$

where the last equality follows by the definition of τ. Moreover,

$$\begin{aligned} &nKL\left(P' \| Q'\right) + nKL\left(Q' \| P'\right) \\ &= n\int_{\mathcal{X}} (P'(x) - Q'(x)) \log \frac{P'(x)}{Q'(x)}\, dx \\ &= n(P'(\mathcal{S}) - Q'(\mathcal{S}))\varepsilon + n\int_{\mathcal{A}} (P'(x) - Q'(x)) \log \frac{P'(x)}{Q'(x)}\, dx \\ &\quad + n(Q'(\mathcal{T}) - P'(\mathcal{T}))\varepsilon' \\ &= \frac{n}{1-\tau}\Big((\tilde{P}(\mathcal{S}) - \tilde{Q}(\mathcal{S}))\varepsilon + \int_{\mathcal{A}} (\tilde{P}(x) - \tilde{Q}(x)) \log \frac{P'(x)}{Q'(x)} \\ &\quad + (\tilde{Q}(\mathcal{T}) - \tilde{P}(\mathcal{T}))\varepsilon'\Big). \end{aligned}$$

Therefore, $\Delta_{\text{gap}} = (1-\tau)n(KL\left(P' \| Q'\right) + KL\left(Q' \| P'\right)) + n\tau\varepsilon + n\tau\varepsilon' \geq 2n\left((1-\tau)H^2(P', Q') + \tau\varepsilon\right)$, where the last inequality follows since $H^2(P, Q) \leq KL\left(P \| Q\right)$ for any distributions P and Q.

We now turn to bounding the variance of the noisy clamped LLR under P and Q. Noting that $\text{Var}_{P^n}[\text{ncLLR}] = \text{Var}_{P^n}[\text{cLLR}] + 8$, by Lemma 2.6 it suffices to show that $\max\{\text{Var}_{Q^n}[\text{cLLR}]+8, \text{Var}_{P^n}[\text{cLLR}]+8\} \leq O(\Delta_{\text{gap}}^2)$; or, equivalently, that $\max\{\text{Var}_{Q^n}[\text{cLLR}], \text{Var}_{P^n}[\text{cLLR}]\} \leq O(\Delta_{\text{gap}}^2)$ and $\Delta_{\text{gap}} = \Omega(1)$. Recall that P'' is a distribution such that $P = \tau P'' + (1-\tau)P'$ and the support of P'' is contained in $\mathcal{S}$. Thus,

$$\begin{aligned} \mathop{\text{Var}}_{P^n}[\text{cLLR}] &\leq n\int_{\mathcal{X}} P(x) \log^2 \frac{P'(x)}{Q'(x)}\, dx \\ &= n\int_{\mathcal{X}} (\tau P''(x) + (1-\tau)P'(x)) \log^2 \frac{P'(x)}{Q'(x)}\, dx \\ &= n\Big(\tau P''(\mathcal{S})\varepsilon^2 + (1-\tau)\int_{\mathcal{X}} P'(x) \log^2 \frac{P'(x)}{Q'(x)}\, dx\Big) \end{aligned}$$

Since $\log \frac{P'(x)}{Q'(x)} \leq \varepsilon \leq 1$, $|\log \frac{P'(x)}{Q'(x)}| \leq 3 \cdot |1 - \sqrt{Q'(x)/P'(x)}|$. Therefore, $\int_{\mathcal{X}} P'(x) \log^2 \frac{P'(x)}{Q'(x)}\, dx \leq 9\int_{\mathcal{X}} P'(x)(1 - \sqrt{Q'(x)/P'(x)})^2\, dx = 18 \cdot H^2(P', Q')$. Also, $P''(\mathcal{S}) = 1$ and thus we have that $\text{Var}_{P^n}[\text{cLLR}] = O(n\tau\varepsilon^2 + (1-\tau)nH^2(P', Q')) = O(n\tau\varepsilon + (1-\tau)nH^2(P', Q')$. Similarly, $\text{Var}_{Q^n}[\text{cLLR}] \leq O(n\tau\varepsilon'^2 + (1-\tau)H^2(P', Q')) \leq O(n\tau\varepsilon + (1-\tau)nH^2(P', Q'))$. Finally, $\max\{\text{Var}_{Q^n}[\text{cLLR}], \text{Var}_{P^n}[\text{cLLR}]\} = O\left(\frac{(n\tau\varepsilon + n(1-\tau)H^2(P', Q'))^2}{n\tau\varepsilon + n(1-\tau)H^2(P', Q')}\right) = O\big(\Delta_{\text{gap}}^2 / n(\tau\varepsilon + (1-\tau)H^2(P', Q'))\big)$. Therefore, if $n \geq \frac{C}{\tau\varepsilon + (1-\tau)H^2(P', Q'))}$ for some suitably large constant $C > 0$, this implies that $\max\{\text{Var}_{Q^n}[\text{cLLR}], \text{Var}_{P^n}[\text{cLLR}]\} \leq O(\Delta_{\text{gap}}^2)$ and $\Delta_{\text{gap}} = \Omega(1)$, which concludes our proof. □

3 LOWER BOUND ON THE SAMPLE COMPLEXITY OF PRIVATE TESTING

We now prove the lower bound in Theorem 1.2. We do so by constructing an appropriate *coupling* between the distributions P^n and Q^n, which implies lower bounds for privately distinguishing P^n from Q^n. This style of analysis was introduced in [3], though we require a strengthening of their statement. Specifically, the lower bound of Acharya et al. involves $d'_{\varepsilon}(X, Y) = \varepsilon d_H(X, Y)$, whereas we have $d_{\varepsilon}(X, Y) = \min(\varepsilon d_H(X, Y), 1)$.

For $X, Y \in \mathcal{X}^n$, let $d_H(X, Y)$ be the Hamming distance between X, Y (i.e., $|\{ i : x_i \neq y_i \}|$). Given a metric $d : \mathcal{X}^n \times \mathcal{X}^n \to \mathbb{R}_{\geq 0}$, we define the Wasserstein distance $W_d(P, Q)$ to be $W_d(P, Q) = \inf_{\rho} \mathbb{E}_{(X,Y)\sim\rho}[d(X, Y)]$, where the inf is over all couplings ρ of P^n and Q^n. Let $d_{\varepsilon}(X, Y) = \min\{\varepsilon d_H(X, Y), 1\}$.

Lemma 3.1. *For every ε-DP algorithm $M : \mathcal{X}^n \to \{0, 1\}$, if X and Y are neighboring datasets then $\mathbb{E}[M(X)] \leq e^{\varepsilon}\mathbb{E}[M(Y)]$, where the expectations are over the randomness of the algorithm M.*

Lemma 3.2. *For every ε-DP algorithm $M : \mathcal{X}^n \to \{P, Q\}$, we have the following bound on the advantage: $\left|\mathbb{P}_{X\sim P}[M(X) = P] - \mathbb{P}_{X\sim Q}[M(X) = P]\right| \leq O(W_{d_{\varepsilon}}(P, Q))$.*

Proof. Let $\rho : \mathcal{X}^n \times \mathcal{X}^n \to \mathbb{R}_{\geq 0}$ be a coupling of P^n and Q^n and M be an ε-DP algorithm. We have

$$\begin{aligned} &\mathop{\mathbb{P}}_{X\sim P}[M(X) = P] - \mathop{\mathbb{P}}_{X\sim Q}[M(X) = P] \\ &= \int_{\mathcal{X}^n}\int_{\mathcal{X}^n} \left(\rho(X, Y)\mathop{\mathbb{P}}_{M}[M(X) = P] - \rho(X, Y)\mathop{\mathbb{P}}_{M}[M(Y) = P]\right) dX\, dY \\ &\leq \int_{\mathcal{X}^n}\int_{\mathcal{X}^n} \rho(X, Y) \min\{1, (e^{\varepsilon d_H(X,Y)} - 1)\mathop{\mathbb{P}}_{M}[M(Y) = P]\}\, dX\, dY \\ &\leq 2\int_{\mathcal{X}^n}\int_{\mathcal{X}^n} \rho(X, Y) \min\{1, \varepsilon d_H(X, Y)\}\, dX\, dY \\ &= O(\mathop{\mathbb{E}}_{(X,Y)\sim\rho}[d_{\varepsilon}(X, Y)]), \end{aligned}$$

where $\mathbb{P}_M[\cdot]$ denotes that the probability is over the randomness of the algorithm M, and the first inequality follows from Lemma 3.1. □

The upper bound in Lemma 3.2 is in fact tight. We state the converse (whose proof appears in the full version) below for completeness, although we will not use it in this work.

Lemma 3.3. *There is a ε-DP algorithm $M\colon \mathcal{X}^n \to \{P, Q\}$ such that $\left|\mathbb{P}_{X\sim P}[M(X) = P] - \mathbb{P}_{X\sim Q}[M(X) = P]\right| \geq \Omega(W_{d_\varepsilon}(P, Q))$.*

We will also rely on the following standard fact characterizing total variation distance in terms of couplings:

Fact 1. *Given P and Q, there exists a coupling ρ of P and Q such that $\mathbb{P}_\rho[X \neq Y] = \mathrm{TV}(P, Q)$. We will refer to this coupling as the* total variation coupling *of P and Q.*

We can now prove the lower bound component of Theorem 1.2. Recall that P' and Q' were defined in Equation (5).

Theorem 3.4. *Given P and Q, every ε-DP test K that distinguishes P and Q has the property that $SC^{P,Q}(K) = \Omega\left(\frac{1}{\varepsilon\tau + H^2(P', Q')}\right)$.*

Proof. Consider the following coupling of P^n and Q^n: Given a sample $X \sim P^n$, independently for all $x_i \in \mathcal{S}$[5] label the point $\mathbb{1}$ with probability $\frac{e^\varepsilon Q(x_i)}{P(x_i)}$, otherwise label it . Label all the points in $\mathcal{A}\cup\mathcal{T}$ as $\mathbb{1}$. (In particular, this implies that each x_i is labeled $\mathbb{1}$ with probability $1-\tau$, and with probability τ.) Each point labeled is then independently re-sampled from $\mathcal{T}$ with probability distribution $\frac{Q - e^{\varepsilon'}P}{\tau}\mathbb{1}_{\mathcal{T}}$. Let $\Lambda \subseteq [n]$ be the set of points labeled $\mathbb{1}$, and n' be its size; and note that this set is distributed according to $(P')^{n'}$. We transform this set to a set distributed by $(Q')^{n'}$ using the TV-coupling of $(P')^{n'}$ and $(Q')^{n'}$. The result is a sample from Q^n.

Now, we can rewrite $d_H(X, Y) = \sum_{i\notin\Lambda}\mathbb{1}_{\{X_i=Y_i\}} + \sum_{i\in\Lambda}\mathbb{1}_{\{X_i=Y_i\}} = d_H(X_{\bar{\Lambda}}, Y_{\bar{\Lambda}}) + \mathbb{1}_{\{X_\Lambda=Y_\Lambda\}}\cdot\sum_{i\in\Lambda}\mathbb{1}_{\{X_i=Y_i\}}$, and therefore

$$\begin{aligned}
&\mathbb{E}[\min\{\varepsilon d_H(X, Y), 1\}]\\
&= \mathbb{E}\left[\min\{\varepsilon d_H(X, Y), 1\}\mathbb{1}_{\{X_\Lambda=Y_\Lambda\}}\right] + \mathbb{E}\left[\min\{\varepsilon d_H(X, Y), 1\}\mathbb{1}_{\{X_\Lambda\neq Y_\Lambda\}}\right]\\
&\leq \varepsilon\mathbb{E}\left[d_H(X, Y)\mathbb{1}_{\{X_\Lambda=Y_\Lambda\}}\right] + \mathbb{E}\left[\mathbb{1}_{\{X_\Lambda\neq Y_\Lambda\}}\right]\\
&= \varepsilon\mathbb{E}\left[d_H(X_{\bar{\Lambda}}, Y_{\bar{\Lambda}})\right] + \mathbb{P}[X_\Lambda \neq Y_\Lambda].
\end{aligned}$$

Recalling now that the distribution of (X_Λ, Y_Λ) is that of the TV-coupling of $(P')^{n'}$ and $(Q')^{n'}$, and that $|\bar{\Lambda}| = n - n'$, we get

$$\begin{aligned}
\mathbb{E}[\min\{\varepsilon d_H(X, Y), 1\}] &\leq \mathbb{E}\left[\varepsilon(n - n') + \mathrm{TV}((P')^{n'}, (Q')^{n'})\right]\\
&\leq \varepsilon\tau n + \mathbb{E}\left[\sqrt{n'}\right]H(P', Q') \leq \varepsilon\tau n + \sqrt{(1-\tau)n}\cdot H(P', Q')
\end{aligned}$$

Therefore, by Lemma 3.2, we have that for every ε-DP test M,

$$\left|\mathop{\mathbb{P}}_{X\sim P}[M(X) = P] - \mathop{\mathbb{P}}_{X\sim Q}[M(X) = P]\right| \leq \varepsilon\tau n + \sqrt{(1-\tau)n}\cdot H(P', Q').$$

Thus, in order for the probability of success to be $\Omega(1)$, we need either $\varepsilon\tau n$ or $\sqrt{(1-\tau)n}H(P', Q')$ to be $\Omega(1)$. That is, $n \geq \Omega\left(\min\left\{\frac{1}{\varepsilon\tau}, \frac{1}{(1-\tau)H^2(P', Q')}\right\}\right)$. □

[5] Recall the definitions of $\mathcal{S}$, $\mathcal{T}$ and $\mathcal{A}$ from Section 2.3 (Equation (7)).

4 APPLICATION: DIFFERENTIALLY PRIVATE CHANGE-POINT DETECTION

In this section, we give an application of our method to differentially private change-point detection (CPD). In the change-point detection problem, we are given a time-series of data. Initially, it comes from a known distribution P, and at some unknown time step, it begins to come from another known distribution Q. The goal is to approximate when this change occurs. More formally, we have the following definition.

Definition 4.1. In the *offline change-point detection problem*, we are given distributions P, Q and a data set $X = \{x_1, \ldots, x_n\}$. We are guaranteed that there exists $k^* \in [n]$ such that $x_1, \ldots, x_{k^*-1} \sim P$ and $x_{k^*}, \ldots, x_n \sim Q$. The goal is to output $\hat{k}$ such that $|\hat{k} - k^*|$ is small.

In the *online change-point detection problem*, we are given distributions P, Q, a stream of data points $X = \{x_1, \ldots\}$. We are guaranteed that there exists k^* such that $x_1, \ldots, x_{k^*-1} \sim P$ and $x_{k^*}, \cdots \sim Q$. The goal is to output $\hat{k}$ such that $|\hat{k} - k^*|$ is small.

We study the parameterization of the private change-point detection problem recently introduced by Cummings *et al.* [22].

Definition 4.2 (Change-Point Detection). An algorithm for a (online) change-point detection problem is (α, β)-accurate if for any input dataset (data stream), with probability at least $1 - \beta$ outputs a $\hat{k}$ such that $|\hat{k} - k^*| \leq \alpha$, where the probability is with respect to the randomness in the sampling of the data set and the random choices made by the algorithm.

Our main result is the following:

Theorem 4.3. *There exists an efficient ε-differentially private and (α, β)-accurate algorithm for offline change-point detection from distribution P to Q with $\alpha = \Theta\left(\frac{1}{\varepsilon\tau(P,Q)+H^2(P',Q')}\cdot\log(1/\beta)\right)$.*

Furthermore, there exists an efficient ε-differentially private and (α, β)-accurate algorithm for online change-point detection from distribution P to Q with the same value of α. This latter algorithm also requires as input a value n such that $n = \Omega\left(SC_\varepsilon^{P,Q}\cdot\log\left(\frac{k^*}{n\beta}\right)\right)$. If the algorithm is accurate, it will observe at most $k^* + 2n$ data points, and with high probability observe $k^* + O(n\log n)$ data points.

For constant β, the accuracy of our offline algorithm is optimal up to constant factors, since one can easily show that the best accuracy achievable is $\Omega(SC_\varepsilon^{P,Q})$. A similar statement holds for our online algorithm when the algorithm is given an estimate n of k^* such that $n = \mathrm{poly}(k^*)$.

As one might guess, this problem is intimately related to the hypothesis testing question studied in the rest of this paper. Indeed, our change-point detection algorithm will use the hypothesis testing algorithm of Theorem 1.2 as a black box, in order to reduce to a simpler Bernoulli change-point detection problem (see Lemma 4.4 in Section 4.1). We then give an algorithm to solve this simpler problem (Lemma 4.5 in Section 4.2), completing the proof of Theorem 4.3. In Section 4.3, we show that our reduction is applicable more generally, as we describe an algorithm change-point detection in a goodness-of-fit setting.

4.1 A Reduction to Bernoulli CPD

In this section, we provide a reduction from private change-point detection with arbitrary distributions to non-private change-point detection with Bernoulli distributions.

Lemma 4.4. *Suppose there exists an (α, β)-accurate algorithm which can solve the following restricted change-point detection problem: we are guaranteed that there exists $\tilde{k}^*$ such that $z_1, \dots, z_{\tilde{k}^*-1} \sim 2\,\mathrm{Ber}(\tau_0) - 1$ for some $\tau_0 > 2/3$, $z_{\tilde{k}^*+1}, \dots \sim 2\,\mathrm{Ber}(\tau_1) - 1$ for some $\tau_1 < 1/3$, and $z_{\tilde{k}^*} \in \{\pm 1\}$ is arbitrary.*

Then there exists an ε-differentially private and $((\alpha+1)\cdot SC_\varepsilon^{P,Q}, \beta)$-accurate algorithm which solves the change-point detection problem, where $SC_\varepsilon^{P,Q}$ is as defined in Theorem 1.2.

Proof. We describe the reduction for the offline version of the problem, the reduction in the online setting is identical. The reduction is easy to describe.

We partition the samples into intervals of length $SC_\varepsilon^{P,Q}$. Let $Y_j = \{x_{(i-1)SC_\varepsilon^{P,Q}+1}, \dots, x_{iSC_\varepsilon^{P,Q}}\}$, for $j = 1$ to $\lfloor n/SC_\varepsilon^{P,Q} \rfloor$, and disregard the remaining "tail" of x_i's. We run the algorithm of Theorem 1.2 on each Y_j, and produce a bit $z_j = 1$ if the algorithm outputs that the distribution is P, and a $z_j = -1$ otherwise.

We show that this forms a valid instance of the change-point detection problem in the lemma statement. Let k^* be the change-point in the original problem, and suppose it belongs to $Y_{\tilde{k}^*}$. For every $j < \tilde{k}^*$, all the samples are from P, and by Theorem 1.2, each z_j will independently be 1 with probability $\tau_0 \geq 2/3$. Similarly for every $j > \tilde{k}^*$, all the samples are from Q, and each z_j will independently be -1 with probability at least $1 - \tau_1 \geq 2/3$.

Finally, we show that the existence of an (α, β)-accurate algorithm that solves this problem also solves the original problem. Suppose that the output of the algorithm on the restricted change-point detection problem is j. To map this to an answer to the original problem, we let $\hat{k} = (j-1)SC_\varepsilon^{P,Q}$.

First, note that $\hat{k}$ will be ε-differentially private. We claim that the sequence of z_j's is ε-differentially private. This is because the algorithm of Theorem 1.2 is ε-differentially private, we apply the algorithm independently to each component of the partition, and each data point affects only one component (since they are disjoint). Privacy of $\hat{k}$ follows since privacy is closed under post-processing.

Finally, we establish accuracy. In the restricted change-point detection problem, with probability at least $1 - \beta$, the output j will be such that $|j - \tilde{k}^*| \leq \alpha$. In the original problem's domain, this corresponds to a $\hat{k}$ such that $|\hat{k} - k^*| \leq (\alpha+1)SC_\varepsilon^{P,Q}$, as desired. □

4.2 Solving Bernoulli CPD

In this section, we show that there is a $(\Theta(\log(1/\beta)) + 1, \beta)$-accurate algorithm for the restricted change-point detection problem. Combined with Lemma 4.4, this implies Theorem 4.3.

Lemma 4.5. *There exists an efficient $(O(\log(1/\beta), \beta)$-accurate algorithm for the offline restricted change-point detection problem (as defined in Lemma 4.4).*

Similarly, there exists an efficient $(O(\log(1/\beta), \beta)$-accurate algorithm for the online restricted change-point detection problem. This algorithm requires as input a value n such that $n = \Omega\left(\log\left(\frac{k^*}{n\beta}\right)\right)$. If the algorithm is accurate, it will observe at most $k^* + 2n$ data points, and with high probability observe $k^* + O(n \log n)$ data points.

Proof. We start by describing the algorithm for the offline version of the problem. We then discuss how to reduce from the online problem to the offline problem.

Offline Change-Point Detection. We define the function $\ell(t) = \sum_{j=t}^{n} z_j$. The algorithm's output will be $\hat{k} = \arg\min_{1 \leq t \leq n} \ell(t)$.

Let k^* be the true change-point index. To prove correctness of this algorithm, we show that $\ell(k^*+1) - \ell(t) < 0$ for all $t \geq k^* + 1 + \Theta(\log(1/\beta))$, and that $\ell(k^*-1) - \ell(t) < 0$ for all $t \leq k^* - 1 - \Theta(\log(1/\beta))$. Together, these will show that $\arg\min_{1 \leq t \leq n} \ell(t) \in [k^* - 1 - \Theta(\log(1/\beta)), k^* + 1 + \Theta(\log(1/\beta))]$, proving the result. For the remainder of this proof, we focus on the former case, the latter will follow symmetrically. Specifically, we will show that $\ell(k^*+1) - \ell(t) < 0$ for all $t \geq k^* + 1 + c\log(1/\beta)$, where $c > 0$ is some large absolute constant.

Observe that for $t \geq k^* + 1$, $\ell(k^*+1) - \ell(t) = \sum_{j=k^*+1}^{t-1} z_j$ forms a biased random walk which takes a +1-step with probability $\tau_1 \leq 1/3$ and a -1-step with probability $1 - \tau_1 \geq 2/3$. Define $M_i = \ell(k^*+1) - \ell(k^*+1+i) + i(1-2\tau_1)$ for $i = 0$ to $n - k^* - 1$, and note that this forms a martingale sequence. We will use Theorem 4 of [7], which provides a finite-time law of the iterated logarithm result. Specialized to our setting, we obtain the following maximal inequality, bounding how far this martingale deviates away from 0.

Theorem 4.6 (Follows from Theorem 4 of [7]). *Let $c > 0$ be some absolute constant. With probability at least $1 - \beta$, for all $i \geq c\log(1/\beta)$ simultaneously, $|M_i| \leq O\left(\sqrt{i\,(\log\log(i) + \log(1/\beta))}\right)$.*

This implies that, with probability at least $1 - \beta$, we have that for all $i \geq c\log(1/\beta)$, $\ell(k^*+1) - \ell(k^*+1+i) = M_i - i(1-2\tau_1) \leq O\left(\sqrt{i\,(\log\log(i) + \log(1/\beta))}\right) - \frac{i}{3}$. Note that the right-hand side is non-increasing in i, so it is maximized at $i = c\log(1/\beta)$, and thus $\ell(k^*+1) - \ell(k^*+1+i) \leq O\left(\sqrt{\log(1/\beta)\,(\log\log\log(1/\beta) + \log(1/\beta))}\right) - \frac{c\log(1/\beta)}{3} < 0$, where the last inequality follows for a sufficiently large choice of c.

Online Change-Point Detection. The algorithm will be as follows. Partition the stream into consecutive intervals of length n, which we will draw in batches. If an interval has more -1's than +1's, then call the offline change-point detection algorithm on the final $2n$ data points with failure probability parameter set to $\beta/4$, and output whatever it says.

Let k^* be the true change-point index. First, we show that with probability $\geq 1 - \beta/4$, the algorithm will not see more -1's than +1's in any interval before the one containing k^*. The number of +1's in this interval will be distributed as $\mathrm{Binomial}(n, \tau_0)$ for $\tau_0 > 2/3$. By a Chernoff bound, the probability that we have $> n/2$ -1's is at most $\exp(-\Theta(n))$. Taking a union bound over all $O(k^*/n)$ intervals before the change point gives a failure probability of $\frac{k^*}{n}\exp(-\Theta(n)) \leq \beta/4$, where the last inequality follows by our condition on n.

Next, note that the interval following the one containing k^* will have a number of +1's which is distributed as $\mathrm{Binomial}(n, \tau_1)$ for $\tau_1 < 1/3$. By a similar Chernoff bound, the probability that we have $> n/2$ +1's is at most $\exp(-\Theta(n)) \ll \beta/4$. Thus, with probability

$1-\beta/2$, the algorithm will call the offline change-point detection algorithm on an interval containing the true change point k^*.

We conclude by the correctness guarantees of the offline change-point detection algorithm. Note that we chose the failure probablity parameter to be $\beta/4$, as the offline algorithm may either be called at the interval containing k^*, or the following one, and we take a union bound over both of them. □

4.3 Private Goodness-of-Fit CPD

Our reduction above is rather general, and it can apply to more general change-point detection settings. For instance, the above discussion assumes we know both the initial and final distributions P and Q. Instead, one could imagine a setting where one knows the initial distribution P but not the final distribution Q, which we term goodness-of-fit change-point detection.

Definition 4.7. In the *offline γ-goodness-of-fit change-point detection problem*, we are given a distribution P over domain $\mathcal{X}$ and a data set $X = \{x_1, \ldots, x_n\}$. We are guaranteed that there exists $k^* \in [n]$ such that $x_1, \ldots, x_{k^*-1} \sim P$, and $x_{k^*}, \ldots, x_n \sim Q$, for some fixed (but unknown) distribution Q over domain $\mathcal{X}$, such that $\mathrm{TV}(P, Q) \geq \gamma$. The goal is to output $\hat{k}$ such that $|\hat{k} - k^*|$ is small.

We note that analogous definitions and results hold for the online version of this problem, as in the previous sections.

We omit the full details of the proof, but it proceeds by a very similar argument to that in Sections 4.1 and 4.2. In particular, it is possible to prove an analogue of Lemma 4.4, at which point we can apply Lemma 4.5. The only difference is that we need an algorithm for private goodness-of-fit testing, rather than Theorem 1.2 for hypothesis testing. We use the following result from [3].

Theorem 4.8 (Theorem 13 of [3]). *Let P be a known distribution over $\mathcal{X}$, and let $\mathbf{Q}$ be the set of all distributions Q over $\mathcal{X}$ such that $\mathrm{TV}(P, Q) \geq \gamma$. Given n samples from an unknown distribution which is either P, or some $Q \in \mathbf{Q}$, there exists an efficient ε-differentially private algorithm which distinguishes between the two cases with probability $\geq 2/3$ when $n = \Theta\left(\frac{|\mathcal{X}|^{1/2}}{\gamma^2} + \frac{|\mathcal{X}|^{1/2}}{\gamma\varepsilon^{1/2}} + \frac{|\mathcal{X}|^{1/3}}{\gamma^{4/3}\varepsilon^{2/3}} + \frac{1}{\gamma\varepsilon}\right)$.*

With this in hand, we have the following result for goodness-of-fit changepoint detection.

Theorem 4.9. *There exists an efficient ε-differentially private and (α, β)-accurate algorithm for offline γ-goodness-of-fit change-point detection with $\alpha = \Theta\left(\left(\frac{|\mathcal{X}|^{1/2}}{\gamma^2} + \frac{|\mathcal{X}|^{1/2}}{\gamma\varepsilon^{1/2}} + \frac{|\mathcal{X}|^{1/3}}{\gamma^{4/3}\varepsilon^{2/3}} + \frac{1}{\gamma\varepsilon}\right) \cdot \log(1/\beta)\right)$.*

ACKNOWLEDGMENTS

CC was supported by a Motwani Postdoctoral Fellowship. GK was supported as a Microsoft Research Fellow, as part of the Simons-Berkeley Research Fellowship program. AM was supported by NSF award AF-1763786, a Sloan Foundation Research Award, and a postdoctoral fellowship from BU's Hariri Institute for Computing. AS was supported by NSF awards IIS-1447700 and AF-1763786 and a Sloan Foundation Research Award. JU was supported by NSF grants CCF-1718088, CCF-1750640, and CNS-1816028, and a Google Faculty Research Award. We are grateful to Salil Vadhan for valuable discussions in the early stages of this work.

REFERENCES

[1] Jayadev Acharya, Clément L. Canonne, Cody Freitag, and Himanshu Tyagi. 2019. Test without Trust: Optimal Locally Private Distribution Testing. *Proceedings of the 22nd International Conference on Artificial Intelligence and Statistics, AISTATS 2019 (To appear). Full version available at arXiv:1808.02174.* (2019).
[2] Jayadev Acharya, Gautam Kamath, Ziteng Sun, and Huanyu Zhang. 2018. INSPECTRE: Privately Estimating the Unseen. In *Proceedings of the 35th International Conference on Machine Learning (ICML '18)*. JMLR, Inc., 30–39.
[3] Jayadev Acharya, Ziteng Sun, and Huanyu Zhang. 2018. Differentially Private Testing of Identity and Closeness of Discrete Distributions. In *Advances in Neural Information Processing Systems 31 (NeurIPS '18)*. Curran Associates, Inc.
[4] Maryam Aliakbarpour, Ilias Diakonikolas, and Ronitt Rubinfeld. 2018. Differentially Private Identity and Closeness Testing of Discrete Distributions. In *Proceedings of the 35th International Conference on Machine Learning (ICML '18)*. JMLR, Inc., 169–178.
[5] Jordan Awan and Aleksandra Slavkovic. 2018. Differentially Private Uniformly Most Powerful Tests for Binomial Data. In *Advances in Neural Information Processing Systems 31 (NeurIPS '18)*. Curran Associates, Inc., 4212–4222.
[6] Borja Balle, Gilles Barthe, and Marco Gaboardi. 2018. Privacy Amplification by Subsampling: Tight Analyses via Couplings and Divergences. In *Advances in Neural Information Processing Systems 31 (NeurIPS '18)*. Curran Associates, Inc., 6280–6290.
[7] Akshay Balsubramani. 2015. Sharp Finite-Time Iterated-Logarithm Martingale Concentration. *arXiv preprint arXiv:1405.2639* (2015).
[8] Ziv Bar-Yossef. 2002. *The Complexity of Massive Data Set Computations.* Ph.D. Dissertation. UC Berkeley. Adviser: Christos Papadimitriou. Available at http://webee.technion.ac.il/people/zivby/index_files/Page1489.html.
[9] Raef Bassily, Kobbi Nissim, Adam Smith, Thomas Steinke, Uri Stemmer, and Jonathan Ullman. 2016. Algorithmic Stability for Adaptive Data Analysis. In *Proceedings of the 48th Annual ACM Symposium on the Theory of Computing (STOC '16)*. ACM, Cambridge, MA, 1046–1059.
[10] Amos Beimel, Kobbi Nissim, and Uri Stemmer. 2013. Characterizing the Sample Complexity of Private Learners. In *Proceedings of the 4th Conference on Innovations in Theoretical Computer Science (ITCS '13)*. ACM, 97–110.
[11] Aditya Bhaskara, Daniel Dadush, Ravishankar Krishnaswamy, and Kunal Talwar. 2012. Unconditional differentially private mechanisms for linear queries. In *Proceedings of the forty-fourth annual ACM symposium on Theory of computing*. ACM, 1269–1284.
[12] Aleksandr Alekseevich Borovkov. 1999. *Mathematical Statistics.* CRC Press.
[13] Hai Brenner and Kobbi Nissim. 2014. Impossibility of Differentially Private Universally Optimal Mechanisms. *SIAM J. Comput.* 43, 5 (2014), 1513–1540.
[14] Mark Bun and Thomas Steinke. 2016. Concentrated Differential Privacy: Simplifications, Extensions, and Lower Bounds. In *Proceedings of the 14th Conference on Theory of Cryptography (TCC '16-B)*. Springer, Berlin, Heidelberg, 635–658.
[15] Mark Bun, Jonathan Ullman, and Salil Vadhan. 2014. Fingerprinting Codes and the Price of Approximate Differential Privacy. In *Proceedings of the 46th Annual ACM Symposium on the Theory of Computing (STOC '14)*. ACM, New York, NY, USA, 1–10.
[16] Bryan Cai, Constantinos Daskalakis, and Gautam Kamath. 2017. Priv'IT: Private and Sample Efficient Identity Testing. In *Proceedings of the 34th International Conference on Machine Learning (ICML '17)*. JMLR, Inc., 635–644.
[17] Zachary Campbell, Andrew Bray, Anna Ritz, and Adam Groce. 2018. Differentially Private ANOVA Testing. In *Proceedings of the 2018 International Conference on Data Intelligence and Security (ICDIS '18)*. IEEE Computer Society, Washington, DC, USA, 281–285.
[18] Clément L. Canonne. 2017. A short note on distinguishing discrete distributions. https://github.com/ccanonne/probabilitydistributiontoolbox/blob/master/testing.pdf. (2017).
[19] Clément L. Canonne, Gautam Kamath, Audra McMillan, Adam Smith, and Jonathan Ullman. 2018. The Structure of Optimal Private Tests for Simple Hypotheses. *arXiv preprint arXiv:1811.08382* (2018).
[20] Kamalika Chaudhuri, Claire Monteleoni, and Anand D. Sarwate. 2011. Differentially Private Empirical Risk Minimization. *Journal of Machine Learning Research* 12 (2011), 1069–1109.
[21] Simon Couch, Zeki Kazan, Kaiyan Shi, Andrew Bray, and Adam Groce. 2019. Differentially Private Nonparametric Hypothesis Testing. *arXiv preprint arXiv:1903.09364* (2019).
[22] Rachel Cummings, Sara Krehbiel, Yajun Mei, Rui Tuo, and Wanrong Zhang. 2018. Differentially Private Change-Point Detection. In *Advances in Neural Information Processing Systems 31 (NeurIPS '18)*. Curran Associates, Inc., 10848–10857.
[23] Rachel Cummings, Katrina Ligett, Kobbi Nissim, Aaron Roth, and Zhiwei Steven Wu. 2016. Adaptive learning with robust generalization guarantees. In *Conference on Learning Theory, COLT*. 772–814.
[24] Aref N. Dajani, Amy D. Lauger, Phyllis E. Singer, Daniel Kifer, Jerome P. Reiter, Ashwin Machanavajjhala, Simson L. Garfinkel, Scot A. Dahl, Matthew Graham, Vishesh Karwa, Hang Kim, Philip Lelerc, Ian M. Schmutte, William N. Sexton, Lars Vilhuber, and John M. Abowd. 2017. The Modernization of Statistical Disclosure

Limitation at the U.S. Census Bureau. (2017). Presented at the September 2017 meeting of the Census Scientific Advisory Committee.
[25] Differential Privacy Team, Apple. 2017. Learning with Privacy at Scale. https://machinelearning.apple.com/docs/learning-with-privacy-at-scale/appledifferentialprivacysystem.pdf. (December 2017).
[26] John C. Duchi, Michael I. Jordan, and Martin J. Wainwright. 2013. Local Privacy and Statistical Minimax Rates. In *IEEE 57th Annual Symposium on Foundations of Computer Science (FOCS '13)*. 429–438.
[27] John C. Duchi and Feng Ruan. 2018. The Right Complexity Measure in Locally Private Estimation: It is not the Fisher Information. *arXiv preprint arXiv:1806.05756* (2018).
[28] Cynthia Dwork, Vitaly Feldman, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Aaron Roth. 2015. Generalization in Adaptive Data Analysis and Holdout Reuse. In *Advances in Neural Information Processing Systems (NIPS)*.
[29] Cynthia Dwork, Vitaly Feldman, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Aaron Roth. 2015. Preserving Statistical Validity in Adaptive Data Analysis. In *Proceedings of the 47th Annual ACM Symposium on the Theory of Computing (STOC '15)*. ACM, 1046–1059.
[30] Cynthia Dwork, Vitaly Feldman, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Aaron Roth. 2015. The reusable holdout: Preserving validity in adaptive data analysis. *Science* 349, 6248 (June 2015), 636–638.
[31] Cynthia Dwork, Krishnaram Kenthapadi, Frank McSherry, Ilya Mironov, and Moni Naor. 2006. Our Data, Ourselves: Privacy Via Distributed Noise Generation. In *Advances in Cryptology - EUROCRYPT 2006, 25th Annual International Conference on the Theory and Applications of Cryptographic Techniques, St. Petersburg, Russia, May 28 - June 1, 2006, Proceedings*. 486–503.
[32] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. 2006. Calibrating Noise to Sensitivity in Private Data Analysis. In *3rd IACR Theory of Cryptography Conference (TCC '06)*. Springer, New York, NY, 265–284.
[33] Cynthia Dwork and Aaron Roth. 2014. The Algorithmic Foundations of Differential Privacy. *Foundations and Trends in Theoretical Computer Science* 9, 3-4 (2014), 211–407.
[34] Cynthia Dwork and Guy N Rothblum. 2016. Concentrated Differential Privacy. *arXiv preprint arXiv:1603.01887* (2016).
[35] Cynthia Dwork, Adam Smith, Thomas Steinke, Jonathan Ullman, and Salil Vadhan. 2015. Robust Traceability from Trace Amounts. In *Proceedings of the 56th Annual IEEE Symposium on Foundations of Computer Science (FOCS '15)*. IEEE Computer Society, Berkeley, CA, 650–669.
[36] Úlfar Erlingsson, Vasyl Pihur, and Aleksandra Korolova. 2014. RAPPOR: Randomized aggregatable privacy-preserving ordinal response. In *ACM SIGSAC Conference on Computer and Communications Security, CCS*.
[37] Vitaly Feldman and Thomas Steinke. 2017. Generalization for Adaptively-chosen Estimators via Stable Median. In *COLT 2017 - The 30th Annual Conference on Learning Theory*.
[38] Vitaly Feldman and Thomas Steinke. 2018. Calibrating Noise to Variance in Adaptive Data Analysis. In *Proceedings of the 31st Annual Conference on Learning Theory (COLT '18)*. 535–544.
[39] Vitaly Feldman and David Xiao. 2014. Sample complexity bounds on differentially private learning via communication complexity. In *Proceedings of the 27th Annual Conference on Learning Theory (COLT '14)*. 1000–1019.
[40] Marco Gaboardi, Hyun-Woo Lim, Ryan M Rogers, and Salil P Vadhan. 2016. Differentially Private Chi-Squared Hypothesis Testing: Goodness of Fit and Independence Testing.. In *ICML*. 2111–2120.
[41] Marco Gaboardi and Ryan Rogers. 2018. Local Private Hypothesis Testing: Chi-Square Tests. In *Proceedings of the 35th International Conference on Machine Learning (ICML '18)*. JMLR, Inc., 1626–1635.
[42] Alison L. Gibbs and Francis E. Su. 2002. On choosing and bounding probability metrics. *International Statistical Review* 70, 3 (dec 2002), 419–435.
[43] Moritz Hardt and Kunal Talwar. 2010. On the geometry of differential privacy. In *Proceedings of the 42nd ACM Symposium on Theory of Computing, STOC*.
[44] Kazuya Kakizaki, Jun Sakuma, and Kazuto Fukuchi. 2017. Differentially Private Chi-squared Test by Unit Circle Mechanism. In *Proceedings of the 34th International Conference on Machine Learning (ICML '17)*. JMLR, Inc., 1761–1770.
[45] Vishesh Karwa and Salil Vadhan. 2018. Finite Sample Differentially Private Confidence Intervals. In *Proceedings of the 9th Conference on Innovations in Theoretical Computer Science (ITCS '18)*. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, Dagstuhl, Germany, 44:1–44:9.
[46] Shiva Prasad Kasiviswanathan, Homin K. Lee, Kobbi Nissim, Sofya Raskhodnikova, and Adam Smith. 2008. What Can We Learn Privately?. In *FOCS*. IEEE, 531–540.
[47] Shiva Prasad Kasiviswanathan and Adam D. Smith. 2008. A Note on Differential Privacy: Defining Resistance to Arbitrary Side Information. *CoRR* abs/0803.3946 (2008).
[48] Assimakis Kattis and Aleksandar Nikolov. 2016. Lower Bounds for Differential Privacy from Gaussian Width, In 33rd International Symposium on Computational Geometry. *arXiv preprint arXiv:1612.02914*, 45:1–45:16.
[49] Daniel Kifer and Ryan M. Rogers. 2017. A New Class of Private Chi-Square Tests. In *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS '17)*. JMLR, Inc., 991–1000.
[50] Martin Kulldorff. 2001. Prospective Time Periodic Geographical Disease Surveillance Using a Scan Statistic. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 164, 1 (2001), 61–72.
[51] Tze Leung Lai. 1995. Sequential Changepoint Detection in Quality Control and Dynamical Systems. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 57, 4 (1995), 613–658.
[52] Jingbo Liu, Paul W. Cuff, and Sergio Verdú. 2015. E_γ-Resolvability. *CoRR* abs/1511.07829 (2015). arXiv:1511.07829 http://arxiv.org/abs/1511.07829
[53] Gary Lorden. 1971. Procedures for Reacting to a Change in Distribution. *The Annals of Mathematical Statistics* 42, 6 (1971), 1897–1908.
[54] Frank McSherry and Kunal Talwar. 2007. Mechanism Design via Differential Privacy. In *48th Annual IEEE Symposium on Foundations of Computer Science (FOCS '07)*. 94–103.
[55] Yajun Mei. 2006. Sequential Change-Point Detection when Unknown Parameters are Present in the Pre-Change Distribution. *The Annals of Statistics* 34, 1 (2006), 92–122.
[56] George V. Moustakides. 1986. Optimal Stopping Times for Detecting Changes in Distributions. *The Annals of Statistics* 14, 4 (1986), 1379–1387.
[57] Aleksandar Nikolov. 2015. An Improved Private Mechanism for Small Databases. In *Automata, Languages, and Programming - 42nd International Colloquium, ICALP*. 1010–1021.
[58] Aleksandar Nikolov, Kunal Talwar, and Li Zhang. 2016. The Geometry of Differential Privacy: The Small Database and Approximate Cases. *SIAM J. Comput.* 45, 2 (2016), 575–616. https://doi.org/10.1137/130938943
[59] Kobbi Nissim, Sofya Raskhodnikova, and Adam Smith. 2007. Smooth sensitivity and sampling in private data analysis. In *Proceedings of the 30th annual ACM Symposium on Theory of Computing, STOC*. 75–84.
[60] Ewan S. Page. 1954. Continuous Inspection Schemes. *Biometrika* 41, 1/2 (1954), 100–115.
[61] Ewan S. Page. 1955. A Test for a Change in a Parameter Occurring at an Unknown Point. *Biometrika* 42, 3/4 (1955), 523–527.
[62] Moshe Pollak. 1985. Optimal Detection of a Change in Distribution. *The Annals of Statistics* 13, 1 (1985), 206–227.
[63] Moshe Pollak. 1987. Average Run Lengths of an Optimal Method of Detecting a Change in Distribution. *The Annals of Statistics* 15, 2 (1987), 749–779.
[64] Ryan Rogers, Aaron Roth, Adam Smith, and Om Thakkar. 2016. Max-information, differential privacy, and post-selection hypothesis testing. In *IEEE 57th Annual Symposium on Foundations of Computer Science, FOCS*. 487–494.
[65] Daniel Russo and James Zou. 2016. Controlling bias in adaptive data analysis using information theory. In *Artificial Intelligence and Statistics, AISTATS*. 1232–1240.
[66] Or Sheffet. 2018. Locally Private Hypothesis Testing. In *Proceedings of the 35th International Conference on Machine Learning (ICML '18)*. JMLR, Inc., 4605–4614.
[67] Walter Andrew Shewhart. 1931. *Economic Control of Quality of Manufactured Product*. ASQ Quality Press.
[68] Albert N. Shiryaev. 1963. On Optimum Methods in Quickest Detection Problems. *Theory of Probability & Its Applications* 8, 1 (1963), 22–46.
[69] Adam Smith. 2011. Privacy-Preserving Statistical Estimation with Optimal Convergence Rates. In *Proceedings of the 43rd Annual ACM Symposium on the Theory of Computing (STOC '11)*. ACM, New York, NY, USA, 813–822.
[70] Social Science One. 2018. SocialScienceOne. `https://socialscience.one/`. (2018).
[71] Marika Swanberg, Ira Globus-Harris, Iris Griffith, Anna Ritz, Adam Groce, and Andrew Bray. 2019. Improved Differentially Private Analysis of Variance. *Proceedings on Privacy Enhancing Technologies* 2019, 3 (2019).
[72] Caroline Uhler, Aleksandra Slavković, and Stephen E. Fienberg. 2013. Privacy-preserving Data Sharing for Genome-wide Association Studies. *The Journal of Privacy and Confidentiality* 5, 1 (2013), 137–166.
[73] Venugopal V. Veeravalli and Taposh Banerjee. 2014. Quickest Change Detection. *Academic Press Library in Signal Processing* 3 (2014), 209–255.
[74] Duy Vu and Aleksandra Slavković. 2009. Differential Privacy for Clinical Trial Data: Preliminary Evaluations. In *2009 IEEE International Conference on Data Mining Workshops (ICDMW '09)*. IEEE, 138–143.
[75] Yue Wang, Daniel Kifer, Jaewoo Lee, and Vishesh Karwa. 2018. Statistical Approximating Distributions Under Differential Privacy. *The Journal of Privacy and Confidentiality* 8, 1 (2018), 1–33.
[76] Yue Wang, Jaewoo Lee, and Daniel Kifer. 2015. Revisiting Differentially Private Hypothesis Tests for Categorical Data. *arXiv preprint arXiv:1511.03376* (2015).
[77] Yu-Xiang Wang, Jing Lei, and Stephen E. Fienberg. 2016. A Minimax Theory for Adaptive Data Analysis. *CoRR* abs/1602.04287 (2016).
[78] Aolin Xu and Maxim Raginsky. 2017. Information-Theoretic Analysis of Generalization Capability of Learning Algorithms. In *Advances in Neural Information Processing Systems 30 (NIPS '17)*. Curran Associates, Inc., 2524–2533.