Privately Detecting Changes in Unknown Distributions

Rachel Cummings [1] **Sara Krehbiel** [2] **Yuliia Lut** [1] **Wanrong Zhang** [1]

Abstract

The change-point detection problem seeks to identify distributional changes in streams of data. Increasingly, tools for change-point detection are applied in settings where data may be highly sensitive and formal privacy guarantees are required, such as identifying disease outbreaks based on hospital records, or IoT devices detecting activity within a home. Differential privacy has emerged as a powerful technique for enabling data analysis while preventing information leakage about individuals. Much of the prior work on change-point detection—including the only private algorithms for this problem—requires complete knowledge of the pre-change and post-change distributions, which is an unrealistic assumption for many practical applications of interest. This work develops differentially private algorithms for solving the change-point detection problem when the data distributions are unknown. Additionally, the data may be sampled from distributions that change smoothly over time, rather than fixed pre-change and post-change distributions. We apply our algorithms to detect changes in the linear trends of such data streams. Finally, we also provide experimental results to empirically validate the performance of our algorithms.

1. Introduction

The *change-point detection problem* seeks to identify distributional changes in streams of data. It models data points as initially being sampled from a pre-change distribution P_0 and then switching to being sampled from a post-change distribution P_1 at an unknown change-point time k^*. The task is to quickly and accurately identify the change-point time k^*. The change-point problem has been widely studied in theoretical statistics (Shewhart, 1931; Page, 1954; Shiryaev, 1963; Pollak, 1987; Mei, 2008) as well as practical applications including climatology (Lund & Reeves, 2002), econometrics (Bai & Perron, 2003), and DNA analysis (Zhang & Siegmund, 2012).

Much of the previous work on change-point detection focused on the *parametric* setting, where the distributions P_0 and P_1 are perfectly known to the analyst. In this structured setting, the analyst could use algorithms tailored to details of these distributions, such as computing the maximum log-likelihood estimator (MLE) of the change-point time. In this work, we consider the *nonparametric* setting, where these distributions are unknown to the analyst. This setting is closer to practice, as it removes the unrealistic assumption of perfect distributional knowledge. In practice, an analyst may only have sample access to the current (pre-change) distribution, and may wish to detect a change to any distribution that is sufficiently far away, without making specific parametric assumptions on the future (post-change) distribution. The nonparametric setting requires different test statistics, as common approaches like computing the MLE do not work without full knowledge of P_0 and P_1.

In many applications, change-point detection algorithms are applied to sensitive data, and may require formal privacy guarantees. For example, the Center for Disease Control (CDC) may wish to analyze hospital records to detect disease outbreaks, or the Census Bureau may wish to analyze income records to detect changes in employment rates. We will use *differential privacy* (Dwork et al., 2006) as our privacy notion, which has been well-established as the predominant privacy notion in theoretical computer science. Informally, differential privacy bounds the effect of any individual's data in a computation, and ensures that very little can be inferred about an individual from seeing the output of a differentially private analysis. Differential privacy is typically achieved algorithmically by adding noise that scales with the *sensitivity* of a computation, which is the maximum change in the function's value that can be caused by changing a single entry in the database. Privacy for high sensitivity analyses require large amounts of noise, which yield high statistical error (see Section 2.2 for more details).

[1]H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, Georgia, USA [2]Department of Mathematics and Computer Science, Santa Clara University, Santa Clara, California, USA. Correspondence to: Rachel Cummings <rachelc@gatech.edu>; Sara Krehbiel <skrehbiel@scu.edu>; Yuliia Lut <yuliia.lut@gatech.edu>; Wanrong Zhang <wanrongz@gatech.edu>.

Proceedings of the 37^{th} International Conference on Machine Learning, Online, PMLR 119, 2020.

Unfortunately, most nonparametric estimation procedures are not amenable to differential privacy. Indeed, all prior work on private change-point detection has been in the parametric setting, where P_0 and P_1 are known (Cummings et al., 2018; Canonne et al., 2019). A standard approach in the nonparametric setting is to first estimate a parametric model, and then perform parametric change-point detection using the estimated model. Common nonparametric estimation techniques include kernel methods and spline methods (Parzen, 1962; Rosenblatt, 1956) or nonparametric regression (Azzalini et al., 1989). These methods are difficult to make private in part because of the complexity of finite sample error bounds combined with the effect of injecting additional noise for privacy. In contrast, simple rank-based statistics, which order samples by their value, have sensitivity that is easy to analyze.

In this work, we estimate nonparametric change-points using the Mann-Whitney test (Wilcoxon, 1945; Mann & Whitney, 1947), which is a rank-based test statistic, presented formally in Section 2.1. This test picks an index k and measures the fraction of points before k that are greater than points after k. For the change-point problem, the expectation of this statistic is largest around the true change-point k^* (under mild non-degeneracy conditions on the pre- and post-change distributions). This statistic simply computes pairwise comparisons of the observed data, and it does not require any additional knowledge of P_0 or P_1 beyond the assumption that a data point from P_0 is larger than a data point from P_1 with probability $> 1/2$. The test statistic has sensitivity $O(1/n)$ for a database of size n, which is lower than most other test statistics for the same task (Mann & Whitney, 1947).

1.1. Our Results

In this paper, we provide differentially private algorithms for accurate nonparametric change-point detection in both the offline and online settings. We also apply our results to settings where data are not sampled i.i.d., but are instead sampled from distributions changing smoothly over time.

In the offline case, the entire database $X = \{x_1, \ldots, x_n\}$ is given up front, and the analyst seeks to estimate the change-point with small additive error. We use the Mann-Whitney rank-sum statistic and its extension to the change-point setting (Darkhovsky, 1976). At every possible change-point time k, the test measures the fraction of points before k that are greater than points after k using statistic $V(k) = \frac{\sum_{j=k+1}^{n} \sum_{i=1}^{k} I(x_i > x_j)}{k(n-k)}$. The test then outputs the index $\hat{k}$ that maximizes this statistic. Even before adding privacy, we improve the best previously-known finite sample accuracy guarantees of this estimation procedure. The previous non-private accuracy guarantee has $O(n^{2/3})$ additive error (Darkhovsky, 1976), whereas our Theorem 1 in Section 3.1 achieves $O(1)$ additive error.

With these improved accuracy bounds, we give Algorithm 1, PNCPD, in Section 3.2 to make this estimation procedure differentially private. Our algorithm uses the REPORTMAX framework of (Dwork & Roth, 2014). The REPORTMAX algorithm takes in a collection of queries, computes a noisy answer to each query, and returns the index of the query with the largest noisy value. We instantiate this framework with our test statistics $V(k)$ as queries to privately select the argmax of the statistics. One challenge is ensuring that the test statistics $V(k)$ have low enough sensitivity that the additional noise required for privacy does not harm the estimation error by too much. We show that our PNCPD algorithm is differentially private (Theorem 2) and has $O(\frac{1}{\epsilon^{1.01}})$ additive accuracy (Theorem 3), meaning that the accuracy affect of adding privacy is independent of the database size n.

In the online case, the analyst starts with an initial database of size n and receives a stream of additional data points, arriving online. The analyst's goal is to accurately estimate the change-point quickly after it occurs. This is a more challenging setting because the analyst will have very little post-change data if they want to detect changes quickly. In this setting, we give Algorithm 2, ONLINEPNCPD in Section 4. This algorithm uses the ABOVETHRESHOLD framework of (Dwork et al., 2009; 2010). The ABOVETHRESHOLD algorithm 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. Our algorithm computes the test statistic $V(k)$ for the middle index k of each sliding window of the last n data points. Once the algorithm finds a window with a high enough test statistic, it waits for enough additional data points to meet the requirements of our *offline* algorithm PNCPD for accuracy, and then calls PNCPD on the n most recent data points to estimate the change-point time. One technical challenge in the online setting is finding a threshold that is high enough to prevent false positives before a change occurs, and low enough that a true change will trigger a call to the offline algorithm. We show that our ONLINEPNCPD algorithm is differentially private (Theorem 4) and has $O(\log n)$ additive error (Theorem 5).

In Section 5 we report experimental results that empirically validate our theoretical results. We start by applying our PNCPD algorithm to a real-world dataset of stock price time-series data that appear by visual inspection to contain a change-point, and we find that our algorithm finds the correct change-point with minimal error, even for small ϵ values. We then apply our PNCPD algorithm to simulated datasets sampled from Gaussian distributions, varying the parameters corresponding to the size of the distributional change, the location of the change-point in the dataset, and ϵ. We also perform simulations for our application to drift change detection. Finally, we apply our ONLINEPNCPD algorithm to streaming simulated datasets drawn from Gaussian distributions, again varying the parameters that correspond to the size of the distributional change, the location of the change-point in the dataset, and ϵ. In all cases, the empirical accuracy supports our theoretical guarantees.

In Section 6, we apply our results to privately solve the problem of *drift change detection*, where points are sampled from smoothly changing distributions whose means are shifting linearly with time, and the linear drift parameter changes at unknown change-time k^*. We also suggest extensions of this reduction technique to apply our algorithm for non-linear drift change detection.

1.2. Related work

Change-point detection is a canonical problem in statistics that has been studied for nearly a century; selected results include (Shewhart, 1931; Page, 1954; Shiryaev, 1963; Roberts, 1966; Lorden, 1971; Pollak, 1985; 1987; Moustakides, 1986; Lai, 1995; 2001; Kulldorff, 2001; Mei, 2006; 2008; 2010; Chan, 2017). The problem originally arose from industrial quality control, and has since been applied in a wide variety of other contexts including climatology (Lund & Reeves, 2002), econometrics (Bai & Perron, 2003), and DNA analysis (Zhang & Siegmund, 2012). In the parametric setting where pre-change and post-change distributions P_0 and P_1 are perfectly known, the Cumulative Sum (CUSUM) procedure (Page, 1954) is among the most commonly used algorithms for solving the change-point detection problem. It follows the generalized log-likelihood ratio principle, calculating $\ell(k) = \sum_{i=k}^{n} \log \frac{P_1(x_i)}{P_0(x_i)}$ for each $k \in [n]$, and declaring that a change occurs if and only if $\ell(\hat{k}) \geq T$ for MLE $\hat{k} = \text{argmax}_k\ \ell(k)$ and appropriate threshold $T > 0$. Nonparametric change-point detection has also been well-studied in the statistics literature (Darkhovsky, 1976; Carlstein, 1988; Bhattacharyya & Johnson, 1968), and requires different test statistics that do not rely on exact knowledge of the distributions P_0 and P_1.

The only two prior works on differentially private change-point detection (Cummings et al., 2018; Canonne et al., 2019) both considered the parametric setting and employed differentially private variants of the CUSUM procedure and the change-point MLE underlying it. (Cummings et al., 2018) directly privatized non-private procedures for the offline and online settings. (Canonne et al., 2019) gave private change-point detection as an instantiation of a solution to the more general problem of private hypothesis testing, partitioning time series data into batches of size equal to the sample complexity of the hypothesis testing problem, and then outputs the batch number most consistent with a change-point. Both works assumed that the pre- and post-distributions were fully known in advance.

In our nonparametric setting, we use the Mann-Whitney test (Wilcoxon, 1945; Mann & Whitney, 1947) instead of the MLE that the CUSUM procedure is built on. The Mann-Whitney test was originally proposed as a rank-based nonparametric two-sample test, to test whether two samples were drawn from the same distribution using the null hypothesis that after randomly selecting one point from each sample, each point is equally likely to be the larger of the two. It was extended to the change-point setting by (Darkhovsky, 1976), for testing whether samples from before and after the hypothesized change-point were drawn from the same distribution. Given a database $X = \{x_1, \ldots, x_n\}$, for each possible change-point k, the test statistic $V(k) = \frac{\sum_{j=k+1}^{n} \sum_{i=1}^{k} I(x_i > x_j)}{k(n-k)}$ counts the proportion of index pairs (i, j) with $i \leq k < j$ for which $x_i > x_j$. This is a nonparametric test because it does not require any additional knowledge of the distributions from which data are drawn. Additionally, the Mann-Whitney test is known to be more efficient (Gibbons & Chakraborti, 2011) and have lower sensitivity (Mann & Whitney, 1947) than most other test statistics for the same task, including the Wald statistic (Wald & Wolfowitz, 1940) and the Kolmogorov-Smirnov test (Lilliefors, 1967). Differentially private versions of related test statistics have been used in recent unpublished work in the context of hypothesis testing, but they have not been applied to the change-point problem (Couch et al., 2018; 2019).

Despite structural similarities with (Cummings et al., 2018), the analyses for this paper's algorithms are vastly different due to new challenges introduced by the nonparametric setting. Most test statistics for nonparametric estimation have high sensitivity, and therefore require large amounts of noise to be added to satisfy differential privacy. This means that off-the-shelf applications of nonparametric test statistics to the differentially private change-point framework of (Cummings et al., 2018) would result in high error. Indeed, even with our use of the Mann-Whitney test statistic which was chosen for its low sensitivity, an immediate application of the best known finite-sample accuracy bounds (Darkhovsky, 1976) yielded additive error $O(n^{2/3})$ in the offline setting for databases of size n. To achieve our much tighter $O(\epsilon^{-1.01})$ error bounds required a new analysis.

2. Preliminaries

This section provides the necessary background for interpreting our results for the problem of private nonparametric change-point detection. Section 2.1 defines the nonparametric change-point detection problem, Section 2.2 describes the differentially private tools that will be brought to bear. Additional preliminaries can be found in Appendix **??**.

2.1. Change-point background

Let $X = \{x_1, \ldots, x_n\}$ be n real-valued data points. The *change-point detection problem* is parametrized by two distributions, P_0 and P_1. The data points in X are hypothesized to initially be sampled i.i.d. from P_0, but at some unknown change time $k^* \in [n]$, an event may occur (e.g., epidemic disease outbreak) and change the underlying distribution to P_1. The goal of a data analyst is to announce that a change has occurred as quickly as possible after k^*. Since the x_i may be sensitive information—such as individuals' medical information or behaviors inside their home—the analyst will wish to announce the change-point time in a privacy-preserving manner.

In the standard non-private offline change-point literature, the analyst wants to test the null hypothesis $H_0 : k^* = n$, where $x_1, \ldots, x_n \sim_{\text{iid}} P_0$, against the composite alternate hypothesis $H_1 : k^* < n$, where $x_1, \ldots, x_{k^*} \sim_{\text{iid}} P_0$ and $x_{k^*+1}, \ldots, x_n \sim_{\text{iid}} P_1$. For known P_1 and P_0, the log-likelihood ratio of $k^* = \infty$ against $k^* = k$ is

$$\ell(k, X) = \textstyle\sum_{i=k+1}^{n} \log \frac{P_1(x_i)}{P_0(x_i)}.$$

The maximum likelihood estimator (MLE) of the change time k^* is $\text{argmax}_{k\in[n]} \ell(k, X)$. However, note that to perform this test, the analyst must have complete knowledge of distributions P_0 and P_1 to compute the log-likelihood ratio.

In this paper, we consider the situation that we do not know both the pre-change distribution and the post-change distribution. We require no knowledge of the pre- and post-change distributions, and assume only that the probability of an observation from P_0 exceeding an observation from P_1 is different than the probability of an observation from P_1 exceeding an observation from P_0, which is necessary for technical reasons. The Mann-Whitney test (Wilcoxon, 1945) is a commonly used nonparametric test of the null hypothesis that it is equally likely that a randomly selected value from one sample will be less than or greater than a randomly selected value from a second sample. (Darkhovsky, 1976) proposed a modification of the Mann-Whitney test to solve the change-point estimation problem. For each possible change-point k, a test statistic counting the proportion of index pairs (i, j) with $i \le k, j > k$ for which $x_i > x_j$ is calculated as follows:

$$V(k, X) = \frac{\sum_{j=k+1}^{n} \sum_{i=1}^{k} I(x_i > x_j)}{k(n-k)} \tag{1}$$

For data X drawn according to the change-point model with distributions P_0, P_1, this statistic is largest or smallest in expectation at the true change-point k^* depending on the value $a = \Pr_{x_0 \sim P_0, x_1 \sim P_1}[x_0 > x_1]$. If $a > 1/2$, we estimate the change-point by taking the arg max of the Mann-Whitney statistics; otherwise we take the arg min. When X is clear from context, we will simply write $V(k)$. The estimator $\hat{k}$ is understood to denote the argmax or argmin of $V(k)$ depending on whether $a > 1/2$.

We will measure the additive error of our estimations of the true change-point as follows.

Definition 1 ((α, β)-accuracy)**.** *A change-point detection algorithm that produces a change-point estimator $\tilde{k}$ is* (α, β)-accurate *if* $\Pr[|\tilde{k} - k^*| > \alpha] \le \beta$*, where the probability is taken over randomness of the data X sampled according to the change-point model with true change-point k^* and (possibly) the randomness of the algorithm.*

2.2. Differential privacy background

Differential privacy bounds how much a single data entry can affect analysis performed on the database. Databases X, X' are *neighboring* if they differ in at most one entry.

Definition 2 (Differential Privacy (Dwork et al., 2006))**.** *An algorithm $\mathcal{M} : \mathbb{R}^n \to \mathcal{R}$ is* ϵ-differentially private *if for every pair of neighboring databases $X, X' \in \mathbb{R}^n$, and for every subset of possible outputs $\mathcal{S} \subseteq \mathcal{R}$,*

$$\Pr[\mathcal{M}(X) \in \mathcal{S}] \le \exp(\epsilon) \Pr[\mathcal{M}(X') \in \mathcal{S}].$$

One common technique for achieving differential privacy is by adding Laplace noise. The *Laplace distribution* with scale b is the distribution with probability density function: $\text{Lap}(x|b) = \frac{1}{2b} \exp\left(-\frac{|x|}{b}\right)$. We will write $\text{Lap}(b)$ to denote the Laplace distribution with scale b, or (with a slight abuse of notation) to denote a random variable sampled from $\text{Lap}(b)$. The *sensitivity* of a function or query f is defined as $\Delta(f) = \max_{\text{neighbors } X,X'} |f(X) - f(X')|$, and it determines the scale of noise that must be added to satisfy differential privacy. The Laplace Mechanism of (Dwork et al., 2006) takes in a function f, database X, and privacy parameter ϵ, and outputs $f(X) + \text{Lap}(\Delta(f)/\epsilon)$.

One helpful property of differential privacy is that it *composes*, meaning that the privacy parameter degrades gracefully as additional computations are performed on the same database. (Theorem **??** in the appendix.)

Our algorithms use REPORTMAX (Dwork & Roth, 2014), which takes in a collection of queries, computes a noisy answer to each query, and returns the index of the query with the largest noisy value. We use this as the framework for our offline private nonparametric change-point detector PNCPD in Section 3 to privately select the time k with the highest Mann-Whitney statistics $V(k)$.

The ABOVETHRESHOLD algorithm of (Dwork et al., 2009; 2010), refined to its current form by (Dwork & Roth, 2014), 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. We use this algorithm as a framework for our online private nonparametric change-point detector ONLINEPNCPD in Section 4 when new data points arrive online in a streaming fashion. Both ABOVETHRESHOLD and REPORTMAX are reviewed in detail in Appendix **??**.

3. Offline private nonparametric change-point detection

In this section, we give an offline private algorithm for change-point detection when the pre- and post-change distributions are unknown. In Section 3.1, we first offer the finite sample accuracy guarantee for the non-private nonparametric algorithm given by $\hat{k} = \text{argmax}\, V(k)$ for the test statistic $V(k)$ given in Equation (1), which will serve as the baseline for evaluating the utility of our private algorithm. Then in Section 3.2 we present our private algorithm, and give privacy and accuracy guarantees. All omitted proofs are given in Appendix **??**.

3.1. Finite sample accuracy guarantee for the non-private nonparametric estimator

In this section, we provide error bounds for the non-private nonparametric change-point estimator when the data are drawn from two unknown distributions P_0, P_1 with true change-point $k^* \in \{\lceil \gamma n \rceil, \dots, \lfloor (1-\gamma)n \rfloor\}$, for some known $\gamma < 1/2$. This γ bounds away from the change-point occurring too early or too late in the sample, and is necessary to ensure sufficient number of samples from both the pre-change and post-change distributions. Without loss of generality, we assume that $a := \Pr_{x_0 \sim P_0, x_1 \sim P_1}[x_0 > x_1] > 1/2$.

For the non-private task, we use the following estimation procedure of (Darkhovsky, 1976), which calculates the estimated change-point $\hat{k}$ as the argmax of $V(k)$ over all k in the range permitted by γ:

$$\hat{k} = \text{argmax}_{k \in \{\lceil \gamma n \rceil, \dots, \lfloor (1-\gamma)n \rfloor\}} V(k),$$

for test statistic $V(k)$ defined in Equation (1). We show in Theorem 1 that the additive error of this procedure is constant with respect to the sample size n.

Our result is much tighter that the previously known finite-sample accuracy result in (Darkhovsky, 1976), which gave an estimation error bound of $O(n^{2/3})$. This sublinear result comes from a connection between the accuracy and the maximal deviation of $V(k)$ from the expected value over $[\gamma n, (1-\gamma)n]$. To bound the maximal deviation, (Darkhovsky, 1976) first analyzed the variance approximation of $V(k)$ to bound the deviation for a single point k. Then they utilized a Lipschitz property to partition $[\gamma n, (1-\gamma)n]$ to small intervals, and took a union bound over these intervals to yield a high probability guarantee. In contrast, we better leverage the connection between $V(k)$ and $V(k^*)$ for improved accuracy and a simplified proof. At a high level, we show that the expectation of $V(k)$ is single-peaked around k^*, and $V(k) - V(k^*)$ is subgaussian. We carefully analyze the discrete derivative as a function of $|k^* - k|$, γ, and n to use a concentration bound yielding our constant error result.

Theorem 1. *For data $X = \{x_1, \dots, x_n\}$ drawn according to the change-point model with any distributions P_0, P_1 with $a = \Pr_{x \sim P_0, y \sim P_1}[x > y] > 1/2$, constraint $\gamma \in (0, 1/2)$, and change-point $k^* \in \{\lceil \gamma n \rceil, \dots, \lfloor (1-\gamma)n \rfloor\}$, we have that the estimator*

$$\hat{k} = \text{argmax}_{k \in [\lceil \gamma n \rceil, \lfloor (1-\gamma)n \rfloor]} \frac{\sum_{j=k+1}^{n} \sum_{i=1}^{k} I(x_i > x_j)}{k(n-k)}$$

is (α, β)-accurate for any $\beta > 0$ and

$$\alpha = C \cdot \left(\frac{1}{\gamma^4 (a - 1/2)^2} \right)^c \cdot \log \frac{1}{\beta}$$

for any constant $c > 1$ and for $C > 0$ depending only on c.

If $a < 1/2$ we achieve the same error bound using

$$\hat{k} = \text{argmin}_{k \in [\lceil \gamma n \rceil, \lfloor (1-\gamma)n \rfloor]} \frac{\sum_{j=k+1}^{n} \sum_{i=1}^{k} I(x_i > x_j)}{k(n-k)}.$$

3.2. Private offline algorithm

We now give a differentially private version of the nonparametric estimation procedure of (Darkhovsky, 1976), in Algorithm 1. Our algorithm uses REPORTMAX as a private subroutine, instantiated with queries $V(k)$ to privately compute $\text{argmax}\, V(k)$. We show that our algorithm is differentially private (Theorem 2) and produces an estimator with additive accuracy that is constant with respect to the sample size n (Theorem 3).

Algorithm 1 Private Nonparametric Change-Point Detector: PNCPD(X, ϵ, γ)

Input: Database $X = \{x_1, \dots, x_n\}$, privacy parameter ϵ, contraint parameter γ.
for $k = \{\lceil \gamma n \rceil, \dots, \lfloor (1-\gamma)n \rfloor\}$ **do**
  Compute $V(k) = \frac{1}{k(n-k)} \sum_{j=k+1}^{n} \sum_{i=1}^{k} I(x_i > x_j)$
  Sample $Z_k \sim \text{Lap}(\frac{2}{\epsilon \gamma n})$
end for
Output $\tilde{k} = \underset{k \in \{\lceil \gamma n \rceil, \dots, \lfloor (1-\gamma)n \rfloor\}}{\text{argmax}} \{V(k) + Z_k\}$

The crux of the privacy proof involves analyzing the sensitivity of the Mann-Whitney statistic to ensure that sufficient noise is added for the REPORTMAX algorithm to maintain its privacy guarantees. The low sensitivity of this test statistic plays a critical role in requiring only small amounts of noise to preserve privacy. The accuracy proof extends Theorem 1 for the non-private estimator to incorporate the additional error due to the Laplace noise added for privacy. Since the event $V(k) > V(k^*)$ is less probable for k that are further away from k^*, our analysis permits larger values of Laplace noise Z_k for k far from k^*, allowing privacy "for free" with respect to accuracy, for small constants ϵ.

Theorem 2. *For arbitrary data $X = \{x_1, \ldots, x_n\}$, privacy parameter $\epsilon > 0$, and constraint $\gamma \in (0, 1/2)$,* PNCPD*(X, ϵ, γ) is ϵ-differentially private.*

Next we provide an accuracy guarantee for our private algorithm PNCPD when the data are drawn according to the change-point model. The first term in the error bound of Theorem 3 comes from the randomness of the n data points, and the second term is the additional cost that comes from the randomness of the sampled Laplace noises, which quantifies the cost of privacy. To ensure that the cost of privacy is as small as possible, we use *k-specific thresholds* t_k in the proof to optimize the trade-off between how much to tolerate the Laplace noise added for privacy versus the randomness of the data. As $|k - k^*|$ increases, $V(k)$ is less likely to be close to $V(k^*)$, so we can allow more Laplace noise rather than set a universal tolerance for all k.

Theorem 3. *For data $X = \{x_1, \ldots, x_n\}$ drawn according to the change-point model with any distributions P_0, P_1 with $a = \Pr_{x\sim P_0, y\sim P_1}[x > y] > 1/2$, constraint $\gamma \in (0, 1/2)$, change-point $k^* \in \{\lceil \gamma n \rceil, \ldots, \lfloor (1-\gamma)n \rfloor\}$, and privacy parameter $\epsilon > 0$, we have that* PNCPD*(X, ϵ, γ) is (α, β)-accurate for any $\beta > 0$ and*

$$\alpha = \max\left\{ C_1 \cdot \left(\frac{1}{\gamma^4 (a - 1/2)^2}\right)^c \cdot \log\frac{1}{\beta}, \right. \\ \left. C_2 \cdot \left(\frac{1}{\epsilon\gamma(a - 1/2)}\right)^c \cdot \log\frac{1}{\beta} \right\},$$

for any constant $c > 1$ and some constants $C_1, C_2 > 0$ depending on c.

As with our analysis of the non-private estimator, we can take the argmin and get the same error bounds (with $a - 1/2$ replaced by $|a - 1/2|$) if $\Pr_{x\sim P_0, y\sim P_1}[x > y] < 1/2$.

4. Online change point detection

In this section, we show how to extend our results for change-point detection with unknown distributions to the *online setting*, where the database X is not given in advance, but instead data points arrive one-by-one. We assume the analyst starts with a database of size n, and receives one new data point per unit time.

Our algorithm uses the Above Noisy Threshold algorithm of (Dwork et al., 2009; 2010) (AboveThreshold, Algorithm **??**) instantiated with queries of the Mann-Whitney test statistic $V(k)$ in the center of a sliding window of the most recent n points. With each new data point $k > n$, the algorithm computes $V(k)$ for database $X = \{x_{k-n/2+1}, \ldots x_{k+n/2}\}$, and compares this statistic against a noisy threshold for significance. When this statistic is sufficiently high, the online algorithm calls the offline algorithm PNCPD on this window to estimate k^*. For simplicity in indexing and to avoid confusion with the notation of the previous section, we define $U(k) = V(k)$ when $V(k)$ is taken over database X for each $k > n/2$. Since the algorithm only checks for a change-point in the middle of the window, we assume that $k^* \geq n/2$ to ensure that the change-point does not occur too early to be detected.

We note that the offline subroutine PNCPD assumes that a change point occurs sometime after the first γn and before the last γn of the n data points on which it is called. We will show that for an appropriate choice of T, OnlinePNCPD exceeds $\hat{T}$ for some k such that $k^* \in [k, k+n/2]$. Therefore, by waiting for an additional γn data points, we ensure that the assumptions of PNCPD are met as long as $\gamma < 1/4$.

Algorithm 2 Online Private Nonparametric Change-Point Detector: OnlinePNCPD$(X, n, \epsilon, \gamma, T)$

Input: Data stream X, starting size n , privacy parameter ϵ, constraint parameter γ, threshold T.
Let $\hat{T} = T + \text{Lap}\left(\frac{8}{\epsilon n}\right)$
for each new data point $x_{k+n/2}$, $k > n/2$ **do**
  Let $U(k) = \frac{4}{n^2}\sum_{j=k+1}^{k+n/2}\sum_{i=k-n/2+1}^{k} I(x_i > x_j)$
  Sample $Z_k \sim \text{Lap}(\frac{16}{\epsilon n})$
  if $U(k) + Z_k > \hat{T}$ **then**
    Wait for γn new data points to arrive
    Output PNCPD$(\{x_{k-\frac{n}{2}+1+\gamma n}, \ldots x_{k+\frac{n}{2}+\gamma n}\}, \frac{\epsilon}{2}, \gamma)$
    Halt
  end if
end for

Privacy follows immediately from the privacy guarantees of AboveThreshold and PNCPD.

Theorem 4. *For arbitrary data stream X with starting size n, privacy parameter $\epsilon > 0$, and constraint $\gamma \in (0, 1/2)$,* OnlinePNCPD*(X, n, ϵ, γ) is ϵ-differentially private.*

To give accuracy bounds on the performance of OnlinePNCPD, we need to bound several sources of error. First we need to set the threshold T such that the algorithm will not raise a false alarm before the change-point occurs (i.e., control the false positive rate) and that the algorithm will not fail to raise an alarm on a window containing the true change-point (i.e., control the false negative rate). This must be done taking into account the additional error from the private AboveThreshold subroutine. Finally, we can use the accuracy guarantees of PNCPD to show that conditioned on calling a window that contains the true change-point, we are likely to output an estimator $\hat{k}$ that is close to the true change-point k^*.

Theorem 5. *For data stream X with starting size n drawn according to the change-point model with any distributions P_0, P_1 with $a = \Pr_{x \sim P_0, y \sim P_1}[x > y] > 1/2$, constraint $\gamma \in (0, 1/4)$, change-point $k^* \geq n/2$, privacy parameter $\epsilon > 0$, and threshold $T \in [T_L, T_U]$ such that*

$$T_L = \frac{1}{2} + \sqrt{\frac{2}{n}\log(\frac{8(k^* - n/2)}{\beta})} + \frac{32\log((k^* - n/2)/\beta)}{n\epsilon}$$

$$T_U = a - \sqrt{\frac{2}{n}\log\left(\frac{8}{\beta}\right)} - \frac{32\log(8(k^* - n/2)/\beta)}{n\epsilon},$$

we have that ONLINEPNCPD*$(X, n, \epsilon, \gamma, T)$ is (α, β)-accurate for any $\beta > 0$ and*

$$\alpha = \max\left\{C_1 \cdot \left(\frac{1}{\gamma^4(a - 1/2)^2}\right)^c \cdot \log\frac{n}{\beta},\right.$$
$$\left. C_2 \cdot \left(\frac{1}{\epsilon\gamma(a - 1/2)}\right)^c \cdot \log\frac{n}{\beta}\right\},$$

for any constant $c > 1$ and some constants $C_1, C_2 > 0$ which depend only on c.

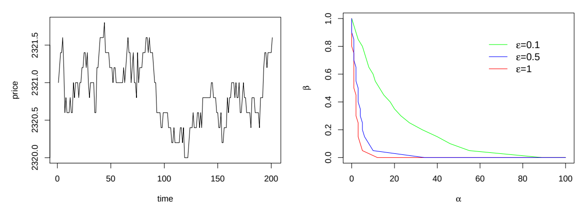
For any starting database size that is at least this large (only $n = \Omega((\frac{\log(k^*/\beta)}{\epsilon(a-1/2)})^2)$), the acceptable region $[T_L, T_U]$ for a threshold T will be non-empty. Moreover, the $\log k^*$ dependence of T_L and T_U means that only a rough estimate of the true change-point is necessary in practice to choose an acceptable threshold T.

5. Empirical Results

We now report the results of an experiment on real data followed by Monte Carlo experiments designed to validate the theoretical results of previous sections. We only consider our accuracy guarantees because differential privacy provides a worst-case guarantee for all hypothetical databases. Our simulations consider both offline and online settings for detecting a change in the mean of Gaussian distribution.

5.1. Results of Offline Algorithm with Real Data

First we illustrate the effectiveness of our offline algorithm on real data by applying it to a window of stock price data including a sudden drop in price, and we use it to determine approximately when this change-point occurred. We use a dataset from (Cao et al., 2018), which contains stock price data over time, with prices collected every second over a span of 5 hours on October 9, 2012. We identified by visual inspection a window of $n = 200$ seconds (indexed 6900 to 7100 in the dataset, reindexed 0 to 200 here) that appears to include a discrete change in distribution from higher mean price to lower mean price. We then calculated the argmax of the Mann-Whitney statistic $V(k)$ to identify the most likely change-point as time $\hat{k} = 92$, assuming the pre-change data were drawn i.i.d. from one distribution and the post-change data were drawn i.i.d. from a distribution with lower mean. We used this estimate as the ground truth ($k^* = \hat{k} = 92$) in error analysis of our private offline algorithm. We ran our PNCPD algorithm with $\gamma = 0.1$ on the selected dataset 10^3 times for each privacy value $\epsilon = 0.1, 0.5, 1$. Figure 1(a) plots the data in our selected window, and Figure 1(b) plots the empirical accuracy $\beta = \Pr[|\tilde{k} - k^*| > \alpha]$ as a function of α for our PNCPD simulations.



(a) Data trajectory

(b) Accuracy of PNCPD on data from (a)

Figure 1. Real data and accuracy results.

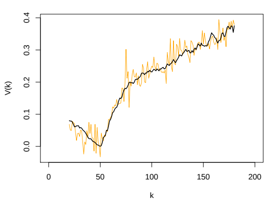
5.2. Offline Results with Synthetic Data

We now provide simulations of our algorithms using synthetic datasets drawn according to the change-point model. We use an initial distribution of $\mathcal{N}(0, 1)$ and post-change distributions of the form $\mathcal{N}(\mu_1, 1)$, considering both a small change $\mu_1 = 1$ and a large change $\mu_1 = 5$. We use $n = 200$ observations with true change $k^* = 50, 100, 150$. This process is repeated 10^3 times for each value of k^* and μ_1. We consider the performance of our algorithm for $\gamma = 0.1$ and $\epsilon = 0.1, 1, 5, \infty$, where $\epsilon = \infty$ corresponds to the non-private problem, which serves as our baseline. The empirical probabilities $\beta = \Pr[|\tilde{k} - k^*| > \alpha]$ as a function of α are summarized in Appendix **??** in Figure **??**.

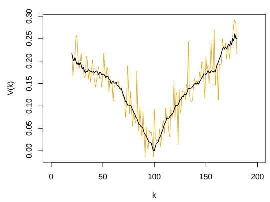
As expected, the algorithm finds the change-point accurately, with better performance when the distributional change is larger or the ϵ value is larger. Performance is slightly diminished when the change-point is at the center of the window, corresponding to $k^* = 100$ in our experiments. This is due to the scaling factor $\frac{1}{k(n-k)}$ in the expression of $V(k)$ as seen in Equation (1), which places relatively higher weight on k that are close to the beginning and end of the window.

We also note that our simulations use slightly larger ϵ values and distributional changes than previous work on *parametric* private change-point detection, where the pre- and post-change distributions are given explicitly as input to the algorithm (Cummings et al., 2018).This is to be expected since the nonparametric problem is information theoretically harder to solve.

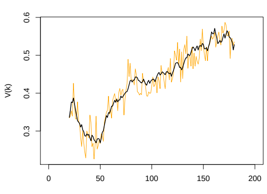
Figure 2 illustrates these accuracy guarantees, showing the values of the true test statistic $V(k)$ and the noisy test statistic $V(k) + Z_k$ for the same distributions. We still use $n = 200$ observations and $k^* = 50, 100$ and $\mu_1 = 1, 5$, and run the process only once for each pair of parameter values. The smoother black line in the figures corresponds to the noiseless test statistic $V(k)$ and the more jagged orange line corresponds to the noisy test statistic $V(k) + Z_k$ for $\epsilon = 5$. Figure 2 shows that in all cases, $V(k)$ is minimized at k^*. This is even more prominent when the distributional change is larger ($\mu_1 = 5$), tolerating more noise. This illustrates the structure of the proof of Theorem 3, and in particular Equation (**??**), where we separate out the failure probability of the algorithm into two terms: the probability of bad data and the probability of bad draws from the Laplace distribution.



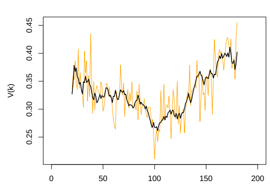
(a) $k^* = 50, \mu_1 = 5$



(b) $k^* = 100, \mu_1 = 5$



(c) $k^* = 50, \mu_1 = 1$



(d) $k^* = 100, \mu_1 = 1$

Figure 2. Value for statistics $V(k)$ with (orange) and without (black) Laplace noise with privacy parameter $\epsilon = 5$ for varying settings for the size change and location of a change point.

5.3. Online Results with Synthetic Data

We also perform simulations for our online private change-point detection algorithm ONLINEPNCPD when the data points arrive sequentially. We use an initial distribution of $\mathcal{N}(5, 1)$ and post-change distribution of $\mathcal{N}(0, 1)$, where the true change occurs at time $k^* = 5000$. To help ensure that the range of the appropriate threshold T in ONLINEPNCPD is non-empty, we choose a larger window size $n = 500$, and larger privacy parameter $\epsilon = 1, 5, 10, \infty$.

We choose the appropriate threshold T by setting a constraint that an algorithm must have positive and negative false alarm rates both at most 0.1, which can be ensured by setting $\beta = 0.4$. (Recall from the proof of Theorem 5 that our false alarm rates are each $\beta/4$.)

Since we know k^* and a, we can compute the theoretical upper and lower bounds on the threshold exactly for the distributions used in our simulations using the expressions given in the statement of Theorem 5. The resulting lower bounds are $T_L = 1.28, 0.80, 0.74, 0.69$ and the upper bounds are $T_U = 0.16, 0.74, 0.81, 0.89$ for $\epsilon = 1, 5, 10, \infty$, respectively. Although the theoretical range of T is empty for $\epsilon = 1, 5$, our empirical results show that $T = 0.8$ is sufficient to control both false alarm rates, as the theoretical bounds are overly conservative. We choose $T = 0.8$ for all $\epsilon = 1, 5, 10, \infty$. In practice when a and k^* are unknown, the analyst should set a to be the smallest interesting magnitude of distributional change, and k^* to be the analyst's estimate of the time of the change, and similarly compute T_L and T_U using these estimates. We also note the analyst can also choose the lower and upper bounds of T via numerical methods as in (Cummings et al., 2018).

We run our ONLINEPNCPD algorithm 10^3 times with $\gamma = 0.1$ and privacy parameters $\epsilon = 1, 5, 10, \infty$. Figure **??** in Appendix **??** shows these simulation results. As in the proof of Theorem 5, we can separate the error into two possible sources within the algorithm: halting on an incorrect window, and producing an incorrect estimate of the change-point, even conditioned on halting on the correct window. Figure **??**(a) shows the error from both of these sources, and Figure **??**(b) shows the error from only the latter source. These figures show that our algorithm works well with privacy parameters $\epsilon = 5, 10, \infty$. For $\epsilon = 1$, we can control the overall error rate to be less than 0.4 as desired, but not much lower. Figure **??**(b) shows that this error mainly comes from the failure to halt on the window that contains the true change-point, because the error decreases dramatically after conditioning on the algorithm halting on a correct window that contains the true change-point.

6. Application: Drift Change Detection

In this section, we extend our consideration of the change-point problem to the setting where data are not sampled i.i.d. from fixed pre- and post-change distributions, but instead are sampled from distributions that are changing smoothly over time. In particular, we consider distributions with *drift*, where the parameter of the distribution changes linearly with time, and the rate of linear drift changes at the change-point. Since the samples are not i.i.d., we consider differences between successive pairs of samples in order to apply the algorithms from the previous sections.

The *drift change detection problem* is parametrized by error terms e_t independently sampled from a mean-zero distribution $\mathcal{S}$, two drift terms ξ_0 and ξ_1, a drift change-point $t^* \in [n]$, and a mean η associated with t^*. Independent random variables $X = \{x_1, \ldots, x_n\}$ are said to be drawn from the drift change detection model if we can write,

$$x_t = \mu_t + e_t,$$

for μ_t piecewise linear as follows:

$$\mu_t = \begin{cases} \eta - (t^* - t)\xi_0 & t \le t^* \\ \eta + (t - t^*)\xi_1 & t > t^* \end{cases}.$$

Our goal is to estimate t^* with the smallest possible error.

To apply our algorithms that require i.i.d. samples, we will transform the sample X by considering differences of consecutive pairs of x_i. These differences are i.i.d. with mean ξ_0 before t^*, and i.i.d. with mean ξ_1 after t^*, and we can now apply PNCPD to this instance of change-point detection. For simplicity, we assume n is even and t^* is odd.

Formally, define a new sample $Y = \{y_1, \ldots, y_{n/2}\}$ with sample points $y_t = x_{2t} - x_{2t-1}$, for $t = 1, \ldots n/2$. Then,

$$y_t = \begin{cases} \xi_0 + e_{2t} - e_{2t-1}, & \text{for } t = 1, \ldots, \frac{t^*-1}{2}, \\ \xi_1 + e_{2t} - e_{2t-1}, & \text{for } t = \frac{t^*+2}{2}, \ldots, \frac{N}{2}. \end{cases}$$

Note that random variables $(e_{2t} - e_{2t-1})$ are independent and identically distributed. Thus y_t are independent and drawn from a fixed distribution before the change-point, and from another distribution after the change. We can apply the PNCPD algorithm and privately estimate the drift change-point $\hat{t}$ as twice the output of PNCPD($\{y_1, \ldots, y_{n/2}\}, \epsilon, \gamma$). This estimation procedure inherits the privacy and accuracy guarantees of Theorems 2 and 3.[1]

As a concrete example, consider points sampled from a Gaussian distribution with mean $\mu_t = \xi_0 t + \eta_0$ and standard deviation σ for $t \le t^*$, and from a Gaussian distribution with mean $\mu_t = \xi_1 t + \eta_1$ and standard deviation σ for $t > t^*$. Then $y_t = x_{2t} - x_{2t-1}$ will be Gaussian with variance $2\sigma^2$ and mean ξ_0 before the change-point and ξ_1 after it. If any of the parameters ξ_0, ξ_1, or σ are unknown, this would require nonparametric change-point estimation.

Corollary 6. *For data $X = \{x_1, \ldots, x_n\}$ drawn according to the drift change model with drift terms $\xi_0 > \xi_1$, constraint $\gamma \in (0, 1/2)$, drift change time $t^* \in (\lceil \frac{\gamma}{2} n \rceil \ldots \lceil (1 - \frac{\gamma}{2}) n \rceil)$, and privacy parameter $\epsilon > 0$, there exists an ϵ-differentially private nonparametric change point estimator that is (α, β)-accurate for any $\beta > 0$ and*

$$\alpha = \max\left\{ C_1 \cdot \left(\frac{1}{\gamma^4 (a - 1/2)^2} \right)^c \cdot \log \frac{1}{\beta}, \right.$$
$$\left. C_2 \cdot \left(\frac{1}{\epsilon \gamma (a - 1/2)} \right)^c \cdot \log \frac{1}{\beta} \right\},$$

for any constant $c > 1$ and some constants $C_1, C_2 > 0$ depending only on c.

We note that this approach is not restricted solely to offline linear drift detection. The same reduction in the online setting would allow us to use ONLINEPNCPD to detect drift changes online. Additionally, a similar approach could be used to for other types of smoothly changing data, as long as the smooth changes exhibited enough structure to allow for reduction to the i.i.d. setting. For example, if data were sampled of the form $x_t = f(\mu_t + e_t)$ for any one-to-one function $f : \mathbb{R} \to \mathbb{R}$, we could define $y_t = f^{-1}(x_{2t}) - f^{-1}(x_{2t-1})$, and these y_ts would again be i.i.d.. This includes random variables of the form $\exp(\mu_t + e_t)$, $\log(\mu_t + e_t)$, and arbitrary polynomials $(\mu_t + e_t)^k$ (where even-degree polynomials must be restricted to, e.g., only have positive range).

We use evaluate the performance of our algorithm for the drift change detection problem on synthetic data with parameters $\eta = 1$, $\xi_0 = 0$, $\xi_1 = 5$, and $e_t \sim_{i.i.d.} \mathcal{N}(0, 1)$. We use $n = 200$ observations where the true drift change occurs at time $t^* = 100$, and repeat the process 10^3 times. We modify the observations X to create a new sample $Y = \{y_1, \ldots, y_{n/2}\}$, and apply our PNCPD algorithm to this new sample. Figure 3 plots the empirical accuracy $\beta = \Pr[|\tilde{t} - k^*| > \alpha]$ as a function of α for $\gamma = 0.1$ and $\epsilon = 0.1, 1, 5, \infty$, where $\epsilon = \infty$ is our non-private baseline.

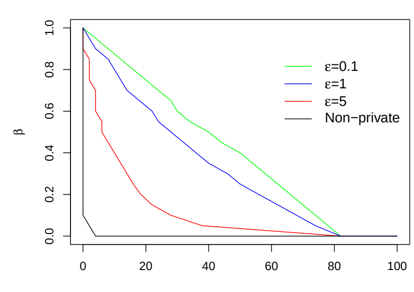


Figure 3. Empirical accuracy $\beta = \Pr[|\tilde{t} - t^*| > \alpha]$ of PNCPD for drift change detection.

[1] This procedure finds a change-point in the sample Y, which corresponds to a pair (x_{2t-1}, x_{2t}) such that one of them is the estimated change point. Under our assumption that t^* is odd, we should output $\hat{t} = 2t - 1$. If t^* is even, then the estimated change-point may be off by one, and $y_{t^*/2}$ is distributed differently than other data points. However, since the PNCPD algorithm is differentially private, its performance is guaranteed be in insensitive to a single outlier in the database, so this fact will not affect the result of the algorithm by too much.

Acknowledgements

R.C. supported in part by a Mozilla Research Grant, a Google Research Fellowship, NSF grants CNS-1850187 and CNS-1942772, and a JPMorgan Chase Faculty Research Award. S.K. supported in part by a Mozilla Research Grant. Y.L. and W.Z. supported in part by a Mozilla Research Grant, NSF grant CNS-1850187, and ARC-TRIAD Fellowships from the Georgia Institute of Technology. This work was completed in part while R.C., Y.L., and W.Z. were visiting the Simons Institute for the Theory of Computing, and while S.K. was visiting the Stanford University Graduate School of Business.

References

Azzalini, A., Bowman, A. W., and Härdle, W. On the use of nonparametric regression for model checking. *Biometrika*, 76(1):1–11, 1989.

Bai, J. and Perron, P. Computation and analysis of multiple structural change models. *Journal of Applied Econometrics*, 18(1):1–22, 2003.

Bhattacharyya, G. and Johnson, R. A. Nonparametric tests for shift at an unknown time point. *The Annals of Mathematical Statistics*, 39(5):1731–1743, 1968.

Canonne, C. L., Kamath, G., McMillan, A., Smith, A., and Ullman, J. The structure of optimal private tests for simple hypotheses. In *Proceedings of the 51st ACM Symposium on Theory of Computing*, STOC '19, pp. 310–321, 2019.

Cao, Y., Xie, Y., and Gebraeel, N. Multi-sensor slope change detection. *Annals of Operations Research*, 263(1-2):163–189, 2018.

Carlstein, E. Nonparametric change-point estimation. *The Annals of Statistics*, 16(1):188–197, 1988.

Chan, H. P. Optimal sequential detection in multi-stream data. *The Annals of Statistics*, 45(6):2736–2763, 2017.

Couch, S., Kazan, Z., Shi, K., Bray, A., and Groce, A. A differentially private wilcoxon signed-rank test. arXiv pre-print 1809.01635, 2018.

Couch, S., Kazan, Z., Shi, K., Bray, A., and Groce, A. Differentially private nonparametric hypothesis testing. arXiv pre-print 1903.09364, 2019.

Cummings, R., Krehbiel, S., Mei, Y., Tuo, R., and Zhang, W. Differentially private change-point detection. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, NeurIPS '18, pp. 10848–10857, 2018.

Darkhovsky, B. A nonparametric method for the a posteriori detection of the "disorder" time of a sequence of independent random variables. *Theory of Probability & Its Applications*, 21(1):178–183, 1976.

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

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

Dwork, C., Naor, M., Reingold, O., Rothblum, G. N., and Vadhan, S. P. 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, pp. 381–390, 2009.

Dwork, C., Naor, M., Pitassi, T., and Rothblum, G. N. Differential privacy under continual observation. In *Proceedings of the 42nd ACM Symposium on Theory of Computing*, STOC '10, 2010.

Gibbons, J. D. and Chakraborti, S. *Nonparametric statistical inference*. Springer, 2011.

Kulldorff, M. Prospective time periodic geographical disease surveillance using a scan statistic. *Journal of the Royal Statistical Society, Series A*, 164(1):61–72, 2001.

Lai, T. L. Sequential changepoint detection in quality control and dynamical systems. *Journal of the Royal Statistical Society, Series B*, 57(4):613–658, 1995.

Lai, T. L. Sequential analysis: some classical problems and new challenges. *Statistica Sinica*, 11(2):303–408, 2001.

Lilliefors, H. W. On the kolmogorov-smirnov test for normality with mean and variance unknown. *Journal of the American statistical Association*, 62(318):399–402, 1967.

Lorden, G. Procedures for reacting to a change in distribution. *The Annals of Mathematical Statistics*, 42(6): 1897–1908, 1971.

Lund, R. and Reeves, J. Detection of undocumented changepoints: A revision of the two-phase regression model. *Journal of Climate*, 15(17):2547–2554, 2002.

Mann, H. B. and Whitney, D. R. On a test of whether one of two random variables is stochastically larger than the other. *The Annals of Mathematical Statistics*, pp. 50–60, 1947.

McDiarmid, C. On the method of bounded differences. In *Surveys in Combinatorics*, pp. 148–188. Cambridge University Press, 1989.

Mei, Y. Sequential change-point detection when unknown parameters are present in the pre-change distribution. *The Annals of Statistics*, 34(1):92–122, 2006.

Mei, Y. Is average run length to false alarm always an informative criterion? *Sequential Analysis*, 27(4):354–419, 2008.

Mei, Y. Efficient scalable schemes for monitoring a large number of data streams. *Biometrika*, 97(2):419–433, 2010.

Moustakides, G. V. Optimal stopping times for detecting changes in distributions. *The Annals of Statistics*, 14(4): 1379–1387, 1986.

Page, E. S. Continuous inspection schemes. *Biometrika*, 41 (1/2):100–115, 1954.

Parzen, E. On estimation of a probability density function and mode. *The annals of mathematical statistics*, 33(3): 1065–1076, 1962.

Pollak, M. Optimal detection of a change in distribution. *The Annals of Statistics*, 13(1):206–227, 1985.

Pollak, M. Average run lengths of an optimal method of detecting a change in distribution. *The Annals of Statistics*, 15(2):749–779, 1987.

Roberts, S. W. A comparison of some control chart procedures. *Technometrics*, 8(3):411–430, 1966.

Rosenblatt, M. Remarks on some nonparametric estimates of a density function. *The Annals of Mathematical Statistics*, pp. 832–837, 1956.

Shewhart, W. A. *Economic Control of Quality of Manufactured Product*. D. Van Norstrand Company, Inc., 1931.

Shiryaev, A. N. On optimum methods in quickest detection problems. *Theory of Probability & Its Applications*, 8(1): 22–46, 1963.

Wald, A. and Wolfowitz, J. On a test whether two samples are from the same population. *The Annals of Mathematical Statistics*, 11(2):147–162, 1940.

Wilcoxon, F. Individual comparisons by ranking methods. *Biometrics bulletin*, 1(6):80–83, 1945.

Zhang, N. and Siegmund, D. O. Model selection for high-dimensional, multi-sequence change-point problems. *Statistica Sinica*, 22(4):1507–1538, 2012.