Single and Multiple Change-Point Detection with Differential Privacy

Wanrong Zhang* Sara Krehbiel† Rui Tuo‡ Yajun Mei* Rachel Cummings*

September 2, 2019

Abstract

The *change-point detection problem* seeks to identify distributional changes at an unknown change-point k^* in a stream of data. This problem appears in many important practical settings involving personal data, including biosurveillance, fault detection, finance, signal detection, and security systems. The field of *differential privacy* offers data analysis tools that provide powerful worst-case privacy guarantees. We study the statistical problem of change-point detection through the lens of differential privacy. We give private algorithms for both online and offline change-point detection, analyze these algorithms theoretically, and provide empirical validation of our results.

Keywords: differential privacy, change-point detection, learning theory, online learning, adaptive data analysis

1 Introduction

The *change-point detection problem* seeks to identify distributional changes at an unknown change-point k^* in a stream of data. The estimated change-point should be consistent with the hypothesis that the data are initially drawn from pre-change distribution P_0 but from post-change distribution P_1 starting at the change-point. This problem appears in many important practical settings, including biosurveillance, fault detection, finance, signal detection, and security systems. For example, the CDC may wish to detect a disease outbreak based on real-time data about hospital visits, or smart home IoT devices may want to detect changes changes in activity within the home. In both of these applications, the data contain sensitive personal information.

The field of *differential privacy* offers data analysis tools that provide powerful worst-case privacy guarantees. Informally, an algorithm that is ϵ-differentially private ensures that any particular output of the algorithm is at most e^ϵ more likely when a single data entry is changed. In the past decade, the theoretical computer science community has developed a wide variety of differentially private algorithms for many statistical tasks. The private algorithms most relevant to this work are based on the simple output perturbation principle that to produce an ϵ-differentially private estimate of some statistic on the database, we should add to the exact statistic noise proportional to Δ/ϵ, where Δ indicates the *sensitivity* of the statistic, or how much it can be influenced by a single data entry.

*School of Industrial and Systems Engineering, Georgia Institute of Technology. Email: {wanrongz, ymei3, rachelc}@gatech.edu. W.Z. supported in part by an ARC-TRIAD fellowship from the Georgia Institute of Technology. Y.M. supported in part by NSF grants DMS-1613258 and DMS-1830344. W.Z. and Y.M. also supported in part by the National Center for Advancing Translational Sciences of the National Institutes of Health under Award Number UL1TR002378. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. R.C. supported in part by a Mozilla Research Grant, a Google Research Fellowship, and NSF grant CNS-1850187. Part of this work was completed while R.C. and W.Z. were visiting the Simons Institute for the Theory of Computing.

†Department of Math and Computer Science, Santa Clara University. Email: skrehbiel@scu.edu. Most of this work was completed while S.K. was employed at the University of Richmond. Supported in part by a Mozilla Research Grant.

‡School of Industrial and Systems Engineering, Texas A&M University. Email: ruituo@tamu.edu. Most of this work was completed while R.T. was visiting Georgia Institute of Technology. Supported in part by NSF grant DMS-156443.

We study the statistical problem of change-point detection through the lens of differential privacy. We give private algorithms for both online and offline change-point detection, analyze these algorithms theoretically, and then provide empirical validation of these results.

1.1 Related work

The change-point detection problem originally arose from industrial quality control, and has since been applied in a wide variety of other contexts including climatology [LR02], econometrics [BP03], and DNA analysis [ZS12]. The problem is studied both in the *offline setting*, in which the algorithm has access to the full dataset $X = \{x_1, \dots, x_n\}$ up front, and in the *online setting*, in which data points arrive one at a time $X = \{x_1, \dots\}$. Change-point detection is a canonical problem in statistics that has been studied for nearly a century; selected results include [She31, Pag54, Shi63, Rob66, Lor71, Pol85, Pol87, Mou86, Lai95, Lai01, Kul01, Mei06, Mei08, Mei10, Cha17].

Our approach is inspired by the commonly used Cumulative Sum (CUSUM) procedure [Pag54]. 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 $\hat{k} = \operatorname{argmax}_k \ell(k)$ and appropriate threshold $T > 0$. The existing change-point literature works primarily in the asymptotic setting when $k_n^*/n \to r$ for some $r \in (0, 1)$ as $n \to \infty$ (see, e.g., [Hin70, Car88]). In contrast, we consider finite databases and provide the first accuracy guarantees for the MLE from a finite sample ($n < \infty$).

In offering the first algorithms for *private* change-point detection, we primarily use two powerful tools from the differential privacy literature. ReportMax [DR14] calculates noisy approximations of a stream of queries on the database and reports which query produced the largest noisy value. We instantiate this with partial log-likelihood queries to produce a private approximation of the the change-point MLE in the offline setting. AboveThresh [DNPR10] calculates noisy approximations on a stream of queries on the database iteratively and aborts as soon as a noisy approximation exceeds a specified threshold. We extend our offline results to the harder online setting, in which a bound on k^* is not known a priori, by using AboveThresh to identify a window of fixed size n in which a change is likely to have occurred so that we can call our offline algorithm on that window to estimate the true change-point.

Recently, [CKM$^+$19] also provided a private change-point detection algorithm based on the more general problem of private hypothesis testing. Their algorithm partitions 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. Their bound gives the minimum number of data points needed to distinguish between two distributions with constant advantage but does not necessarily imply the closest possible approximation of the true change-point. Their accuracy guarantees and ours alike are quantified with respect to distance measures between modified versions of the hypothesized distributions, and comparability of the bounds depends on the specific distributions from which data are drawn.

1.2 Our results

We use existing tools from differential privacy to solve the change-point detection problem in both offline and online settings, neither of which have been studied in the private setting before.

Private offline change-point detection. We develop an offline private change-point detection algorithm OfflinePCPD (Algorithm 3) that is accurate under one of two assumptions about the distributions from which data are drawn. As is standard in the privacy literature, we give accuracy guarantees that bound the additive error of our estimate of the true change-point with high probability. Our accuracy theorem statements (Theorems 6 and 8) also provide guarantees for the non-private estimator for comparison. Since traditional statistics typically focuses on the the asymptotic consistency and unbiasedness of the estimator, ours are the first finite-sample accuracy guarantees for the standard (non-private) MLE. As expected, MLE accuracy decreases with the sensitivity of the measured quantity but increases as the pre- and post-change distribution grow apart. Interestingly, it is constant with respect to the size of the database. In providing MLE bounds alongside accuracy guarantees for our private algorithms, we are able to quantify the cost of privacy as roughly $D_{KL}(P_0||P_1)/\epsilon$.

We are able to prove ϵ-differential privacy under both distributional assumptions, by instantiating the general-purpose ReportMax algorithm from the privacy literature with our log-likelihood queries (Theorem 5). Noting that when the measured quantity has unbounded sensitivity, we introduce a clamping function so that the sensitivity is still bounded by a certain threshold. Importantly and in contrast to our accuracy results, the distributional assumption need only apply to the hypothesized distributions from which data are drawn; privacy holds for arbitrary input databases.

Private online change-point detection. In OnlinePCPD (Algorithm 6), we extend our offline results to the online setting by using the AboveThresh framework to first identify a window in which the change is likely to have happened and then call the offline algorithm to identify a more precise approximation of when it occurred. Standard ϵ-differential privacy under our first distributional assumption follows from composition of the underlying privacy mechanisms (Theorem 9). Accuracy of our online mechanism relies on appropriate selection of the threshold that identifies a window in which a change-point has likely occurred, at which point the error guarantees are inherited from the offline algorithm (Theorem 10).

Empirical validation. Finally, we run several Monte Carlo experiments to validate our theoretical results for both the online and offline settings. We consider data drawn from Bernoulli distribution, which satisfies our first distributional assumption, as well as Gaussian and Gamma distributions, which satisfy our second distributional assumptions. Our offline experiments are summarized in Figure 1, which shows that change-point detection is easier when P_0 and P_1 are further apart and harder when the privacy requirement is stronger (ϵ is smaller). Additionally, these experiments enhance our theoretical results, finding that OfflinePCPD performs well even when we relax the assumptions required for our theoretical accuracy bounds by running our algorithm on imperfect hypotheses P_0 and P_1 that are closer together than the true distributions from which data are drawn. Figure 3 shows that OnlinePCPD also performs well, consistent with our theoretical guarantees.

2 Preliminaries

Our work considers the statistical problem of change-point detection through the lens of differential privacy. Section 2.1 defines the change-point detection problem, Section 2.2 describes the differentially private tools that will be brought to bear, and Section 2.3 give several concentration inequalities which will be used in our proofs.

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 $H_0 : k^* = \infty$, where $x_1, \ldots, x_n \sim_{\text{iid}} P_0$, against the composite alternate hypothesis $H_1 : k^* \in [n]$, where $x_1, \ldots, x_{k^*-1} \sim_{\text{iid}} P_0$ and $x_{k^*}, \ldots, x_n \sim_{\text{iid}} P_1$. The log-likelihood ratio of $k^* = \infty$ against $k^* = k$ is given by

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

The maximum likelihood estimator (MLE) of the change time k^* is given by

$$\hat{k}(X) = \text{argmax}_{k \in [n]} \ell(k, X). \tag{2}$$

When X is clear from context, we will simply write $\ell(k)$ and $\hat{k}$.

We always use log to refer to the natural logarithm, and when necessary, we interpret $\log \frac{0}{0} = 0$. An important quantity in our accuracy analysis will be the Kullback-Leibler distance between probability distributions P_0 and P_1, defined as $D_{KL}(P_1||P_0) = \int_{-\infty}^{\infty} P_1(x) \log \frac{P_1(x)}{P_0(x)} dx = \mathbb{E}_{x \sim P_1}[\log \frac{P_1(x)}{P_0(x)}]$. For given distributions P_0, P_1, our proofs will use the following three variations of KL-divergence:

$$C = \min \{D_{KL}(P_0||P_1), D_{KL}(P_1||P_0)\} \tag{3}$$

$$C_M = \min \left\{ D_{KL}(P_0|| \frac{P_0 + P_1}{2}), D_{KL}(P_1|| \frac{P_0 + P_1}{2}) \right\} = \min_{i=0,1} \mathbb{E}_{x \sim P_i} \left[\log \frac{2P_i(x)}{P_0(x) + P_1(x)} \right] \tag{4}$$

$$C_A = \min \left\{ -\mathbb{E}_{x \leftarrow P_0} \left[\log \frac{P_1(x)}{P_0(x)} \right]_{-A/2}^{A/2}, \mathbb{E}_{x \leftarrow P_1} \left[\log \frac{P_1(x)}{P_0(x)} \right]_{-A/2}^{A/2} \right\}, \tag{5}$$

where A is a pre-specified (input) truncation parameter.

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 $\tilde{k}(X)$ where a distribution change occurred at time k^* is* (α, β)-accurate *if* $\Pr[|\tilde{k} - k^*| < \alpha] \geq 1 - \beta$*, where the probability is taken over randomness of the algorithm and sampling of X.*

2.2 Differential privacy background

Differential privacy bounds the maximum amount that a single data entry can affect analysis performed on the database. Two databases X, X' are *neighboring* if they differ in at most one entry.

Definition 2 (Differential Privacy [DMNS06])**.** *An algorithm $\mathcal{M} : \mathbb{R}^n \rightarrow \mathcal{R}$ is* ϵ-differentially private *if for every pair of neighboring databases $X, X' \in \mathbb{R}^n$, and for every subset of possible outputs $\mathcal{S} \subseteq \mathcal{R}$,*

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

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 $\Delta(f) = \max_{\text{neighbors } X, X'} |f(X) - f(X')|$. The Laplace Mechanism of [DMNS06] takes in a function f, database X, and privacy parameter ϵ, and outputs $f(X) + \text{Lap}(\Delta(f)/\epsilon)$. Since our algorithms estimate a change-point based on log-likelihood ratios, it will be useful to denote the sensitivity of the log-likelihood function given distributions P_0, P_1 as follows:

$$\Delta(\ell) \quad = \quad \max_{x \in \mathbb{R}} \log \frac{P_1(x)}{P_0(x)} - \min_{x' \in \mathbb{R}} \log \frac{P_1(x')}{P_0(x')}. \tag{6}$$

Our algorithms rely on two existing differentially private algorithms, ReportMax [DR14] and AboveThresh [DNPR10]. 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 use this as the framework for our offline private change-point detector OfflinePCPD in Section 3 to privately select the time k with the highest log-likelihood ratio $\ell(k)$.

Algorithm 1 Report Noisy Max: ReportMax$(X, \Delta, \{f_1, \ldots, f_m\}, \epsilon)$

```
Input: database $X$, set of queries $\{f_1, \ldots, f_m\}$ each with sensitivity $\Delta$, privacy parameter $\epsilon$
for $i = 1, \ldots, m$ do
    Compute $f_i(X)$
    Sample $Z_i \sim \text{Lap}(\frac{\Delta}{\epsilon})$
end for
Output $i^* = \operatorname{argmax}_{i \in [m]} (f_i(X) + Z_i)$
```

Theorem 1 ([DR14])**.** ReportMax *is ϵ-differentially private.*

The AboveThresh algorithm, first introduced by [DNPR10] and refined to its current form by [DR14], 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 change-point detector OnlinePCPD in Section 4 when new data points arrive online in a streaming fashion.

Algorithm 2 Above Noisy Threshold: AboveThresh$(X, \Delta, \{f_1, f_2, \ldots\}, T, \epsilon)$

```
Input: database $X$, stream of queries $\{f_1, f_2, \ldots\}$ each with sensitivity $\Delta$, threshold $T$, privacy parameter $\epsilon$
Let $\hat{T} = T + \mathrm{Lap}(\frac{2\Delta}{\epsilon})$
for each query $i$ do
    Let $Z_i \sim \mathrm{Lap}(\frac{4\Delta}{\epsilon})$
    if $f_i(X) + Z_i > \hat{T}$ then
        Output $a_i = \top$
        Halt
    else
        Output $a_i = \bot$
    end if
end for
```

Theorem 2 ([DNPR10])**.** AboveThresh *is ϵ-differentially private.*

Theorem 3 ([DNPR10])**.** *For any sequence of m queries $f_1, \ldots, f_m$ with sensitivity Δ such that $|\{i < m : f_i(X) \geq T - \alpha\}| = 0$,* AboveThresh *outputs with probability at least $1-\beta$ a stream of $a_1, \ldots, a_m \in \{\top, \bot\}$ such that $a_i = \bot$ for every $i \in [m]$ with $f(i) < T - \alpha$ and $a_i = \top$ for every $i \in [m]$ with $f(i) > T + \alpha$ as long as*

$$\alpha \geq \frac{8\Delta \log(2m/\beta)}{\epsilon}.$$

2.3 Concentration inequalities

Our proofs will use the following bounds on partial sums of independent random variables.

Lemma 1 (Ottaviani's inequality [VDVW96])**.** *For independent random variables $U_1, \ldots, U_m$, for $S_k = \sum_{i \in [k]} U_i$ for $k \in [m]$, and for $\lambda_1, \lambda_2 > 0$, we have*

$$\Pr\left[\max_{1 \leq k \leq m} |S_k| > \lambda_1 + \lambda_2\right] \leq \frac{\Pr\left[|S_m| > \lambda_1\right]}{1 - \max_{1 \leq k \leq m} \Pr\left[|S_m - S_k| > \lambda_2\right]}.$$

If we additionally assume the U_i above are i.i.d. with mean 0 and take values from an interval of bounded length L, we can apply Hoeffding's inequality for the following corollary:

Corollary 2. *For independent and identically distributed random variables $U_1, \ldots, U_m$ with mean zero and support strictly bounded by an interval of length L, for $S_k = \sum_{i \in [k]} U_i$ for $k \in [m]$, and for $\lambda_1, \lambda_2 > 0$, we have*

$$\Pr[\max_{k \in [m]} |S_k| > \lambda_1 + \lambda_2] \leq \frac{2\exp(-2\lambda_1^2/(mL^2))}{1 - 2\exp(-2\lambda_2^2/(mL^2))}.$$

When our random variables do not come from a bounded-length interval, we will require Bernstein's inequality instead of Hoeffding's to attain a similar result on their partial sums.

Lemma 3 (Bernstein's inequality [VDVW96])**.** *For independent random variables $Y_1, \ldots, Y_m$ with mean zero such that $\mathbb{E}\left[e^{|Y_i|/M} - 1 - \frac{|Y_i|}{M}\right] M^2 \leq \frac{1}{2} v_i$ for constants M and v_i for all $i \in [m]$, we have*

$$\Pr[|Y_1 + \ldots + Y_n| > x] \leq 2\exp\left(-\frac{1}{2}\frac{x^2}{v + Mx}\right),$$

for $v \geq v_1 + \ldots + v_m$.

Corollary 4. *For independent and identically distributed random variables $Y_1, \ldots, Y_m$ with mean zero such that $\mathbb{E}\left[e^{|Y_i|} - 1 - |Y_i|\right] \le \frac{1}{2}v$, for constant v for all $i \in [m]$, and for $S_k = \sum_{i\in[k]} Y_i$ for $k \in [m]$, and for $\lambda_1, \lambda_2 > 0$, we have*

$$\Pr[\max_{k\in[m]} |S_k| > \lambda_1 + \lambda_2] \le \frac{2\exp(-\lambda_1^2/(2mv + 2\lambda_1))}{1 - 2\exp(-\lambda_2^2/(2mv + 2\lambda_2))}.$$

3 Offline private change-point detection

In this section, we investigate the differentially private change-point detection problem in the setting that n data points $X = \{x_1, \ldots, x_n\}$ are known to the algorithm in advance. Given two hypothesized distributions P_0 and P_1, our algorithms privately approximate the MLE $\hat{k}$ of the change time k^*. We consider accuracy of change-point estimation with and without the assumption that the distributions have uniformly bounded likelihood ratios.

First, we provide finite-sample accuracy guarantees for the MLE in each of these cases in Section 3.1. Second, we offer an algorithm OfflinePCPD in Section 3.2 that achieves privacy by introducing noise proportional to the sensitivity of the log-likelihood calculation. To detect changes in certain distributions such as Gaussians, our OfflinePCPD algorithm requires infinite noise and therefore provides no accuracy. Therefore, we finally provide a second private algorithm OfflinePTCPD in Section 3.3, which has no restriction on the distributions and instead uses a truncation parameter $A > 0$ to control the sensitivity of the log-likelihood calculation. In Table 1 we summarize accuracy bounds for both the MLE and the output of our algorithms under these assumptions.

Quantity	Accuracy guarantee α
MLE $\hat{k}$	$\min\left\{\frac{2\Delta(\ell)^2}{C^2}\log\frac{32}{3\beta}, \frac{35}{C_M^2}\log\frac{32}{3\beta}\right\}$
OfflinePCPD	$\max\left\{\frac{8\Delta(\ell)^2}{C^2}\log\frac{64}{3\beta}, \frac{4\Delta(\ell)}{C\epsilon}\log\frac{16}{\beta}\right\}$
OfflinePTCPD	$\max\left\{\frac{8A^2}{C_A^2}\log\frac{64}{3\beta}, \frac{4A}{C_A\epsilon}\log\frac{16}{\beta}\right\}$

Table 1: Summary of accuracy guarantees for non-private and private offline change-point detection under the alternate hypothesis H_1. The expressions $\hat{k}$, $\Delta(\ell)$, C, C_M and C_A are defined in (2), (6), (3), (4), (5), respectively.

Although our algorithms only guarantee accuracy if the analyst supplies the true distributions P_0, P_1 from which data are drawn, it is important to note that the algorithms are ϵ-differentially private for any *hypothesized* distributions P_0, P_1 and privacy parameter $\epsilon > 0$ *regardless of the distributions from which X is drawn.* In the change-point or statistical process control (SPC) literature, when the pre- and post- change distributions are unknown in practical settings, researchers often choose hypotheses P_0, P_1 with the smallest justifiable distance. While it is easier to detect and accurately estimate a larger change, larger changes are often associated with a higher-sensitivity MLE, requiring more noise (and therefore additional error) or truncation (and therefore information loss) to preserve privacy. We propose that practitioners using our private change-point detection algorithm choose input hypotheses accordingly. This practical setting is considered in our numerical studies, presented in Section 5.

3.1 Finite sample accuracy guarantees for the MLE

Here we provide two accuracy bounds for the standard (non-private) MLE. These are the first finite-sample accuracy guarantees for this estimator. Such non-asymptotic properties have not been previously studied in traditional statistics, which typically focuses on consistency and unbiasedness of the estimator, with less attention to the convergence rate. We show that the additive error of the MLE is constant with respect to the sample size, which means that the convergence rate is $O_P(1)$. These results provide a baseline for quantifying the cost of privacy, since the techniques used in the theorem below mirror those used later in the accuracy proofs for our private algorithms.

A technical challenge that arises in proving accuracy of the estimator is that the x_i are not identically distributed when the true change-point $k^* \in (1, n]$, and so the partial log-likelihood ratios $\ell(k)$ are dependent across k. Hence we need to investigate a sequence of $\ell(k)$ that may be neither independent nor identically distributed. Fortunately, the differences $\ell(k) - \ell(k+1) = \log \frac{P_1(x_k)}{P_0(x_k)}$ are piecewise i.i.d. This property is key in our proof. Moreover, we show that we can divide the possible outputs of the algorithm into regions of doubling size with exponentially decreasing probability of being selected by the algorithm, resulting in accuracy bounds that are independent of the number of data points n.

Note that our first accuracy guarantee depends on two measures $\Delta(\ell)$ and C of the distances between distributions P_0 and P_1. Accuracy is best for distributions for which $\Delta(\ell)$ is small relative to KL-divergence, which is consistent with the intuition that larger changes are easier to detect but output sensitivity degrades the robustness of the estimator, harming accuracy. This will be true for our first private algorithm OfflinePCPD, whose accuracy is additionally harmed by the extra noise required to protect privacy when output sensitivity is higher.

This dependence on $\Delta(\ell)$ is not inherent, however. Allowing $\Delta(\ell)$ to be infinite precludes our use of the same concentration inequalities in obtaining the accuracy guarantee, but the main idea in the proof can be salvaged by decomposing the change from P_0 to P_1 into a change from P_0 to the average distribution $(P_0 + P_1)/2$ and then the average distribution to P_1. Correspondingly, our second accuracy guarantee will use the alternative distance measure

$$C_M = \min \left\{ D_{KL}(P_0 || \frac{P_0 + P_1}{2}), D_{KL}(P_1 || \frac{P_0 + P_1}{2}) \right\},$$

which will allow us provide an MLE accuracy guarantee for arbitrary distributions.

Theorem 4. *For n data points drawn from P_0, P_1 such that $\Delta(\ell) < \infty$ with true change time $k^* \in (1, n]$, the MLE $\hat{k}$ is (α, β)-accurate for any $\beta > 0$ and*

$$\alpha = \frac{2\Delta(\ell)^2}{C^2} \log \frac{32}{3\beta}. \tag{7}$$

For n data points drawn from arbitrary P_0, P_1 with true change time $k^* \in (1, n)$, the MLE $\hat{k}$ is (α, β)-accurate for any $\beta > 0$ and

$$\alpha = \frac{35}{C_M^2} \log \frac{32}{3\beta}, \tag{8}$$

where C and C_M are defined in (3) and (4), respectively.

Proof. Given some true change-point k^* and error tolerance $\alpha > 0$, we can partition the set of bad possible outputs $\hat{k}$ into sub-intervals of exponentially increasing size as follows. For $i \geq 1$, let

$$\begin{aligned} R_i^- &= [k^* - 2^i\alpha, k^* - 2^{i-1}\alpha) \\ R_i^+ &= (k^* + 2^{i-1}\alpha, k^* + 2^i\alpha] \\ R_i &= R_i^- \cup R_i^+ \end{aligned}$$

Then we can bound the probability of the bad event as follows:

$$\beta = \Pr[|\hat{k} - k^*| > \alpha] \leq \sum_{i \geq 1} \Pr[\max_{k \in R_i} \{\ell(k) - \ell(k^*)\} > 0] \tag{9}$$

This requires us to reason about the probability that the log-likelihood ratios for the data are not too far away from their expectation. Although the $\ell(k)$ are not independent across k, their pairwise differences $\ell(k+1) - \ell(k)$ are. When $\Delta(\ell) < \infty$ we can apply our corollary of Ottaviani's inequality (Corollary 2) to bound the probability that $\ell(k)$ exceeds $\ell(k^*)$ by appropriately defining several random variables corresponding to a data stream X drawn according to the change-point model.

Specifically, we can decompose the empirical log-likelihood difference between the true change-point k^* and any candidate k into the expected value of this difference and the sum of i.i.d. random variables with mean zero as follows:

$$U_j := \begin{cases} -\log \frac{P_0(x_j)}{P_1(x_j)} + D_{KL}(P_0||P_1), & j < k^* \\ -\log \frac{P_1(x_j)}{P_0(x_j)} + D_{KL}(P_1||P_0), & j \geq k^* \end{cases}$$

$$\ell(k) - \ell(k^*) = \begin{cases} \sum_{j=k}^{k^*-1} U_j - (k^* - k) D_{KL}(P_0||P_1), & k < k^* \\ \sum_{j=k^*}^{k-1} U_j - (k - k^*) D_{KL}(P_1||P_0), & k \geq k^* \end{cases}$$

We also define random variable S_m to denote the sum of m i.i.d. random variables as follows, noting that S_m is distributed like $\sum_{j=k^*+m}^{k^*-1} U_j$ for $m < 0$ and like $\sum_{j=k^*}^{k^*+m-1} U_j$ for $m > 0$.

$$S_m = \begin{cases} \sum_{k^*+m \leq j < k^*} U_j, & m < 0 \\ \sum_{k^* \leq j < k^*+m} U_j & m > 0 \end{cases}$$

With these random variables, we bound the probability that the MLE lives in any particular bad subinterval $R_i, i \geq 1$ as follows:

$$\begin{aligned}
&\Pr[\max_{k \in R_i}\{\ell(k) - \ell(k^*)\} > 0] \\
&\quad = \Pr[\max_{k \in R_i^-}\{\sum_{j=k}^{k^*-1} U_j - (k^* - k) D_{KL}(P_0||P_1)\} > 0] \\
&\quad\quad + \Pr[\max_{k \in R_i^+}\{\sum_{j=k^*}^{k-1} U_j - (k - k^*) D_{KL}(P_1||P_0)\} > 0] \\
&\quad \leq \Pr[\max_{k \in [2^{i-1}\alpha]} |S_{-k}| > 2^{i-1}\alpha C] + \Pr[\max_{k \in [2^{i-1}\alpha]} |S_k| > 2^{i-1}\alpha C] \\
&\quad \leq \frac{4 \cdot \exp(-2^{i-2}\alpha C^2/\Delta(\ell)^2)}{1 - 2 \cdot \exp(-2^{i-2}\alpha C^2/\Delta(\ell)^2)} && (10) \\
&\quad \leq 8 \exp(-2^{i-2}\alpha C^2/\Delta(\ell)^2) && (11) \\
&\quad = 8 \left(\exp(\frac{-\alpha C^2}{2\Delta(\ell)^2})\right)^{2^{i-1}}
\end{aligned}$$

where the first inequality comes from the definitions of R_i and C, inequality (10) follows from an application of Corollary 2 with $\lambda_1 = \lambda_2 = 2^{i-2}\alpha C$ and $L = \Delta(\ell)$, and the denominator can be simplified as in (11) under the assumption that $\alpha \geq \frac{2\Delta(\ell)^2 \log 4}{C^2}$, which is satisfied by the final bound on α in (7).

We now consider the sum of these terms over all i, which will be needed for the final bound on Equation (9). We note that this sum is bounded above by a geometric series with ratio $\exp(-\alpha C^2/(2\Delta(\ell)^2))$ since $2^{i-1} \geq i$, yielding the second and third inequalities. For the fourth inequality, the same assumed lower bound on α is used to simplify the denominator as in (11):

$$\begin{aligned}
\sum_{i \geq 1} \Pr[\max_{k \in R_i}\{\ell(k) - \ell(k^*)\} > 0] &\leq 8 \sum_{i \geq 1} \left(\exp(\frac{-\alpha C^2}{2\Delta(\ell)^2})\right)^{2^{i-1}} \\
&\leq 8 \sum_{i \geq 1} \left(\exp(\frac{-\alpha C^2}{2\Delta(\ell)^2})\right)^{i} \\
&\leq \frac{8 \exp(\frac{-\alpha C^2}{2\Delta(\ell)^2})}{1 - \exp(\frac{-\alpha C^2}{2\Delta(\ell)^2})} \\
&\leq \frac{32}{3} \exp\left(\frac{-\alpha C^2}{2\Delta(\ell)^2}\right).
\end{aligned}$$

For α as in (7) in the theorem statement, the expression above is bounded by β as required.

In the case that $\Delta(\ell)$ is infinite, we instead define i.i.d. random variables V_j with mean zero, according to an alternative log-likelihood as follows:

$$V_j := \begin{cases} -\log \frac{P_0(x_j)}{(\frac{P_0+P_1}{2})(x_j)} + D_{KL}(P_0||\frac{P_0+P_1}{2}), & j < k^* \\ -\log \frac{P_1(x_j)}{(\frac{P_0+P_1}{2})(x_j)} + D_{KL}(P_1||\frac{P_0+P_1}{2}), & j \geq k^* \end{cases}$$

This new set of random variables is necessary when $\Delta(\ell)$ is infinite, because the U_j no longer have bounded support, so we cannot apply Corollary 2. Instead we will apply a corollary of Bernstein's inequality (Corollary 4) to get similar bounds.

With these random variables, we can bound the empirical log-likelihood difference between the true change-point k^* and any candidate k by,

$$\frac{1}{2}[\ell(k)-\ell(k^*)] \leq \begin{cases} \sum_{j=k}^{k^*-1} V_j - (k^*-k)D_{KL}(P_0||\frac{P_0+P_1}{2}), & k < k^* \\ \sum_{j=k^*}^{k-1} V_j - (k-k^*)D_{KL}(P_1||\frac{P_0+P_1}{2}), & k \geq k^*. \end{cases}$$

The inequality follows by concavity of the log function, which gives that $\frac{1}{2}\log\frac{P_1(x)}{P_0(x)} \leq \log(\frac{P_0+P_1}{2})(x)P_0(x)$ for any x. Next we bound each term in (9) for any $i \geq 1$ as follows, noting that the 1/2 multiplier has no effect when we are only concerned with the maximum being positive:

$$\begin{aligned}
&\Pr[\max_{k\in R_i}\{\ell(k)-\ell(k^*)\} > 0] \\
&\quad\leq \Pr[\max_{k\in R_i^-}\{\sum_{j=k}^{k^*-1} V_j - (k^*-k)D_{KL}(P_0||\frac{P_0+P_1}{2})\} > 0] \\
&\qquad + \Pr[\max_{k\in R_i^+}\{\sum_{j=k^*}^{k-1} V_j - (k-k^*)D_{KL}(P_1||\frac{P_0+P_1}{2})\} > 0] \\
&\quad\leq \Pr[\max_{k\in[2^{i-1}\alpha]}|\textstyle\sum_{j=k^*-k}^{k^*-1} V_j| > 2^{i-1}\alpha C_M] + \Pr[\max_{k\in[2^{i-1}\alpha]}|\textstyle\sum_{j=k^*}^{k^*+k-1} V_j| > 2^{i-1}\alpha C_M] && (12) \\
&\quad\leq \frac{4\exp\left(-\frac{2^{i-3}\alpha C_M^2}{C_M+8}\right)}{1-2\exp\left(-\frac{2^{i-3}\alpha C_M^2}{C_M+8}\right)} && (13) \\
&\quad\leq 8\exp\left(-\frac{2^{i-3}\alpha C_M^2}{C_M+8}\right) && (14)
\end{aligned}$$

where (12) follows from the definitions of R_i and C_M, and (13) follows from an application of Corollary 4 with $\lambda_1 = \lambda_2 = 2^{i-2}\alpha C_M$ and $v = 4$. The denominator is simplified in (14) using our final bound on α in (8) and direct calculations to show that $\alpha \geq \frac{35}{C_M^2}\log\frac{32}{3\beta} > 82/C_M^2$ implies $2\exp\left(-\frac{2^{i-3}\alpha C_M^2}{C_M+8}\right) < 1/2$.

To verify the application of Corollary 4 used in Equation (13), we need to show that for all j,

$$\mathbb{E}\left[\exp(|V_j|) - 1 - |V_j|\right] \leq 2. \tag{15}$$

To show this, let Y_j be the biased i.i.d. alternative log-likelihood ratio as follows:

$$Y_j = \begin{cases} \frac{(\frac{P_0+P_1}{2})(x_j)}{P_0(x_j)}, & j < k^* \\ \frac{(\frac{P_0+P_1}{2})(x_j)}{P_1(x_j)}, & j \geq k^* \end{cases}$$

Because $P_i(x)/(\frac{P_0+P_1}{2})(x) \leq 2$ for $i = 0, 1$, we have $0 \leq D_{KL}(P_i||(P_0+P_1)/2) \leq \log 2$, and thus $e^{D_{KL}(P_i||(P_0+P_1)/2)} \in [1, 2]$. It suffices to note that $\mathbb{E}\left[\exp(|V_j|)\right] \leq 3$, because $\mathbb{E}\left[\exp(|V_j|) - 1 - |V_j|\right] \leq \mathbb{E}\left[\exp(|V_j|) - 1\right]$. We present the analysis when $j < k^*$, and the following expectation is taken under P_0. Note that the other side $j \geq k^*$ is similar with the expectation taken under P_1.

$$\begin{aligned}
\mathbb{E}\left[\exp(|V_j|)\right] &= \mathbb{E}\left[\exp(|\log Y_j - \mathbb{E}\left[\log Y_j\right]|)\right] \\
&\leq \mathbb{E}\left[\exp(\log Y_j - \mathbb{E}\left[\log Y_j\right])\right] + \mathbb{E}\left[\exp(\mathbb{E}\left[\log Y_j\right] - \log Y_j)\right] \\
&= \mathbb{E}\left[Y_j\right] e^{D_{KL}(P_0||(P_0+P_1)/2)} + \frac{\mathbb{E}\left[1/Y_j\right]}{e^{D_{KL}(P_0||(P_0+P_1)/2)}} \\
&\leq e^{D_{KL}(P_0||(P_0+P_1)/2)} + \frac{2}{e^{D_{KL}(P_0||(P_0+P_1)/2)}} && (16) \\
&\leq 2\sqrt{2} \leq 3, && (17)
\end{aligned}$$

where (16) follows from $\mathbb{E}\left[Y_j\right] = 1$, and $\mathbb{E}\left[1/Y_j\right] = \mathbb{E}\left[P_0(x)/(\frac{P_0+P_1}{2})(x)\right] \leq 2$, and (17) follows from the optimization that $x + 2/x \leq 2\sqrt{2}$ for $x \in [1, 2]$.

Finally, we consider the sum of the terms (14) over all i:

$$\begin{aligned}
\sum_{i\geq 1}\Pr[\max_{k\in R_i}\{\ell(k)-\ell(k^*)\}>0] &\leq \sum_{i\geq 1} 8\left(\exp\left(-\frac{\alpha C_M^2}{4(C_M+8)}\right)\right)^{2^{i-1}}\\
&\leq \sum_{i\geq 1} 8\left(\exp\left(-\frac{\alpha C_M^2}{4(C_M+8)}\right)\right)^{i}\\
&\leq \frac{8\exp\left(-\frac{\alpha C_M^2}{4(C_M+8)}\right)}{1-\exp\left(-\frac{\alpha C_M^2}{4(C_M+8)}\right)}\\
&\leq \frac{32}{3}\exp\left(-\frac{\alpha C_M^2}{35}\right). \qquad (18)
\end{aligned}$$

The denominator is simplified in 18 using the condition that $\exp\left(-\frac{\alpha C_M^2}{4(C_M+8)}\right) < 1/4$. For α as in (8) in the theorem statement, the expression above is bounded by β, completing the proof.

□

3.2 Offline algorithm under the uniform bound assumption

Our first private offline algorithm OfflinePCPD applies the report noisy max algorithm [DR14] to the change-point problem by adding Laplace noise with parameter $\Delta(\ell)/\epsilon$ to each finite-sensitivity partial log-likelihood ratio $\ell(k)$ in order to estimate the private change-point. We note that our algorithm can be easily modified to additionally output an approximation of $\ell(\tilde{k})$ and incur 2ϵ privacy cost by composition.

Algorithm 3 Offline private change-point detector: OfflinePCPD$(X, P_0, P_1, \epsilon, n)$

```
Input: database $X$, distributions $P_0, P_1$, privacy parameters $\epsilon$, database size $n$
Let $\Delta(\ell) = \max_x \log\frac{P_1(x)}{P_0(x)} - \min_{x'} \log\frac{P_1(x')}{P_0(x')}$
for $k = 1, \ldots, n$ do
    Compute $\ell(k) = \sum_{i=k}^{n} \log\frac{P_1(x_i)}{P_0(x_i)}$
    Sample $Z_k \sim \mathrm{Lap}(\frac{\Delta(\ell)}{\epsilon})$
end for
Output $\tilde{k} = \underset{1\leq k\leq n}{\mathrm{argmax}}\{\ell(k) + Z_k\}$
```

Privacy of OfflinePCPD follows by instantiation of ReportMax [DR14] with queries $\ell(k)$ for $k \in [n]$, which have sensitivity $\Delta(\ell)$; this proof is included for completeness.

Theorem 5. *For arbitrary data X and $\epsilon > 0$, OfflinePCPD(X, P_0, P_1, ϵ) is ϵ-differentially private.*

Proof. Fix any two neighboring databases X, X' that differ on index j. For any $k \in [n]$, denote the respective partial log-likelihood ratios as $\ell(k)$ and $\ell'(k)$. By (1), we have

$$\ell'(k) = \ell(k) + \Delta\mathbb{I}\{j \geq k\} \qquad \text{with} \qquad \Delta = \log\frac{P_1(x_j')}{P_0(x_j')} - \log\frac{P_1(x_j)}{P_0(x_j)}. \qquad (19)$$

Next, for a given $1 \leq i \leq n$, fix Z_{-i}, a draw from $[\mathrm{Lap}(\Delta(\ell)/\epsilon)]^{n-1}$ used for all the noisy log likelihood ratio values except the ith one. We will bound from above and below the ratio of the probabilities that the algorithm outputs $\tilde{k} = i$ on inputs X and X'. Define the minimum noisy value in order for i to be select with X:

$$Z_i^* = \min\{Z_i : \ell(i) + Z_i > \ell(k) + Z_k \quad \forall k \neq i\}$$

If $\Delta < 0$, then for all $k \neq i$ we have

$$\ell'(i) + \Delta(\ell) + Z_i^* \geq \ell(i) + Z_i^* > \ell(k) + Z_k \geq \ell'(k) + Z_k.$$

If $\Delta \geq 0$, then for all $k \neq i$ we have

$$\ell'(i) + Z_i^* \geq \ell(i) + Z_i^* > \ell(k) + Z_k \geq \ell'(k) - \Delta(\ell) + Z_k.$$

Hence, $Z_i' \geq Z_i^* + \Delta(\ell)$ ensures that the algorithm outputs i on input X', and the theorem follows from the following inequalities for any fixed Z_{-i}, with probabilities over the choice of $Z_i \sim \mathrm{Lap}(\Delta(\ell)/\epsilon)$.

$$\Pr[\tilde{k} = i \mid X', Z_{-i}] \geq \Pr[Z_i' \geq Z_i^* + A \mid Z_{-i}] \geq e^{-\epsilon} \Pr[Z_i \geq Z_i^* \mid Z_{-i}] = e^{-\epsilon} \Pr[\tilde{k} = i \mid X, Z_{-i}]$$

□

Next we provide an accuracy guarantee for the output $\tilde{k}$ of our private algorithm OfflinePCPD when the data are drawn from P_0, P_1 with true change point $k^* \in (1, n)$. By providing this bound using a technique mirroring that of Theorem 4 to bound the error of the non-private MLE, Theorem 6 quantifies the marginal cost of requiring privacy in change-point detection. This additional cost comes from the fact that not only may the randomness of the n data points X result in an incorrect MLE, but the randomness of the Laplace noise added for privacy may also result in an incorrect noisy estimate of the MLE.

Theorem 6. *For hypotheses P_0, P_1 such that $\Delta(\ell) < \infty$ and n data points X drawn from P_0, P_1 with true change time $k^* \in (1, n]$, and for privacy parameter $\epsilon > 0$,* OfflinePCPD$(X, P_0, P_1, \epsilon, n)$ *is (α, β)-accurate for any $\beta > 0$ and*

$$\alpha = \max\left\{ \frac{8\Delta(\ell)^2}{C^2} \log \frac{64}{3\beta}, \frac{4\Delta(\ell)}{C\epsilon} \log \frac{16}{\beta} \right\}. \tag{20}$$

Proof. Our proof is structured around the observation that the algorithm only outputs a particular incorrect $\tilde{k} \neq k^*$ if there exists some k in with $\ell(k) + Z_k > \ell(k^*) + Z_{k^*}$ for a set of random noise values $\{Z_k\}_{k \in [n]}$ selected by the algorithm. For the algorithm to output an incorrect value, there must either be a k that nearly beats the true change point on the noiseless data or there must be a k that receives much more noise than k^*. Intuitively, this captures the respective scenarios that unusual data causes non-private ERM to perform poorly and that unusual noise draws causes our private algorithm to perform poorly.

As in the proof of Theorem 4, given some true change-point k^* and error tolerance $\alpha > 0$, we partition the set of bad possible outputs k into sub-intervals of exponentially increasing size. For $i \geq 1$, let

$$\begin{aligned} R_i^- &= [k^* - 2^i \alpha, k^* - 2^{i-1}\alpha) \\ R_i^+ &= (k^* + 2^{i-1}\alpha, k^* + 2^i \alpha] \\ R_i &= R_i^- \cup R_i^+ \end{aligned}$$

Then for any range-specific thresholds t_i for $i \geq 1$, our previous observations allow us to bound the probability of the bad event as follows:

$$\beta = \Pr[|\tilde{k} - k^*| > \alpha] \leq \sum_{i \geq 1} \Pr[\max_{k \in R_i}\{\ell(k) - \ell(k^*)\} > -t_i] + \sum_{i \geq 1} \Pr[\max_{k \in R_i}\{Z_k - Z_{k^*}\} \geq t_i] \tag{21}$$

We bound each term in the above expression separately for $t_i = 2^{i-2}\alpha C$, and we will set α to ensure that each term is at most $\beta/2$. As in Theorem 4, we can bound the first set of terms using Corollary 2 to bound the probability that $\ell(k)$ significantly exceeds $\ell(k^*)$ by appropriately defining several random variables corresponding to a data stream X drawn according to the change-point model:

$$U_j = \begin{cases} -\log \frac{P_0(x_j)}{P_1(x_j)} + D_{KL}(P_0||P_1), & j < k^* \\ -\log \frac{P_1(x_j)}{P_0(x_j)} + D_{KL}(P_1||P_0), & j \geq k^* \end{cases}$$

$$\ell(k) - \ell(k^*) = \begin{cases} \sum_{j=k}^{k^*-1} U_j - (k^* - k) D_{KL}(P_0||P_1), & k < k^* \\ \sum_{j=k^*}^{k-1} U_j - (k - k^*) D_{KL}(P_1||P_0), & k \geq k^* \end{cases}$$

We also define random variable S_m to denote the sum of m i.i.d. random variables as follows, noting that S_m is distributed like $\sum_{j=k^*+m}^{k^*-1} U_j$ for $m < 0$ and like $\sum_{j=k^*}^{k^*+m-1} U_j$ for $m > 0$.

$$S_m = \begin{cases} \sum_{k^*+m \leq j < k^*} U_j, & m < 0 \\ \sum_{k^* \leq j < k^*+m} U_j & m > 0 \end{cases}$$

With these random variables, we bound each term in the first set of terms in (21) for any $i \geq 1$ and threshold $t_i = 2^{i-2}\alpha C$ as follows:

$$
\begin{aligned}
&\Pr[\max_{k\in R_i}\{\ell(k) - \ell(k^*)\} > -2^{i-2}\alpha C] \\
&\quad\leq \Pr[\max_{k\in R_i^-}\{\sum_{j=k}^{k^*-1} U_j - (k^*-k)D_{KL}(P_0||P_1)\} > -2^{i-2}\alpha C] \\
&\qquad + \Pr[\max_{k\in R_i^+}\{\sum_{j=k^*}^{k-1} U_j - (k-k^*)D_{KL}(P_1||P_0)\} > -2^{i-2}\alpha C] \\
&\quad\leq \Pr[\max_{k\in[2^{i-1}\alpha]} |S_{-k}| > 2^{i-2}\alpha C] + \Pr[\max_{k\in[2^{i-1}\alpha]} |S_k| > 2^{i-2}\alpha C] \\
&\quad\leq \frac{4\cdot\exp(-2^{i-4}\alpha C^2/\Delta(\ell)^2)}{1-2\cdot\exp(-2^{i-4}\alpha C^2/\Delta(\ell)^2)} && (22)\\
&\quad\leq 8\exp(-2^{i-4}\alpha C^2/\Delta(\ell)^2) && (23)\\
&\quad= 8\left(\exp(\frac{-\alpha C^2}{8\Delta(\ell)^2})\right)^{2^{i-1}}
\end{aligned}
$$

where (22) follows from an application of Corollary 2 with $\lambda_1 = \lambda_2 = 2^{i-3}\alpha C$ and $L = \Delta(\ell)$, and the denominator can be simplified as in (23) under the assumption that $\alpha \geq \frac{8\Delta(\ell)^2 \log 4}{C^2}$, which is satisfied by our final bounds.

We now consider the sum of these terms over all i, which will be needed for the final bound on Equation (21). We note that this sum is bounded above by a geometric series with ratio $\exp(-\alpha C^2/(8\Delta(\ell)^2))$ since $2^{i-1} \geq i$, yielding the second and third inequalities. For the fourth inequality, the same assumed lower bound on α is used to simplify the denominator as in (23):

$$
\begin{aligned}
\sum_{i\geq 1}\Pr[\max_{k\in R_i}\{\ell(k)-\ell(k^*)\} > -2^{i-2}\alpha C] &\leq 8\sum_{i\geq 1}\left(\exp(\frac{-\alpha C^2}{8\Delta(\ell)^2})\right)^{2^{i-1}} \\
&\leq 8\sum_{i\geq 1}\left(\exp(\frac{-\alpha C^2}{8\Delta(\ell)^2})\right)^{i} \\
&\leq \frac{8\exp(\frac{-\alpha C^2}{8\Delta(\ell)^2})}{1-\exp(\frac{-\alpha C^2}{8\Delta(\ell)^2})} \\
&\leq \frac{32}{3}\exp\left(\frac{-\alpha C^2}{8\Delta(\ell)^2}\right) && (24)
\end{aligned}
$$

The first term in (20) in the theorem statement ensures that the expression above is bounded by $\beta/2$. It remains to show that the second term in (20) is enough to guarantee that the Laplace noise added for privacy will not harm accuracy except with probability $\beta/2$.

Next we bound the second set of terms of (21). We can easily bound one term in this set for any $i \geq 1$ since each Z_k and Z_{k^*} are independent draws from a Laplace distribution with parameter $\Delta(\ell)/\epsilon$, allowing us to apply a union bound over all indices in R_i:

$$
\begin{aligned}
\Pr[\max_{k\in R_i}\{Z_k - Z_{k^*}\} \geq 2^{i-2}\alpha C] &\leq \Pr[2\max_{k\in R_i}|Z_k| \geq 2^{i-2}\alpha C] \\
&\leq 2^i\alpha\Pr[|\mathrm{Lap}(\Delta(\ell)/\epsilon)| \geq 2^{i-3}\alpha C] \\
&\leq 2^i\alpha\cdot\exp(\frac{-2^{i-3}\alpha C\epsilon}{\Delta(\ell)}) \\
&= 2^i\alpha\left(\exp(\frac{-\alpha C\epsilon}{4\Delta(\ell)})\right)^{2^{i-1}}
\end{aligned}
$$

Then by summing over all ranges and assuming in (25) that $\alpha \geq \frac{4\Delta(\ell)\ln 2}{C\epsilon}$ to simplify the denominator (which will be satisfied by our final bound on α, we obtain a bound on the probability of large noise applied to any possible k far from k^*.

$$\begin{aligned}
\sum_{i\geq 1} \Pr[\max_{k\in R_i}\{Z_k - Z_{k^*}\} > 2^{i-2}\alpha C] &\leq \alpha \sum_{i\geq 1} 2^i \left(\exp(\frac{-\alpha C\epsilon}{4\Delta(\ell)})\right)^{2^{i-1}} \\
&\leq 2\alpha \sum_{i\geq 1} i \left(\exp(\frac{-\alpha C\epsilon}{4\Delta(\ell)})\right)^{i} \\
&= 2\alpha \frac{\exp(\frac{-\alpha C\epsilon}{4\Delta(\ell)})}{(1-\exp(\frac{-\alpha C\epsilon}{4\Delta(\ell)}))^2} \\
&\leq 8\alpha \exp\left(\frac{-\alpha C\epsilon}{4\Delta(\ell)}\right) \qquad (25)
\end{aligned}$$

Since $x/2 \geq \ln x$, requiring $\alpha \geq \frac{4\Delta(\ell)\log(16/\beta)}{C\epsilon}$ suffices to ensure that (25) is at most $\beta/2$ as required. □

3.3 Offline algorithm for arbitrary distributions

In this subsection, we give an offline private change-point detector OfflinePTCPD that offers guarantees even when $\Delta(\ell)$ is infinite. Relaxing the uniform bound assumption means that we may have a single data point x_j that dramatically increases $\ell(k)$ for $k \geq j$, so we cannot add noise proportional to $\Delta(\ell)$. Instead we truncate the log-likelihood ratio and add noise proportional to the post-truncation range. We compute the A-truncated log-likelihood ratio of k as

$$\ell_A(k) = \sum_{i=k}^{n} \left[\log \frac{P_1(x_i)}{P_0(x_i)}\right]_{-A/2}^{A/2},$$

where $[x]_a^b$ denotes the projection of x onto the interval $[a,b]$. This truncation scheme yields privacy immediately by instantiation of ReportMax [DR14] with queries $\ell_A(k)$ for $k \in [n]$, which have sensitivity $\Delta(\ell_A) = A$.

Algorithm 4 Offline private change-point detector: OfflinePTCPD$(X, P_0, P_1, \epsilon, n, A)$

```
Input: database $X$, distributions $P_0, P_1$, privacy parameter $\epsilon$, database size $n$, truncation parameter $A$
for $k = 1, \ldots, n$ do
    Compute $\ell_A(k) = \sum_{i=k}^{n} \left[\log \frac{P_1(x_i)}{P_0(x_i)}\right]_{-A/2}^{A/2}$
    Sample $Z_k \sim \mathrm{Lap}(\frac{A}{\epsilon})$
end for
Output $\tilde{k} = \operatorname{argmax}_{1\leq k\leq n}\{\ell_A(k) + Z_k\}$        ▷ Report noisy argmax
```

.

Theorem 7. *For arbitrary data X and $\epsilon > 0$,* OfflinePTCPD$(X, P_0, P_1, \epsilon, A)$ *is ϵ-differentially private.*

Since we are no longer able to uniformly bound $P_1(x)/P_0(x)$, our accuracy results include a truncation parameter A in place of $\Delta(\ell)$ since A is the sensitivity of ℓ_A. Rather than C, the distributional difference measure parametrizing our results correspondingly depends on the truncation parameter A, which must be chosen to ensure C_A is positive:

$$C_A = \min\left\{-\mathbb{E}_{x\leftarrow P_0}\left[\log\frac{P_1(x)}{P_0(x)}\right]_{-A/2}^{A/2}, \mathbb{E}_{x\leftarrow P_1}\left[\log\frac{P_1(x)}{P_0(x)}\right]_{-A/2}^{A/2}\right\}.$$

We note that for Gaussian and Gamma distributions, any $A > 0$ ensures $C_A > 0$. In Section 5 we illustrate that for these distributions, it is best to choose small A to avoid excess noise and effectively rely on the sign of the log-likelihood ratio for accuracy. For general $P_0 \neq P_1$, $A > 2$ is a sufficient condition by the following argument. When $A > 2$, we have $[\log x]_{-A/2}^{A/2} \leq x - 1$, and thus $\mathbb{E}_{P_0}\left[\log \frac{P_1(x)}{P_0(x)}\right]_{-A/2}^{A/2} < \mathbb{E}_{P_0}\left[\frac{P_1(x)}{P_0(x)} - 1\right] = 0$ and $\mathbb{E}_{P_1}\left[\log \frac{P_1(x)}{P_0(x)}\right]_{-A/2}^{A/2} = -\mathbb{E}_{P_1}\left[\log \frac{P_1(x)}{P_0(x)}\right]_{-A/2}^{A/2} > 0$.

With these definitions, we are ready to present the accuracy of OfflinePTCPD, in which the quantities A and $\mathbb{E}\left[\log \frac{P_1(x)}{P_0(x)}\right]_{-A/2}^{A/2}$ play roles analogous to $\ell(k)$ and D_{KL} in Theorem 6.

Theorem 8. *For arbitrary hypotheses P_0, P_1 and n data points X drawn from P_0, P_1 with true change time $k^* \in (1, n)$, for privacy parameter $\epsilon > 0$, and for truncation parameter A that satisfies $C_A > 0$,* OfflinePTCPD$(X, P_0, P_1, \epsilon, n, A)$ *is (α, β)-accurate for any $\beta > 0$ and*

$$\alpha = \max\left\{\frac{8A^2}{C_A^2}\log\frac{64}{3\beta}, \frac{4A}{C_A\epsilon}\log\frac{16}{\beta}\right\}, \tag{26}$$

where C_A is defined in (5).

Proof. Given some true change-point k^* and error tolerance $\alpha > 0$, we can partition the set of bad possible outputs k into sub-intervals of exponentially increasing size as follows. Following the notation of Theorem 6, for $i \geq 1$ we let

$$\begin{aligned} R_i^- &= [k^* - 2^i\alpha, k^* - 2^{i-1}\alpha) \\ R_i^+ &= (k^* + 2^{i-1}\alpha, k^* + 2^i\alpha] \\ R_i &= R_i^- \cup R_i^+, \end{aligned}$$

and for range-specific thresholds t_i for $i \geq 1$, we will bound the probability of a bad output as follows:

$$\Pr[|\tilde{k} - k^*| > \alpha] \leq \sum_{i \geq 1}\Pr[\max_{k \in R_i}\{\ell_A(k) - \ell_A(k^*)\} > t_i] + \sum_{i \geq 1}\Pr[\max_{k \in R_i}\{Z_k - Z_{k^*}\} \geq t_i] \tag{27}$$

In order to do this, we decompose the truncated log-likelihood difference between the true change-point k^* and any candidate k into the sum of i.i.d. random variables with mean zero and the expected value of this difference as follows:

$$U_j = \begin{cases} \left[\log \frac{P_1(x_j)}{P_0(x_j)}\right]_{-A/2}^{A/2} - \mathbb{E}_{x \leftarrow P_0}\left[\log \frac{P_1(x)}{P_0(x)}\right]_{-A/2}^{A/2}, & j < k^* \\ -\left[\log \frac{P_1(x_j)}{P_0(x_j)}\right]_{-A/2}^{A/2} + \mathbb{E}_{x \leftarrow P_1}\left[\log \frac{P_1(x)}{P_0(x)}\right]_{-A/2}^{A/2}, & j \geq k^* \end{cases}$$

$$\ell_A(k) - \ell_A(k^*) = \begin{cases} \sum_{j=k}^{k^*-1} U_j + (k^* - k)\mathbb{E}_{x \leftarrow P_0}\left[\log \frac{P_1(x)}{P_0(x_i)}\right]_{-A/2}^{A/2}, & k < k^* \\ \sum_{j=k^*}^{k-1} U_j - (k - k^*)\mathbb{E}_{x \leftarrow P_1}\left[\log \frac{P_1(x)}{P_0(x_i)}\right]_{-A/2}^{A/2}, & k \geq k^* \end{cases}$$

The rest of the proof follows exactly as the proof of the accuracy of OfflinePCPD from Theorem 6 with ℓ replaced with ℓ_A, with $\Delta(\ell)$ replaced with A, and with C replaced with C_A. As before, we set $t_i = 2^{i-2}\alpha C_A$, and the constants are inherited exactly as is because the truncated log-likelihood is applicable to the concentration inequalities in the same way that the non-truncated but uniformly bounded log-likelihood was in Theorem 6.

□

4 Online private change-point detection

In this section, we give new differentially private algorithms for change-point detection in the online setting. In this setting, the algorithm initially receives n data points $x_1, \ldots, x_n$ and then continues to receive data points one at a time. As before, the goal is to privately identify an approximation of the time k^* when the data change from distribution P_0 to P_1, and now we additionally want to identify this change shortly after it occurs. We first give an algorithm OnlinePCPD for detecting a single change-point, and then we show how it can be extended to OnlinePMCPD to detect multiple change-points. Our algorithms use OfflinePCPD as a subroutine, but can be modified in a straightforward way to use log-likelihood truncation and OfflinePTCPD if distributions do not satisfy the assumption of uniform boundedness.

4.1 Single change-point

Even in the single change-point setting, our offline algorithm is not directly applicable because we do not know a priori how many points must arrive before a true change-point occurs. To resolve this, OnlinePCPD works like AboveThresh(presented in Section 2.2), determining after each new data entry arrives whether it is likely that a change occurred in the most recent n entries. When OnlinePCPD at time j detects a sufficiently large (noisy) partial log-likelihood ratio $\ell(k,j) = \sum_{i=k}^{j} \log \frac{P_1(x_i)}{P_0(x_i)}$ for some k within n data points of j, it calls OfflinePCPD to privately determine the most likely change point $\tilde{k}$ in the window $\{x_{j-n+1}, \ldots, x_j\}$.

Privacy of OnlinePCPD is immediate from composition of AboveThresh and OfflinePCPD, each with privacy loss $\epsilon/2$. As before, accuracy requires X to be drawn from P_0, P_1 with some true change point k^*. This algorithm also requires a suitable choice of log-likelihood threshold T to guarantee that OfflinePCPD is called for a window of data that actually contains k^*. Specifically, T should be large enough that the algorithm is unlikely to call OfflinePCPD when $j < k^*$ but small enough so that it is likely to call OfflinePCPD by time $j = k^* + n/2$. When both of these conditions hold, we inherit the accuracy of OfflinePCPD.

With our final bounds, we note that $n \gg \frac{\Delta(\ell)}{C} \log(k^*/\beta)$ suffices for existence of a suitable threshold, and an analyst must have a reasonable approximation of k^* in order to choose such a threshold. Otherwise, the accuracy bound itself has no dependence on the change-point k^*.

Algorithm 5 Online private change-point detector: OnlinePCPD$(X, P_0, P_1, \epsilon, n, T)$

```
Input: database $X$, distributions $P_0, P_1$, privacy parameter $\epsilon$, starting size $n$, threshold $T$
Let $\Delta(\ell) = \max_x \log \frac{P_1(x)}{P_0(x)} - \min_{x'} \log \frac{P_1(x')}{P_0(x')}$
Let $\hat{T} = T + \mathrm{Lap}(4\Delta(\ell)/\epsilon)$
for each new data point $x_j, j \geq n$ do
    Compute $\ell_j = \max_{j-n+1 \leq k \leq j}\{\ell(k,j)\} = \max_{j-n+1 \leq k \leq j}\{\sum_{i=k}^{j} \log \frac{P_1(x_i)}{P_0(x_i)}\}$
    Sample $Z_j \sim \mathrm{Lap}(\frac{8\Delta(\ell)}{\epsilon})$
    if $\ell_j + Z_j > \hat{T}$ then
        Output $(j-n)$ + OfflinePCPD$(\{x_{j-n+1}, \ldots, x_j\}, P_0, P_1, \epsilon/2, n)$
        Halt
    end if
end for
```

Theorem 9. *For arbitrary data X and $\epsilon > 0$, OnlinePCPD$(X, P_0, P_1, \epsilon, n, T)$ is ϵ-differentially private.*

Theorem 10. *For hypotheses P_0, P_1 such that $\Delta(\ell) < \infty$, a stream of data points X with starting size n drawn from P_0, P_1 with true change time $k^* \geq n/2$, privacy parameter $\epsilon > 0$, and threshold $T \in [T_L, T_U]$ with*

$$\begin{aligned} T_L &:= 2\Delta(\ell)\sqrt{2\log\frac{64k^*}{\beta}} - C + \frac{16\Delta(\ell)}{\epsilon}\log\frac{8k^*}{\beta}, \\ T_U &:= \frac{nC}{2} - \frac{\Delta(\ell)}{2}\sqrt{n\log(8/\beta)} - \frac{16\Delta(\ell)}{\epsilon}\log\frac{8k^*}{\beta}, \end{aligned}$$

we have that OnlinePCPD$(X, P_0, P_1, \epsilon, n, T)$ is (α, β)-accurate for any $\beta > 0$ and

$$\alpha = \max\{\frac{8\Delta(\ell)^2}{C^2}\log\frac{128}{3\beta}, \frac{8\Delta(\ell)}{C\epsilon}\log\frac{32}{\beta}\}.$$

In the above expressions, $C = \min\{D_{KL}(P_0||P_1), D_{KL}(P_1||P_0)\}$.

Proof. We first give a range $[T_L, T_U]$ of thresholds that ensure that except with probability $\beta/4$, the randomly sampled data stream satisfies the following two conditions for some α'. These conditions are inherited from the requirements for AboveThreshaccuracy, respectively capturing the requirements that the threshold is not reached too early and that it is reached at least by the time the window is centered around k^*:

1. For $T \geq T_L$, $\max_{k\in[j-n+1,j]} \ell(k,j) < T - \alpha'$ for every $j < k^*$.
2. For $T \leq T_U$, $\max_{k\in[k^*-n/2,k^*+n/2-1]} \ell(k, k+n/2) > T + \alpha'$.

When these conditions are satisfied, the AboveThresh guarantee ensures that except with probability $\beta/4$, the randomness of the online algorithm ensures that it calls the offline algorithm on a window of data containing the true change-point. Then we will argue that our overall accuracy follows from the offline guarantee, where we will allow failure probability $\beta/2$.

We will get the first condition by taking a union bound over all windows tested before the change-point, of the probability that the maximum log-likelihood $\max_k \ell(k)$ for n elements $X = \{x_1, \ldots, x_n\}$ sampled from P_0 exceeds a given threshold. To bound this probability, we first define the following random variables.

$$U_j = -\log\frac{P_0(x_j)}{P_1(x_j)} + D_{KL}(P_0||P_1) \qquad\qquad S_m = \sum_{1\leq j\leq m} U_j$$

We note that each $\ell(k)$ is the sum of i.i.d. random variables, and that the maximum log-likelihood over m consecutive elements is equal in distribution to $\max_{k\in[m]} S_k - kD_{KL}(P_0||P_1)$. This yields the first inequality below. Inequality (28) comes from applying Corollary 2 with $\lambda_1 = \lambda_2 = 2^{i-2}C + t/2$ and interval length $L = \Delta(\ell)$.

$$\begin{aligned}
\Pr\left[\max_{1\leq k\leq n}\{\ell(k)\} > t\right] &\leq \sum_{i\geq 1}\Pr[\max_{k\in[2^{i-1},2^i)}\{S_k - kD_{KL}(P_0||P_1)\} > t] \\
&\leq \sum_{i\geq 1}\Pr[\max_{k\in[2^{i-1}]} S_k > 2^{i-1}C + t] \\
&\leq \sum_{i\geq 1}\frac{2\exp(-(2^{i-2}C + t/2)^2/(2^{i-2}\Delta(\ell)^2))}{1 - 2\exp(-(2^{i-2}C + t/2)^2/(2^{i-2}\Delta(\ell)^2))} && (28)\\
&\leq 4\sum_{i\geq 1}\exp(-(2^{i-2}C + t/2)^2/(2^{i-2}\Delta(\ell)^2)) && (29)\\
&\leq 8\exp(-(2^{-1}C + t/2)^2/(2^{-1}\Delta(\ell)^2)) && (30)\\
&\leq \frac{\beta}{8k^*} && (31)
\end{aligned}$$

Inequalities (29), (30), and (31) follow by plugging in $t = 2\Delta(\ell)\sqrt{2\log\frac{64k^*}{\beta}} - C$. This ensures that $1 - 2\exp(-(2^{i-2}C + t/2)^2/(2^{i-2}\Delta(\ell)^2)) \geq 1/2$, giving Inequality (29), and that the series is increasing exponentially in i, so we can collapse the sum with another factor of 2 by considering only $i = 1$ as in Inequality (30). Plugging in this same value of t to Inequality (30) also immediately gives Inequality (31). Taking a union bound over all the windows prior to the change-point, this shows that Condition 1 holds for $T_L = 2\Delta(\ell)\sqrt{2\log\frac{64k^*}{\beta}} - C + \alpha'$ except with probability $\beta/8$.

To show that the second condition holds except with additional probability $\beta/8$, we consider the window of data with the first half of data drawn from P_0 and the second half drawn from P_1 and bound the probability that $\ell(k^*)$ in this window is less than a given threshold as follows. We note that $\ell(k^*, k^* + n/2 - 1)$ is the sum of $n/2$ i.i.d. random variables $\log\frac{P_1(x_i)}{P_0(x_i)}$, although these variables are not mean-zero. Instead, we define mean-zero random variables $V_j = -\log\frac{P_1(x_j)}{P_0(x_j)} + D_{KL}(P_1||P_0)$, and write $\ell(k^*, k^* + n/2 - 1)$ in terms of these new variables, analogously to above. We can then bound the sum of the V_j using Hoeffding's inequality to get Equation (32):

$$\begin{aligned}\Pr[\max_{k^*-n/2\leq k<k^*+n/2}\{\ell(k,k^*+n/2-1)\}<t] &\leq \Pr[\ell(k^*,k^*+n/2-1)<t]\\ &\leq \Pr[\sum_{j=k^*,\ldots,k^*+n/2-1} V_j > nC/2-t]\\ &\leq \exp(-4(nC/2-t)^2/(n\Delta(\ell)^2)) \qquad (32)\end{aligned}$$

Plugging in $t=\frac{nC}{2}-\frac{\Delta(\ell)}{2}\sqrt{n\log(8/\beta)}$ in this final expression ensures that (32) $\leq \beta/8$. This ensures that Condition 2 is satisfied except with probability $\beta/8$ for $T_U = nC/2-\Delta(\ell)\sqrt{n\log(8/\beta)}-\alpha'$.

Then we can instantiate the AboveThresh accuracy guarantee with privacy parameter $\epsilon/2$ and accuracy parameter $\beta/4$ to ensure that for $\alpha' = \frac{16\Delta(\ell)\log(8k^*/\beta)}{\epsilon}$ when Conditions 1 and 2 are satisfied, AboveThresh will identify a window containing the true change-point except with probability $\beta/4$. Combining this with the $\beta/4$ probability that Conditions 1 and 2 fail to hold when $T\in[T_L,T_U]$, we get that OnlinePCPD calls OfflinePCPD in a window containing the change-point except with probability $\beta/2$ over the randomness of the data and of the online portion of the algorithm.

We next instantiate OfflinePCPD with appropriate parameters to ensure that conditioned on being called in the correct window, it will output a $\tilde{k}$ that is within α of the true change-point k^* with probability at most $\beta/2$. We can then complete the proof by taking a union bound over all the failure probabilities.

Our offline accuracy guarantee requires data points are sampled i.i.d. from P_0 before the change point and from P_1 thereafter. However, it remains to be shown that conditioning on the event that we call the offline algorithm in a correct window does not harm the accuracy guarantee too much. For a window size n, change-point k^*, stream X of at least $k^*+n/2$ data points, set of random coins required by OnlinePCPD and its call to OfflinePCPD, and a stopping index $\nu > n/2$, let $N(\nu)$ denote the event that OnlinePCPD calls OfflinePCPD on a window centered at ν, i.e., $\{x_{\nu-n/2},\ldots,x_{\nu+n/2-1}\}$, and let $F(\nu)$ denote the event that OfflinePCPD on the window centered at ν fails to output an approximation within α of k^*. Our previous arguments bound the probability of all $N(\nu)$ for ν outside of a good range $G=(k^*-n/2,k^*]$, and our offline guarantee bounds the probability of $F(\nu)$ for any $\nu\in G$ as long as the data are truly distributed according to the change-point model.

Failure of the online algorithm can be due to either failure to halt on a correct window or failure of the offline algorithm on a window containing the true change. Thus we can then bound the probability of failure of the online algorithm as:

$$\Pr[|\tilde{k}-k^*|>\alpha] \leq \sum_{\nu\notin G}\Pr[N(\nu)] + \Pr[\bigcup_{\nu\in G} F(\nu)]$$

The first summation is at most $\beta/2$ by our previous arguments on the accuracy of the online portion of the algorithm. It remains to calculate the second term. We can still partition the set of bad possible output into sub-intervals of exponentially increasing size as follows. For $i\geq 1$, let

$$\begin{aligned}R_i^- &= [k^*-2^i\alpha, k^*-2^{i-1}\alpha)\\ R_i^+ &= (k^*+2^{i-1}\alpha, k^*+2^i\alpha]\\ R_i &= R_i^-\cup R_i^+\end{aligned}$$

Then we can bound the probability that the offline algorithm fails on any correct window as:

$$\begin{aligned}
\Pr[\bigcup_{\nu \in G} F(\nu)] &\leq \Pr\left[\max_{\nu \in G}\left\{\max_{\substack{\nu-n/2 \leq k \leq \nu+n/2-1 \\ \text{s.t. } |k-k^*|>\alpha}} \{\ell(k, \nu+n/2-1) + Z_k - \ell(k^*, \nu+n/2-1) - Z_{k^*}\}\right\} > 0\right] \\
&= \Pr\left[\max_{\nu \in G}\left\{\max_{\substack{\nu-n/2 \leq k \leq \nu+n/2-1 \\ \text{s.t. } |k-k^*|>\alpha}} \sum_{j=k}^{\nu+n/2-1} \log\frac{P_1(x_j)}{P_0(x_j)} + Z_i - \sum_{j=k^*}^{\nu+n/2-1} \log\frac{P_1(x_j)}{P_0(x_j)} - Z_{k^*}\right\} > 0\right] \\
&= \Pr\left[\max_{\substack{k^*-n+1 \leq k \leq k^*+n/2-1 \\ \text{s.t. } |k-k^*|>\alpha}} \left\{\sum_{j=k}^{k^*} \log\frac{P_1(x_j)}{P_0(x_j)} + Z_k - Z_{k^*}\right\} > 0\right] \\
&\leq \sum_{i \geq 1} \Pr[\max_{k \in R_i}\{\sum_{j=k}^{k^*} \log\frac{P_1(x_j)}{P_0(x_j)}\} > -t_i] + \sum_{i \geq 1} \Pr[\max_{k \in R_i}\{Z_k - Z_{k^*}\} > t_i]
\end{aligned}$$

Notice that the final line above is identical to Equation (21) in the proof of Theorem 6 for the accuracy of OfflinePCPD: the first term is the empirical log-likelihood difference between the true change-point k^* and any candidate k, and the second term is difference between two independent draws of Laplace noise. Thus the remainder of the analysis follows that of Theorem 6 instantiated with parameters $\beta/2$ and $\epsilon/2$. This instantiation of Theorem 6 gives that $\Pr[\bigcup_{\nu \in G} F(\nu)]$ is also bounded by $\beta/2$ when $\alpha = \max\{\frac{8\Delta(\ell)^2}{C^2} \log \frac{128}{3\beta}, \frac{8\Delta(\ell)}{C\epsilon} \log \frac{32}{\beta}\}$.

Combining this with our previous bound on the $N(\nu)$ terms, we get that $\Pr[|\tilde{k} - k^*| > \alpha] \leq \beta$ for the desired α value in the theorem statement.

□

4.2 Multiple changes

We now show how to extend our OnlinePCPD algorithm to detect multiple change-points. In this setting, the data change from distribution P_0 to P_1, from P_1 to P_2, ..., and from P_{m-1} to P_m at times k_1^*, k_2^*, ..., k_m^*, respectively. As data arrive, OnlinePMCPD makes online determinations about when the current window is sufficiently likely to contain a change-point and calls OfflinePCPD when so. After each private report of a change-point $\tilde{k}_i$ the algorithm simply restarts the remaining stream of data points after the next n data points arrive and resumes scanning for subsequent change-points.

The idea of this algorithm is similar to the extension from AboveThresh to Sparse, but by assuming that the m change-points are separated pairwise by at least the starting database size n and by setting the thresholds to ensure that with high probability a changepoint k_i^* is detected by time $k_i^* + n/2$, we can update our sliding window between change-point detections to ensure that each entry only participates in one call to OnlinePCPD and we never miss a change-point. This means that privacy of OnlinePMCPD is immediate from privacy of OnlinePCPD and Sparse, and the accuracy cost is only $\log m$ rather than $\sqrt{m}$.

Algorithm 6 Online private multiple change-point detector: OnlinePMCPD$(X, P_0, \ldots, P_m, \epsilon, n, T_1, \ldots, T_m)$

```
Input: database $X$, distributions $P_0, \ldots, P_m$, privacy parameter $\epsilon$, starting size $n$, thresholds $T_1, \ldots, T_m$
Let $\Delta_1 = \max_x \log \frac{P_1(x)}{P_0(x)} - \min_{x'} \log \frac{P_1(x')}{P_0(x')}$
Let $\hat{T}_1 = T_1 + \mathrm{Lap}(4\Delta_1/\epsilon)$
Let $i = 1$
for each new data point $x_j, j \geq n$ do
    Compute $\ell_j = \max_{j-n+1 \leq k \leq j}\{\ell_i(k,j)\} = \max_{j-n+1 \leq k \leq j}\{\sum_{\iota=k}^{j} \log \frac{P_i(x_\iota)}{P_{i-1}(x_\iota)}\}$
    Sample $Z_j \sim \mathrm{Lap}(\frac{8\Delta_i}{\epsilon})$
    if $\ell_j + Z_j > \hat{T}_i$ then
        Output $\tilde{k}_i = (j-n) +$ OfflinePCPD$(\{x_{j-n+1}, \ldots, x_j\}, P_{i-1}, P_i, \epsilon/2, n)$
        if $i = m$ then
            Halt
        else
            Let $i = i + 1$
            Let $\Delta_i = \max_x \log \frac{P_i(x)}{P_{i-1}(x)} - \min_{x'} \log \frac{P_i(x')}{P_{i-1}(x')}$
            Let $\hat{T}_i = T_i + \mathrm{Lap}(4\Delta_i/\epsilon)$
            Wait for $n$ new data points, i.e., let $j$ advance by $n$
        end if
    end if
end for
```

Theorem 11. *For arbitrary data X and $\epsilon > 0$,* OnlinePMCPD*$(X, P_0, \ldots, P_m, \epsilon, n, T_1, \ldots, T_m)$ is ϵ-differentially private.*

It remains to prove accuracy for OnlinePMCPD. As before, accuracy requires X to be drawn from P_0, P_1, …, P_m with some true change-points k_1^*, k_2^*, …, k_m^*. To detect each change-point k_i^*, the choice of log-likelihood threshold T_i may need to be modified according to the hypothesized distributions and possibly to the expected time until the next change-point, which depends on the accuracy of the previous output.

Theorem 12. *For hypotheses $P_0, P_1, \ldots, P_m$ such that $\Delta_i = \max_x \log \frac{P_i(x)}{P_{i-1}(x)} - \min_{x'} \log \frac{P_i(x')}{P_{i-1}(x')} < \infty$ for $i = 1, \ldots, m$, a stream of data points X with starting size n drawn from $P_0, P_1, \ldots, P_m$ with true change times $k_0^*, k_1^*, \ldots, k_m^*$ with $k_0^* = 0, k_1^* \geq n/2, k_i^* - k_{i-1}^* \geq 3n/2$ for $i = 2, \ldots, m$, privacy parameter $\epsilon > 0$, and thresholds $T_i \in [T_{L,i}, T_{U,i}]$ with*

$$\begin{aligned} T_{L,i} &:= 2\Delta_i \sqrt{2 \log \frac{64m(k_i^* - k_{i-1}^*)}{\beta}} - C_i + \frac{16\Delta_i}{\epsilon} \log \frac{8m(k_i^* - k_{i-1}^*)}{\beta}, \\ T_{U,i} &:= \frac{nC_i}{2} - \frac{\Delta_i}{2}\sqrt{n \log(8m/\beta)} - \frac{16\Delta_i}{\epsilon} \log \frac{8m(k_i^* - k_{i-1}^*)}{\beta} \end{aligned}$$

for $i = 1, \ldots, m$, we have that OnlinePMCPD*$(X, P_0, \ldots, P_m, \epsilon, n, T_1, \ldots, T_m)$ is (α, β)-accurate for any $\beta > 0$ and*

$$\alpha = \max\left\{\frac{8\Delta^2}{C^2} \log \frac{128m}{3\beta}, \frac{8\Delta}{C\epsilon} \log \frac{32m}{\beta}\right\}.$$

In the above expressions, $\Delta = \max\{\Delta_1, \ldots, \Delta_m\}$, $C_i = \min\{D_{KL}(P_{i-1}||P_i), D_{KL}(P_i||P_{i-1})\}$ and $C = \min\{C_1, \ldots, C_m\}$.

Proof. For α as in the theorem statement, we will decompose the probability that the algorithm fails to output α-approximations for every k_i^* into the sum of m conditional probabilities, each of which can be bounded by β/m by an instantiation of our accuracy theorem for OnlinePCPD. In the proof below, we let S_i for $i \in [m]$ denote the event that OnlinePMCPD calls OnlinePCPD for the ith time with k_i^* in the latter half of the window and OnlinePCPD outputs an α-approximation of k_i^*. Then we have that

$$\Pr[\textsc{Online}\text{PMCPD}(X, P_0, \ldots, P_m, \epsilon, n, T_1, \ldots, T_m) \text{ fails}] \leq \Pr[\bar{S}_1] + \sum_{i=2}^{m} \Pr[\bar{S}_i \cap S_1 \cap \cdots \cap S_{i-1}]$$

$$\leq \Pr[\bar{S}_1] + \sum_{i=2}^{m} \Pr[\bar{S}_i \cap S_{i-1}] \leq \sum_{i \in [m]} \Pr[\textsc{Online}\text{PCPD}(X_i', P_{i-1}, P_i, \epsilon, n, T_i) \text{ fails}], \tag{33}$$

for X_i' drawn according to the single change-point model with initial distribution P_{i-1} and post-change distribution P_i with change-point $k_i^* - k_{i-1}^*$. The third inequality is because the event $\bar{S}_i$ conditioned on $S_1 \cap \cdots \cap S_{i-1}$ is equivalent to the failure of OnlinePCPD on a data stream consistent with the single change-point model, and in particular, failure is most likely when there are as many data points drawn from P_{i-1} as possible. Then bounding each term follows from instantiation of the theorem for OnlinePCPD because we can treat the ending point of the previous detection window as the starting point of a new detection procedure.

To be more mathematically rigorous, for $i = 2, \ldots, m$, we have

$$\Pr[\bar{S}_i \cap S_{i-1}] \leq \Pr[\bar{S}_i | S_{i-1}] = \mathbb{E}[\Pr(\bar{S}_i | S_{i-1}, j_{i-1}) | S_{i-1}] = \mathbb{E}[\Pr(\bar{S}_i | j_{i-1}) | S_{i-1}], \tag{34}$$

where the last equality follows from the fact that S_{i-1} and $\bar{S}_i$ are independent conditional on j_{i-1}. This conditional independence is an immediate consequence of the fact that S_{i-1} depends only on the data $x_1, \ldots, x_{j_{i-1}}$ and $\bar{S}_i$ depends only on the data $x_{j_{i-1}+1}, x_{j_{i-1}+2} \ldots$, which are mutually independent conditional on j_{i-1}, as j_{i-1} is a stopping time.

Our final goal is to bound $\Pr(\bar{S}_i | j_{i-1})$, which can be done by invoking Theorem 10. The only difference here is that the index of the first sample is $j_{i-1}+1$, instead of 1. Thus we have to modify the upper and lower thresholds in Theorem 10. Define $T_{L,i}(j) := 2\Delta_i \sqrt{2 \log \frac{64m(k_i^* - j)}{\beta}} - C_i + \frac{16\Delta_i}{\epsilon} \log \frac{8m(k_i^* - j)}{\beta}$, and $T_{U,i}(j) =: \frac{nC_i}{2} - \frac{\Delta_i}{2}\sqrt{n \log(8m/\beta)} - \frac{16\Delta_i}{\epsilon} \log \frac{8m(k_i^* - j)}{\beta}$. For $T_i \in [T_{L,i}, T_{U,i}]$ where $T_{L,i} = 2\Delta_i \sqrt{2 \log \frac{64m(k_i^* - k_{i-1}^*)}{\beta}} - C_i + \frac{16\Delta_i}{\epsilon} \log \frac{8m(k_i^* - k_{i-1}^*)}{\beta}$ and $T_{U,i} = \frac{nC_i}{2} - \frac{\Delta_i}{2}\sqrt{n \log(8m/\beta)} - \frac{16\Delta_i}{\epsilon} \log \frac{8m(k_i^* - k_{i-1}^*)}{\beta}$, we have $T_{L,i} \leq T_{L,i}(j) \leq T_i \leq T_{U,i}(j) \leq T_{U,i}(j)$ for any $j \in [k_{i-1}^*, k_{i-1}^* + n/2]$. Then by instantiation of Theorem 10, we have that $\Pr(\bar{S}_i | j_{i-1} = j) \leq \beta/m$ provided that $j \in [k_{i-1}^*, k_{i-1}^* + n/2]$. Note that the event S_{i-1} implies $j_{i-1} \in [k_{i-1}^*, k_{i-1}^* + n/2]$. Thus, $\Pr[\bar{S}_i \cap S_{i-1}]$ is bounded above by β/m, and (33) is bounded above by β. □

5 Numerical studies

In this section, we present results from Monte Carlo experiments designed to validate the theoretical results of previous sections. The theoretical privacy guarantees hold in the worst-case over all databases and over all outputs of the algorithm, so it is only necessary to empirically validate the accuracy of our algorithms. Our simulations consider both offline (Section 5.1) and online settings (Section 5.2) for the canonical problems of detecting a change in the mean of Bernoulli or Gaussian distributions. In the offline setting, we additionally show that our algorithms can accurately detect changes in the variance of Gaussian distribution and detect changes in the shape parameter of a Gamma distribution.

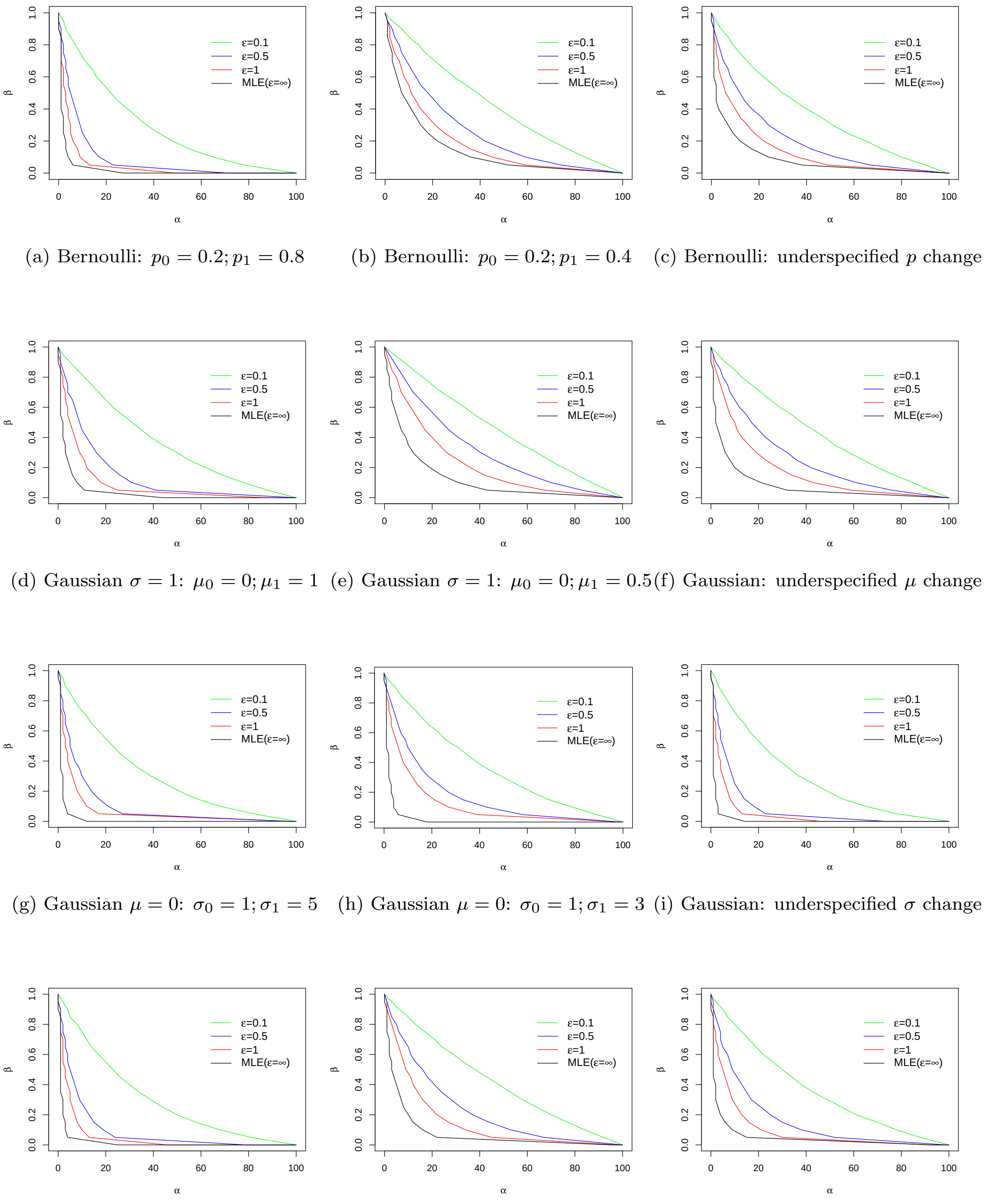
For completeness, we state the PMF of a Bernoulli distribution, and the PDF of Gaussian and Gamma distributions below.

- Bernoulli distribution: $\Pr(x = 1) = p$ and $\Pr(x = 0) = 1 - p$.
- Gaussian distribution: $f(x; \mu, \sigma) = (2\pi\sigma^2)^{1/2} \exp(-(x-\mu)^2/(2\sigma^2))$, where μ is the mean and σ is the standard deviation.
- Gamma distribution: $f(x; k, \theta) = (\Gamma(k)\theta^k)^{-1} x^{k-1} \exp(-x/\theta)$, where θ is the scale parameter and k is the shape parameter.

5.1 Evaluating the offline algorithms

Each simulation is characterized by a probability distribution family (Bernoulli, Gaussian, or Gamma), a distribution parameter that changes (mean, standard deviation, or shape), and a change magnitude (large, small, or underspecified). The large and small change regimes correspond respectively to large and small changes in the distribution parameter of interest. The underspecified regime corresponds to the setting where the true change is large, but the input parameters correspond to a small change. This setting goes beyond our theoretical results to suggest that our algorithm still performs well, even when the distributional parameters are misspecified. All parameters are stated in the caption.

For Bernoulli distributions, the log-likelihood ratio is uniformly bounded and so we use OfflinePCPD; for Gaussian and Gamma distributions, we set $A = 0.1$ (for reasons discussed later in this section) and use OfflinePTCPD. We vary privacy parameter $\epsilon = 0.1, 0.5, 1$ and ∞, representing the non-private case. For each of our simulations, we use $n = 200$ observations where the true change occurs at time $k^* = 100$. This process is repeated 10^4 times. The results of these simulations are presented in Figure 1, which plots the empirical probabilities $\beta = \Pr[|\tilde{k} - k^*| > \alpha]$ as a function of α. All the parameters of each simulation are stated in the caption.



(a) Bernoulli: $p_0 = 0.2; p_1 = 0.8$

(b) Bernoulli: $p_0 = 0.2; p_1 = 0.4$

(c) Bernoulli: underspecified p change

(d) Gaussian $\sigma = 1$: $\mu_0 = 0; \mu_1 = 1$

(e) Gaussian $\sigma = 1$: $\mu_0 = 0; \mu_1 = 0.5$

(f) Gaussian: underspecified μ change

(g) Gaussian $\mu = 0$: $\sigma_0 = 1; \sigma_1 = 5$

(h) Gaussian $\mu = 0$: $\sigma_0 = 1; \sigma_1 = 3$

(i) Gaussian: underspecified σ change

(j) Gamma $\theta = 2$: $k_0 = 3; k_1 = 1$

(k) Gamma $\theta = 2$: $k_0 = 3; k_1 = 2$

(l) Gamma: underspecified k change

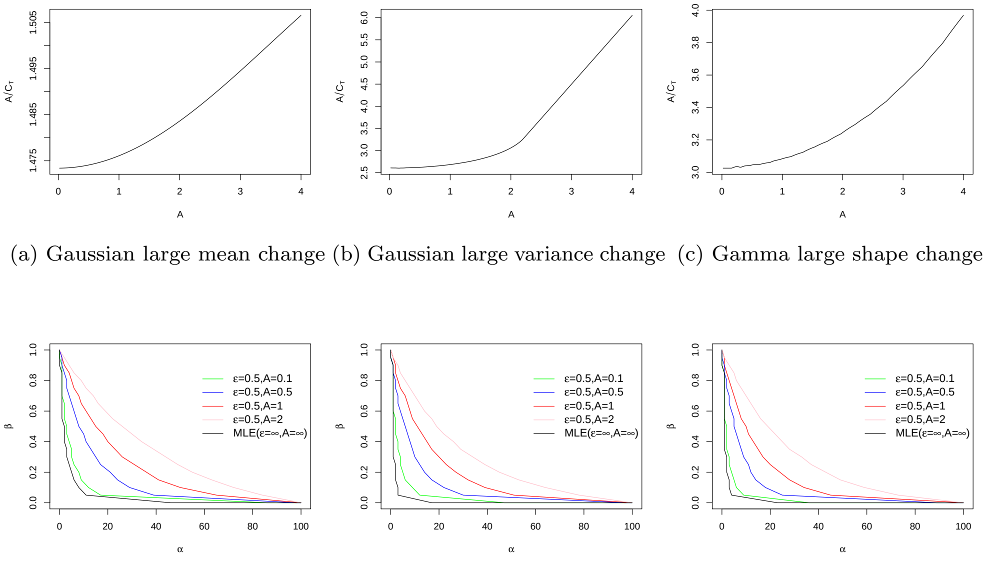
Figure 1: Measured accuracy of offline algorithms on simulated change-point data. For large and small changes (Columns 1 and 2, resp.), parameters specify distributions from which data are drawn and hypothesized distributions given as inputs to the algorithm; for underspecified changes (Column 3), data are drawn according to large change values but algorithm is provided hypothesized distributions consistent with small change values.

Figure 1 illustrates three important results for our offline algorithms when data are drawn from Bernoulli, Gaussian, or Gamma distributions: accuracy deteriorates as privacy improves but performs quite well even for strong privacy guarantees ($\epsilon < 1$), accuracy is best when the true change in distribution is large (Columns 1 vs 2), and the algorithm performs well even when the true change is larger than that hypothesized (Column 3). The performance in the underspecified change experiments bolster our theoretical results substantially, indicating that our hypotheses can be quite far from the distributions of the true data and our algorithms will still identify a change-point relatively accurately.

Choice of truncation parameter A. The OFFLINEPCPD algorithm does not provide meaningful results when the sensitivity of the log-likelihood ratio is infinite, as in the case of Gaussian and Gamma distributions, so we must instead use OFFLINEPTCPD with some truncation parameter A. Theorem 8 shows that accuracy guarantees are strongest when A/C_A is smallest. Since C_A is a function of the hypothesized distributions as well as A, the value of A should be chosen on a case-by-case basis.

The first row of Figure 2 numerically plots A against A/C_A for the large change cases we simulated. The plots suggest that a small A also leads to a small A/C_A, and A/C_A converges to a constant as A goes to 0. The second row verifies optimality of small A by simulation, plotting the empirical probabilities β as a function of accuracy α under different choices of A.

Intuitively, since the mechanism outputs $\text{argmax}_{k\in[n]}\{\sum_{i=k}^{n}\left[\log\frac{P_1(x_i)}{P_0(x_i)}\right]_{-A/2}^{A/2} + \text{Lap}(A/\epsilon)\}$, there is a trade-off between how much information is lost from truncation in the first term and how much noise is added in the second term. As $A \to 0^+$, each data point contributes $\pm A/2$. For natural distributions, it appears that giving some data points more weight than others does not provide enough additional information to offset the additional required noise.



(a) Gaussian large mean change (b) Gaussian large variance change (c) Gamma large shape change

(d) Gaussian large mean change (e) Gaussian large variance change (f) Gamma large shape change

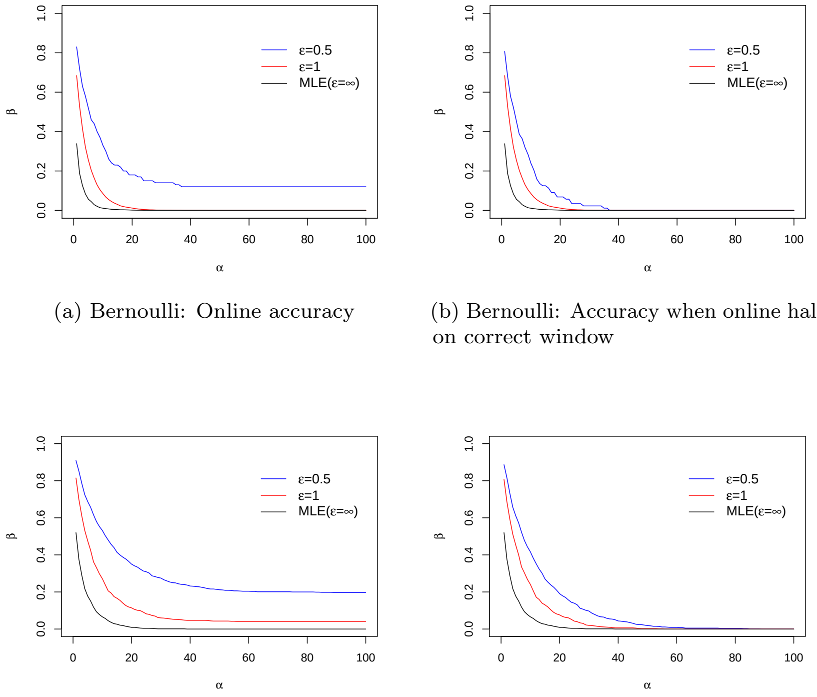
Figure 2: First row plots A/C_A as a function of A varying from 0 to 4 for different types of changes; theoretical accuracy bounds are strongest when A/C_A is smallest. Second row shows simulated accuracy under different choices of A for different types of change. Each simulation involves 10^4 runs of OFFLINEPTCPD on data generated by 200 i.i.d. samples from appropriate distributions with change-point $k^* = 100$.

5.2 Evaluating the online algorithm

We also run Monte Carlo simulations of our online change-point detection algorithm OnlinePCPD when the data points arrive sequentially and the true change occurs at time $k^* = 5000$. We consider only large mean changes in Bernoulli and Gaussian distributions. For the Gaussian distributions, we truncate the log-likelihoods in the main algorithm and call OfflinePTCPD with $A = 0.1$. The new challenge is to choose an appropriate sliding window size n and corresponding threshold T in order to achieve good overall accuracy. The window size of $n = 200$ used in the offline simulations does not permit any threshold that reasonably controls both false positive and false negative rates, so we choose a larger window size of $n = 700$ and restrict our online simulations to $\epsilon = 0.5, 1, \infty$. 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.

For the online simulations, we chose the lower and upper bounds of T via numerical methods in both Bernoulli and Gaussian models instead of using the theoretical bounds, as these bounds are overly conservative for the Bernoulli model and do not immediately apply for truncation method that is necessary in Gaussian model. We use several key ideas from Section 4 to speed up the numerical search of the threshold T. To limit the false positive rate to 0.10 with up to $k^* = 5000$ sliding windows, a conservative lower bound for threshold T is the $1 - 0.10/5000 = 0.99998$ quantile of the noisy versions of $W_n = \max_{1 \le k \le n} \ell(k)$ or $W_n = \max_{1 \le k \le n} \ell_A(k)$ with $n = 700$ under the *pre-change* distribution. To limit the false negative rate, an upper bound for threshold T is the 10% quantile of the noisy versions of CUSUM statistics W_n with $n = 700$ when the change occurs at time 350. This will guarantee that the online algorithms raise an alarm with probability at least 0.9 during the time interval $[4650, 5350]$.

To determine these lower and upper bounds for T, we simulate 10^6 realizations of the CUSUM statistics W_{700} in both the pre-change and post-change cases. In each case, we speed up the computation of W_i by using the recursive form $W_i = \max\{W_{i-1}, 0\} + \log(P_1(X_i)/P_0(X_i))$ or $W_i = \max\{W_{i-1}, 0\} + [\log(P_1(X_i)/P_0(X_i))]_{A/2}^{A/2}$ for $i \ge 1$. The empirical quantiles of the noisy versions of W_{700} under the pre- and post- change cases will yield the lower and upper bounds of the threshold T. When the range of acceptable thresholds T was non-empty, we chose the upper bound. For the Bernoulli model, this resulted in a choice of $T = 220$ for all values of $\epsilon = 0.5, 1, \infty$. In the Gaussian model, we chose $T = 8, 4.5, 100$ for $\epsilon = 0.5, 1, \infty$, respectively. Figure 3 (a and c) indeed show that with these parameters, the algorithm works well except with probability about 0.2, and comparison with plots b and d, we can see that almost all of the error for reasonable values of α is due to failure to abort on a window containing the true change-point. This indicates that the primary challenge in the online setting is determining when to raise an alarm in a sequence of sliding windows of observations. Once such window is identified correctly, the offline estimation algorithm can be used to accurately estimate the change-point.



(a) Bernoulli: Online accuracy

(b) Bernoulli: Accuracy when online halts on correct window

(c) Gaussian: Online accuracy

(d) Gaussian: Accuracy when online halts on correct window

Figure 3: Probability that the online algorithm produces an inaccurate estimate (left) and probability that the online algorithm produces an inaccurate estimate conditioned on halting in a window containing k^* (right) for Bernoulli and Gaussian large mean changes. Each simulation involves 10^6 runs of OnlinePCPD or its 0.1-truncated variant with window size $n = 700$ and varying ϵ on data generated by i.i.d. samples from appropriate distributions with change point $k^* = 5000$. See text for description of choices of threshold T.

References

[BP03] J. Bai and P. Perron. Computation and analysis of multiple structural change models. *Journal of Applied Econometrics*, 18(1):1–22, 2003.

[Car88] E. Carlstein. Nonparametric change-point estimation. *The Annals of Statistics*, 16(1):188–197, 1988.

[Cha17] H. P. Chan. Optimal sequential detection in multi-stream data. *The Annals of Statistics*, 45(6):2736–2763, 2017.

[CKM+19] Clément L Canonne, Gautam Kamath, Audra McMillan, Adam Smith, and Jonathan Ullman. The structure of optimal private tests for simple hypotheses. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, STOC '19, pages 310–321, 2019.

[DMNS06] C. Dwork, F. McSherry, K. Nissim, and A. Smith. Calibrating noise to sensitivity in private data analysis. In *Proceedings of the 3rd Conference on Theory of Cryptography*, TCC '06, pages 265–284, 2006.

[DNPR10] Cynthia Dwork, Moni Naor, Toniann Pitassi, and Guy N. Rothblum. Differential privacy under continual observation. In *Proceedings of the 42nd ACM Symposium on Theory of Computing*, STOC '10, pages 715–724, 2010.

[DR14] Cynthia Dwork and Aaron Roth. The algorithmic foundations of differential privacy. *Foundations and Trends in Theoretical Computer Science*, 9(3–4):211–407, 2014.

[Hin70] D. V. Hinkley. Inference about the change-point in a sequence of random variables. *Biometrika*, 57(1):1–17, 1970.

[Kul01] M. Kulldorff. Prospective time periodic geographical disease surveillance using a scan statistic. *Journal of the Royal Statistical Society, Series A*, 164(1):61–72, 2001.

[Lai95] T. L. Lai. Sequential changepoint detection in quality control and dynamical systems. *Journal of the Royal Statistical Society, Series B*, 57(4):613–658, 1995.

[Lai01] T. L. Lai. Sequential analysis: some classical problems and new challenges. *Statistica Sinica*, 11(2):303–408, 2001.

[Lor71] G. Lorden. Procedures for reacting to a change in distribution. *The Annals of Mathematical Statistics*, 42(6):1897–1908, 1971.

[LR02] R. Lund and J. Reeves. Detection of undocumented changepoints: A revision of the two-phase regression model. *Journal of Climate*, 15(17):2547–2554, 2002.

[McS09] Frank McSherry. Privacy integrated queries: an extensible platform for privacy-preserving data analysis. In *Proceedings of the 2009 ACM SIGMOD International Conference on Management of Data*, SIGMOD '09, pages 19–30, 2009.

[Mei06] Y. Mei. Sequential change-point detection when unknown parameters are present in the pre-change distribution. *The Annals of Statistics*, 34(1):92–122, 2006.

[Mei08] Y. Mei. Is average run length to false alarm always an informative criterion? *Sequential Analysis*, 27(4):354–419, 2008.

[Mei10] Y. Mei. Efficient scalable schemes for monitoring a large number of data streams. *Biometrika*, 97(2):419–433, 2010.

[Mou86] G. V. Moustakides. Optimal stopping times for detecting changes in distributions. *The Annals of Statistics*, 14(4):1379–1387, 1986.

[Pag54] E. S. Page. Continuous inspection schemes. *Biometrika*, 41(1/2):100–115, 1954.

[Pol85] M. Pollak. Optimal detection of a change in distribution. *The Annals of Statistics*, 13(1):206–227, 1985.

[Pol87] M. Pollak. Average run lengths of an optimal method of detecting a change in distribution. *The Annals of Statistics*, 15(2):749–779, 1987.

[Rob66] S. W. Roberts. A comparison of some control chart procedures. *Technometrics*, 8(3):411–430, 1966.

[She31] W. A. Shewhart. *Economic Control of Quality of Manufactured Product.* D. Van Norstrand Company, Inc., 1931.

[Shi63] A. N. Shiryaev. On optimum methods in quickest detection problems. *Theory of Probability & Its Applications*, 8(1):22–46, 1963.

[VDVW96] A. W. Van Der Vaart and J. A. Wellner. *Weak convergence.* Springer, 1996.

[ZS12] N. Zhang and D. O. Siegmund. Model selection for high-dimensional, multi-sequence change-point problems. *Statistica Sinica*, 22(4):1507–1538, 2012.