A. Preliminaries

Theorem 6 (Basic Composition (Dwork et al., 2006))**.** *Let $\mathcal{M}_1$ be an ϵ_1-differentially private algorithm, and let $\mathcal{M}_2$ be an ϵ_2-differentially private algorithm. Then their composition $(\mathcal{M}_1, \mathcal{M}_2)$ is $(\epsilon_1 + \epsilon_2)$-differentially private.*

Algorithm 3 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 Δ, privacy parameter ϵ
for $i = 1, \ldots, m$ **do**
  Compute $f_i(X)$
  Sample $Z_i \sim \text{Lap}(\frac{\Delta}{\epsilon})$
end for
Output $i^* = \underset{i \in [m]}{\text{argmax}} \, (f_i(X) + Z_i)$

Theorem 7 ((Dwork & Roth, 2014))**.** REPORTMAX *is ϵ-differentially private.*

Algorithm 4 Above Noisy Threshold: ABOVETHRESHOLD$(X, \Delta, \{f_1, f_2, \ldots\}, T, \epsilon)$

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

Theorem 8 ((Dwork et al., 2009))**.** ABOVETHRESHOLD *is ϵ-differentially private.*

Theorem 9 ((Dwork et al., 2009))**.** *For any sequence of m queries $f_1, \ldots, f_m$ with sensitivity Δ such that $|\{i < m : f_i(X) \geq T - \alpha\}| = 0$,* ABOVETHRESHOLD *outputs with probability at least $1 - \beta$ a stream of $a_1, \ldots, a_m \in \{\top, \perp\}$ such that $a_i = \perp$ 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*

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

Our proofs use the following concentration inequality.

Theorem 10 (McDiarmid (McDiarmid, 1989))**.** *Define the discrete derivatives of the function $f(X_1, \ldots, X_n)$ of independent random variables $X_1, \ldots, X_n$ as*

$$\begin{aligned} D_i f(x) := &\sup_z f(x_1, \ldots, x_{i-1}, z, x_{i+1}, \ldots, x_n) \\ &- \inf_z f(x_1, \ldots, x_{i-1}, z, x_{i+1}, \ldots, x_n). \end{aligned} \tag{2}$$

Then for $X_1, \ldots, X_n$ independent, $f(X_1, \ldots, X_n)$ is subgaussian with variance proxy $\frac{1}{4}\sum_{i=1}^n ||D_i f||_\infty^2$, and

$$\begin{aligned} &\Pr[f(X_1, \ldots, X_n) - \mathbb{E}[f(X_1, \ldots, X_n)] \geq t] \\ &\qquad \leq \exp\left(-\frac{2t^2}{\sum_{i=1}^n ||D_i f||_\infty^2}\right). \end{aligned}$$

B. 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 detect the drift change-point t^* with the smallest possible error.

In order to apply our algorithms which 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 ease of presentation, we will 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 we have

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

Note that random variables $(e_{2t} - e_{2t-1})$ are independent and identically distributed. Thus the y_t are independent, and they are sampled from a fixed distribution before the change point, and from another distribution after the change-point. We can then 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 will inherit the privacy and accuracy results 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 11. *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*

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

for any constant $c > 1$ and some constant $C_1, C_2 > 0$ depending 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 $f : \mathbb{R} \to \mathbb{R}$,

[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.

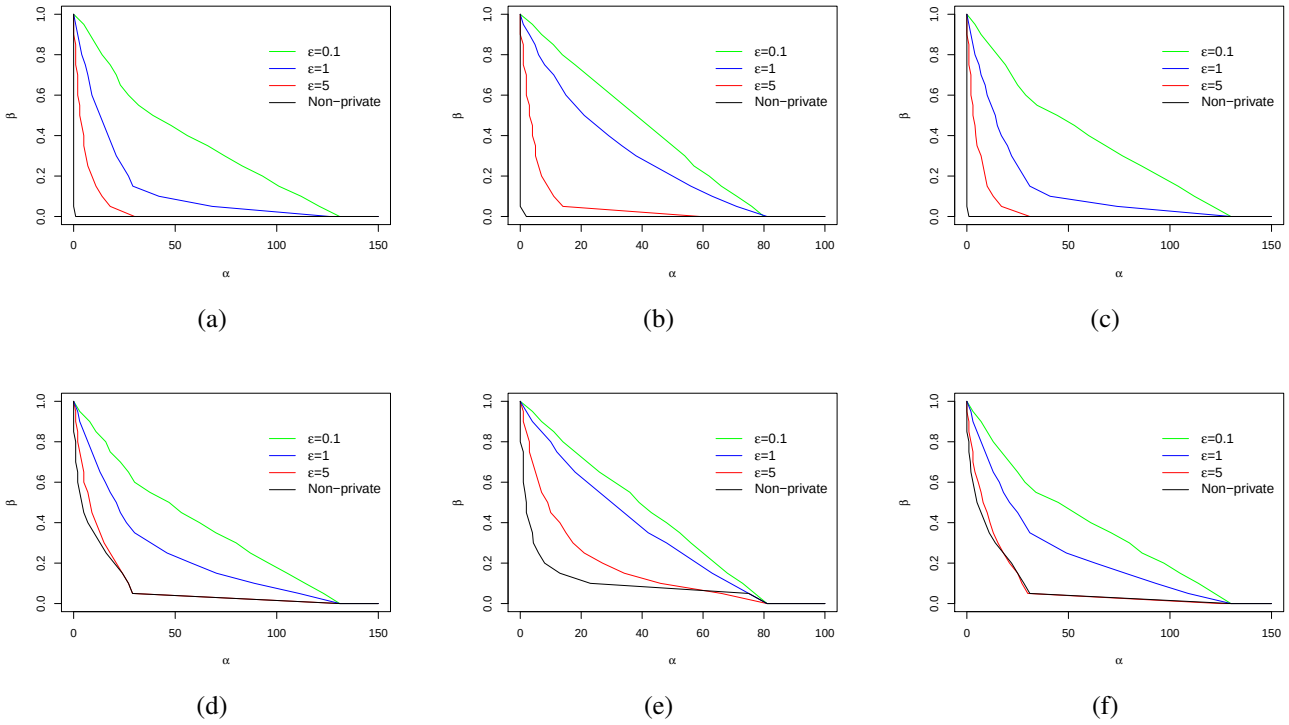


Figure 3. Empirical accuracy $\beta = \Pr[|\tilde{k} - k^*| > \alpha]$ of PNCPD from Monte Carlo simulations using Gaussian data, where pre-change data are drawn from $\mathcal{N}(0, 1)$ and post-change data are drawn from $\mathcal{N}(\mu_1, 1)$. Each simulation involves 10^3 runs of PNCPD with varying ϵ on data generated by 200 i.i.d. samples from appropriate distributions: (a) $k^* = 50$, $\mu_1 = 5$; (b) $k^* = 100$, $\mu_1 = 5$; (c) $k^* = 150$, $\mu_1 = 5$; (d) $k^* = 50$, $\mu_1 = 1$; (e) $k^* = 100$, $\mu_1 = 1$; (f) $k^* = 150$, $\mu_1 = 1$

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).

C. Empirical results

Recall that our drift change detection model involved data points $X = \{x_1, \ldots, x_n\}$ defined as $x_t = \mu_t + e_t$ where

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

for drift change-point t^*, and e_t are mean-zero noise terms. In our simulation we use 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 $Y = \{y_1, \ldots, y_{n/2}\}$ as described in Section B, and apply our PNCPD algorithm to this new sample. Figure 4 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.

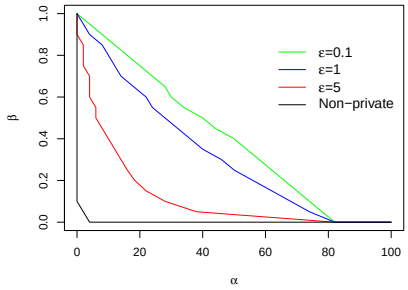
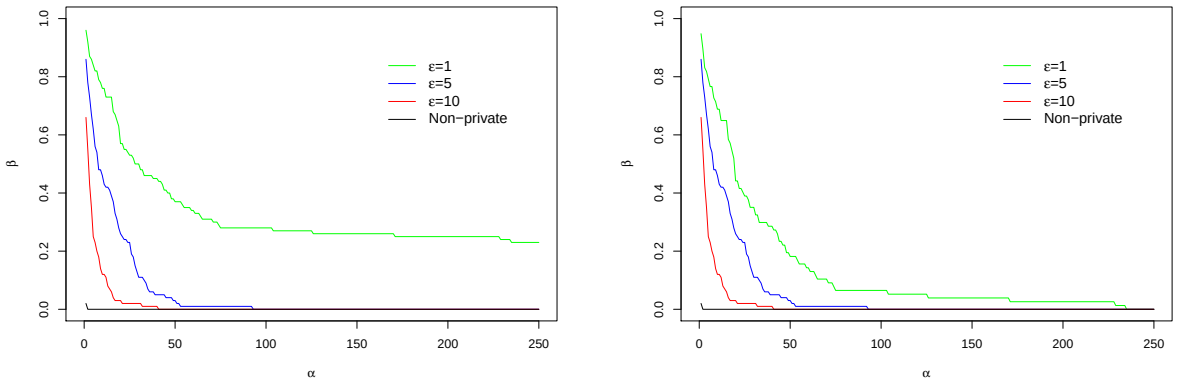


Figure 4. Empirical accuracy $\beta = \Pr[|\tilde{t} - t^*| > \alpha]$ of PNCPD for drift detection. The data are generated from the drift change model with parameters $\eta = 1$, $\xi_0 = 0$, $\xi_1 = 5$, and e_t drawn from $\mathcal{N}(0, 1)$. These data are then modified as described in Section B so that the PNCPD algorithm can be applied.



(a) (b)

Figure 5. Probability of inaccurate estimation and false alarm (a) and probability of inaccurate report conditioned on raising an alarm correctly (b) for Monte Carlo simulations. Data drawn from $\mathcal{N}(5, 1)$ pre-change and $\mathcal{N}(0, 1)$ post-change, with true change-point $k^* = 5000$. Each simulation involves 10^3 runs of ONLINEPNCPD with $\gamma = 0.1$, window size $n = 500$, threshold $T = 0.8$, and varying ϵ.

D. Technical Proofs

D.1. Proof of Theorem 1

Proof. We will show that for $\hat{k} = \operatorname{argmax} V(k)$ and α as in the theorem statement,

$$\Pr[\left|\hat{k} - k^*\right| > \alpha] \leq \sum_{k:|k-k^*|>\alpha} \Pr[V(k) > V(k^*)] \leq \beta.$$

To do this, we fix any $k \in \{\lceil \gamma n \rceil, \ldots, \lfloor (1-\gamma)n \rfloor\}$ and show that $f(X) = V(k) - V(k^*)$ is subgaussian. In particular, for k at least α away from k^*, the expectation of $V(k^*) - V(k)$ is sufficiently large and its discrete derivative is sufficiently small that the probability of $V(k) > V(k^*)$ can be tightly bounded as a function of α by application of Theorem 10.

First we give a lower bound the difference in expectation of $V(k^*)$ and $V(k)$. Observe that

$$\mathbb{E}[V(k)] = \frac{\sum_{i \leq k, j > k} \Pr[x_i > x_j]}{k(n-k)}$$

$$= \begin{cases} \frac{\frac{1}{2}(k^*-k)+a(n-k^*)}{n-k} & k \le k^* \\ \frac{ak^*+\frac{1}{2}(k-k^*)}{k} & k > k^* \end{cases},$$

achieving its maximum at $\mathbb{E}[V(k^*)] = a$. Therefore, we can bound

$$\mathbb{E}[V(k^*) - V(k)] = \begin{cases} (a-\frac{1}{2})\frac{k^*-k}{n-k} & k \le k^* \\ (a-\frac{1}{2})\frac{k-k^*}{k} & k > k^* \end{cases}$$
$$\ge (a-\frac{1}{2})\frac{|k^*-k|}{n}. \tag{3}$$

In the following bounds on the discrete derivative of $f(X) = V(k) - V(k^*)$, we will make use of the fact that f can be written as:

$$f(X) = \frac{\sum_{j=k+1}^{n}\sum_{i=1}^{k} I(x_i > x_j)}{k(n-k)} - \frac{\sum_{j=k^*+1}^{n}\sum_{i=1}^{k^*} I(x_i > x_j)}{k^*(n-k^*)}$$
$$= \left(\frac{1}{k(n-k)} - \frac{1}{k^*(n-k^*)}\right) \sum_{\substack{i\in[1,k] \\ j\in[k+1,n]}} I(x_i > x_j) + \frac{1}{k^*(n-k^*)} \cdot \left(\sum_{\substack{i\in[1,k] \\ j\in[k+1,n]}} I(x_i > x_j) - \sum_{\substack{i\in[1,k^*] \\ j\in[k^*+1,n]}} I(x_i > x_j) \right)$$

We bound the discrete derivative $D_i f$ separately for $i \le \min\{k, k^*\}$, $i \in (\min\{k, k^*\}, \max\{k, k^*\}]$, and $i > \max\{k, k^*\}$. When x_i changes arbitrarily for $i \le \min\{k, k^*\}$, we note that $\sum_{j=k+1}^{n} I(x_i > x_j)$ can change by at most $\pm(n-k)$ and $\sum_{j=k+1}^{k^*} I(x_i > x_j)$ can change by at most $\pm(k^*-k)$. These counts are normalized in f, and the normalization ensures this former count contributes at most $\frac{|k^*-k|}{k^*(n-k^*)} + \frac{|k^*-k|}{kk^*}$ to the discrete derivative. We bound the discrete derivative for $i \le \min\{k, k^*\}$ as follows:

$$\begin{aligned} D_i f &\le \left|\frac{1}{k(n-k)} - \frac{1}{k^*(n-k^*)}\right|(n-k) + \frac{|k^*-k|}{k^*(n-k^*)} \\ &= \left|\frac{1}{k} - \frac{n-k}{k^*(n-k^*)}\right| + \frac{|k^*-k|}{k^*(n-k^*)} \\ &= \left|-\frac{|k-k^*|}{k^*k} + \frac{|k-k^*|}{k^*(n-k^*)}\right| + \frac{|k-k^*|}{k^*(n-k^*)} \\ &\le \frac{|k-k^*|}{\gamma^2 n^2} + \frac{2\,|k-k^*|}{\gamma(1-\gamma)n^2} \\ &\le \frac{3\,|k-k^*|}{\gamma^2 n^2} \end{aligned}$$

We bound the discrete derivative for $i > \max\{k, k^*\}$ similarly, noting that an arbitrary change in x_i changes $\sum_{i'=1}^{k} I(x_{i'} > x_i)$ by at most $\pm k$ and $\sum_{i'=k^*+1}^{k} I(x_{i'} > x_i)$ by at most $\pm(k-k^*)$:

$$\begin{aligned} D_i f &\le \left|\frac{1}{k(n-k)} - \frac{1}{k^*(n-k^*)}\right| \cdot k + \frac{|k^*-k|}{k^*(n-k^*)} \\ &= \left|\frac{1}{n-k} - \frac{k}{k^*(n-k^*)}\right| + \frac{|k^*-k|}{k^*(n-k^*)} \\ &= \left|-\frac{|k-k^*|}{(n-k^*)(n-k)} + \frac{|k-k^*|}{k^*(n-k^*)}\right| + \frac{|k-k^*|}{k^*(n-k^*)} \\ &\le \frac{|k-k^*|}{\gamma^2 n^2} + \frac{2\,|k-k^*|}{\gamma(1-\gamma)n^2} \\ &\le \frac{3\,|k-k^*|}{\gamma^2 n^2} \end{aligned}$$

Finally we bound the discrete derivative for $\min\{k, k^*\} < i \leq \max\{k, k^*\}$. To do this, we note that the first summation in f changes by k if $k < k^*$ or $n - k$ if $k > k^*$, and the difference of summations in the second term changes by at most $n - (k + k^*)$ in either case. Then we achieve our bound as follows:

$$\begin{aligned} D_i f &\leq \left| \frac{1}{k(n-k)} - \frac{1}{k^*(n-k^*)} \right| \cdot \max\{k, n-k\} + \frac{n - (k^* + k)}{k^*(n-k^*)} \\ &\leq \frac{|k - k^*|}{\gamma^2 n^2} + \frac{n}{\gamma(1-\gamma)n^2} \\ &\leq \frac{2}{\gamma^2 n} \end{aligned}$$

Then since $D_i f$ is finite for each i, we have that f is subgaussian with variance proxy as follows:

$$\begin{aligned} \frac{1}{4}\sum_{i=1}^{n}(D_i f)^2 &\leq \frac{n - |k^* - k|}{4} \cdot \frac{9\,|k - k^*|^2}{\gamma^4 n^4} + \frac{|k^* - k|}{4}\left(\frac{|k - k^*|}{\gamma^2 n^2} + \frac{1}{\gamma(1-\gamma)n}\right)^2 \\ &\leq \frac{9\,|k - k^*|^2}{4\gamma^4 n^3} + \frac{|k^* - k|}{\gamma^4 n^2} \\ &\leq \frac{13\,|k^* - k|}{4\gamma^4 n^2} \end{aligned}$$

We can now bound the probability of outputting any particular $k = \lceil \gamma n \rceil, \ldots, \lfloor (1-\gamma)n \rfloor$ as a function of $|k - k^*|$ by applying Theorem 10, recalling our bound on $\mathbb{E}[V(k^*) - V(k)]$ from Equation (3).

$$\begin{aligned} \Pr[V(k) > V(k^*)] &= \Pr\left[V(k) - V(k^*) - \mathbb{E}[V(k) - V(k^*) > \mathbb{E}[V(k^*) - V(k)]\right] \\ &\leq \Pr\left[V(k) - V(k^*) - \mathbb{E}[V(k) - V(k^*)] > (a - \frac{1}{2})\frac{|k - k^*|}{n}\right] \\ &\leq \exp\left(-\frac{2\gamma^4}{13}(a - \frac{1}{2})^2 |k - k^*|\right). \end{aligned}$$

We complete the proof by bounding the probability of any incorrect $\hat{k}$ such that $|\hat{k} - k^*| > \alpha$ by β.

$$\begin{aligned} \Pr[\left|\hat{k} - k^*\right| > \alpha] &\leq 2 \sum_{|k-k^*|=\alpha}^{n} \exp(-\frac{2\gamma^4}{13}(a - \frac{1}{2})^2\,|k - k^*|) \\ &\leq \frac{2\exp(-\frac{2\gamma^4}{13}(a - \frac{1}{2})^2\alpha)}{1 - \exp(-\frac{2\gamma^4}{13}(a - \frac{1}{2})^2)} \\ &\leq \beta \end{aligned}$$

Rearranging shows that our accuracy result will hold for

$$\alpha \geq \frac{13}{2\gamma^4(a - 1/2)^2}\left(\log\frac{2}{\beta} + \log\frac{1}{1 - \exp(-\frac{2\gamma^4}{13}(a - \frac{1}{2})^2)}\right).$$

We achieve our final bound by simplifying the above expression as follows. We observe that $\gamma < 1/2, a < 1$ implies $x = 2\gamma^4(a - 1/2)^2/13 \leq 1/416$, and for small x we have $\log(1/(1 - \exp(-x))) \leq 2\log(1/x)$. For any $c > 0$, we have $\log(1/x) \leq C(1/x)^c$ for any $1/x \geq 416$ and $C \geq (\log 416)/(416^c)$, which can be applied to get our final bound.

□

D.2. Proof of Theorem 2

Proof. Privacy follows by instantiation of REPORTMAX with queries $V(k)$ for $k \in \{\lceil \gamma n \rceil, \ldots, \lfloor (1-\gamma)n \rfloor\}$, which have sensitivity $\Delta(V) = 1/(\gamma n)$, with the observation that noise parameter $2\Delta(V)/\epsilon$ suffices for non-monotonic statistics. We include a proof for completeness.

Fix any two neighboring databases X, X' that differ on index t. For any $k \in \{\lceil \gamma n \rceil, \ldots, \lfloor (1-\gamma)n \rfloor\}$, denote the respective rank statistics as $V(k)$ and $V'(k)$. By the definition of $V(k)$, we have

$$|V(k) - V'(k)| = \begin{cases} \frac{1}{k(n-k)} \left| \sum_{j=k+1}^{n} \mathbb{I}(x_t > x_j) - \mathbb{I}(x'_t > x_j) \right| \leq \frac{1}{k} & \text{if } t \leq k \\ \frac{1}{k(n-k)} \left| \sum_{i=1}^{k} \mathbb{I}(x_i > x_t) - \mathbb{I}(x_i > x'_t) \right| \leq \frac{1}{n-k} & \text{if } t > k, \end{cases}$$

and it follows that $\Delta(V) = 1/(\gamma n)$.

Next, for a given $1 \leq t \leq n$, fix Z_{-t}, a draw from $[\text{Lap}(2/\gamma\epsilon n)]^{n-1}$ used for all the noisy rank statistics values except the lth one. We will bound from above and below the ratio of the probabilities that the algorithm outputs $\tilde{k} = t$ on inputs X and X'. Define the minimum noisy value in order for t to be selected with X:

$$Z_t^* = \min\{Z_t : V(t) + Z_t > V(k) + Z_k \quad \forall k \neq t\}$$

For all $k \neq t$ we have

$$V'(t) + \Delta(V) + Z_t^* \geq V(t) + Z_t^* > V(k) + Z_k \geq V'(k) - \Delta(V) + Z_k.$$

Hence, $Z'_t \geq Z_t^* + 2\Delta(V)$ ensures that the algorithm outputs t on input X', and the theorem follows from the following inequalities for any fixed Z_{-t}, with probabilities over the choice of $Z_t \sim \text{Lap}\,(2/(\gamma\epsilon n))$.

$$\Pr[\tilde{k} = t \mid X', Z_{-t}] \geq \Pr[Z'_i \geq Z_i^* + 2\Delta(V) \mid Z_{-t}] \geq e^{-\epsilon} \Pr[Z_t \geq Z_i^* \mid Z_{-t}] = e^{-\epsilon} \Pr[\tilde{k} = t \mid X, Z_{-i}]$$

□

D.3. Proof of Theorem 3

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 $\Pr_{x \sim P_0, y \sim P_1}[x > y] < 1/2$.

Proof. We will show that for $\tilde{k} = \text{argmax}\,\{V(k) + Z_k\}$ and α as in the theorem statement,

$$\Pr[\left|\tilde{k} - k^*\right| > \alpha] \leq \sum_{k:|k-k^*|>\alpha} \Pr[V(k) + Z_k > V(k^*) + Z_{k^*}] \leq \beta$$

by showing that $V(k) - V(k^*)$ is subgaussian as in Theorem 1, and we will additionally show that the Laplace noise does not introduce too much additional error. For the algorithm to output an incorrect $\tilde{k}$, it must either be the case that the statistic $V(k)$ is nearly as large as $V(k^*)$ because of the randomness of the data points, or that Z_k is much larger than Z_{k^*}. For each value of k, we choose a threshold t_k increasing in $|k - k^*|$ specifying how much to tolerate bad Laplace noise versus bad data, and we bound the probability that the algorithm outputs k as follows:

$$\Pr[V(k) + Z_k > V(k^*) + Z_{k^*}] \leq \Pr[V(k^*) - V(k) < t_k] + \Pr[Z_k - Z_{k^*} > t_k] \tag{4}$$

Setting $t_k = (a - 1/2)|k - k^*|/(2n)$, we can bound the first term as in Theorem 1 using Theorem 10 as follows:

$$\begin{aligned} \Pr[V(k) - V(k^*) > -t_k] &= \Pr\left[V(k) - V(k^*) - \mathbb{E}[V(k) - V(k^*)] > \left(a - \frac{1}{2}\right) \frac{|k - k^*|}{2n}\right] \\ &\leq \exp\left(-\frac{\gamma^4 \left(a - \frac{1}{2}\right)^2 |k - k^*|}{26}\right). \end{aligned}$$

We bound the second term of (4) by analyzing the Laplace noise directly.

$$\begin{aligned}\Pr[Z_k - Z_{k^*} > t_k] &\leq \Pr\left[2\left|\text{Lap}(2/(\epsilon\gamma n))\right| > \left(a - \frac{1}{2}\right)\frac{|k-k^*|}{2n}\right] \\ &\leq \exp\left(-\frac{\left(a-\frac{1}{2}\right)\epsilon\gamma|k-k^*|}{8}\right)\end{aligned}$$

We complete the proof by bounding the probability of any incorrect $\tilde{k}$ such that $\left|\tilde{k} - k^*\right| > \alpha$ by β.

$$\begin{aligned}\Pr\left[\left|\tilde{k} - k^*\right| > \alpha\right] &\leq 2\sum_{k:|k-k^*|=\alpha}^{n} \exp\left(-\frac{\gamma^4\left(a-\frac{1}{2}\right)^2|k-k^*|}{26}\right) + \exp\left(-\frac{\left(a-\frac{1}{2}\right)\epsilon\gamma|k-k^*|}{8}\right) \\ &\leq \frac{2\exp\left(-\frac{\gamma^4}{26}\left(a-\frac{1}{2}\right)^2\alpha\right)}{1-\exp\left(-\frac{\gamma^4}{26}\left(a-\frac{1}{2}\right)^2\right)} + \frac{2\exp\left(-\frac{\epsilon\gamma}{8}\left(a-\frac{1}{2}\right)\alpha\right)}{1-\exp\left(-\frac{\epsilon\gamma}{8}\left(a-\frac{1}{2}\right)\right)} \\ &\leq \beta\end{aligned}$$

We bound each term above by $\beta/2$. Rearranging shows that our accuracy result will hold for

$$\begin{aligned}\alpha \geq \max\Bigg\{&\frac{26}{\gamma^4\left(a-1/2\right)^2}\left(\log\frac{4}{\beta} + \log\frac{1}{1-\exp\left(-\frac{\gamma^4}{26}\left(a-\frac{1}{2}\right)^2\right)}\right), \\ &\frac{8}{\epsilon\gamma\left(a-1/2\right)}\left(\log\frac{4}{\beta} + \log\frac{1}{1-\exp\left(-\frac{\epsilon\gamma}{8}\left(a-\frac{1}{2}\right)\right)}\right)\Bigg\}.\end{aligned}$$

We achieve our final bound by simplifying the above expression as follows. For the first term, we observe that $\gamma < 1/2, a < 1$ implies $x = \gamma^4(a-1/2)^2/26 \leq 1/1664$, and for small x we have $\log(1/(1-\exp(-x))) \leq 2\log(1/x)$. For any $c > 0$, we have $\log(1/x) \leq C(1/x)^c$ for any $1/x \geq 1664$ and $C \geq (\log 1664)/(1664^c)$, which can be applied to get our final bound. For the second term, we observe that $x = \epsilon\gamma(a-\frac{1}{2})/8 \leq \epsilon/32$. When ϵ is small and the corresponding $x \leq 4/5$, we have $\log(1/(1-\exp(-x))) \leq 2\log(1/x)$, and for any $c > 0$, we have $\log(1/x) \leq C(1/x)^c$ for any $1/x \geq 5/4$ and $C \geq (\log 4/5)/((4/5)^c)$. When ϵ is large and the corresponding $x > 4/5$, we have $\log(1/(1-\exp(-x))) \leq \log 2$, which can be incorporated into the constant in our final bound.

□

D.4. Proof of Theorem 4

Proof. By Theorem 8, ABOVETHRESHOLD is ϵ-differentially private, and by Theorem 2, the statistics $V(k)$ and $U(k)$ have sensitivity $2/n$. Also by Theorem 2, PNCPD is ϵ-differentially private. Thus the algorithm ONLINEPNCPD is simply ABOVETHRESHOLD instantiated with privacy parameter $\epsilon/2$, composed with PNCPD also instantiated with privacy parameter $\epsilon/2$. By Basic Composition (Theorem 6), ONLINEPNCPD(X, n, ϵ, γ) is ϵ-differentially private. □

D.5. Proof of Theorem 5

Proof. First, we find an interval $[T_L, T_U]$ for the threshold T that ensures that the algorithm neither calls PNCPD before the true change-point has occurred nor fails to call PNCPD on the window containing k^* somewhere in the middle $(1-2\gamma)n$ data points.

For now we will ignore the error from ABOVETHRESHOLD, and use T'_L, T'_U to denote the desired thresholds ignoring this additional source of noise. For ease of notation and reindexing, we define $U(k) = V(k)$ when $V(k)$ is computed over database $X = \left\{x_{k-n/2+1}, \ldots x_{k+n/2}\right\}$ for the Mann-Whitney test statistic $V(\cdot)$ as defined in Equation (1).

Thus we aim to find a range $[T'_L, T'_U]$ such that

$$\Pr[U(k) > T'_L | X_{k-n/2+1}, \ldots X_{k+n/2} \sim P_0] \leq \frac{\beta}{8(k^* - n/2)}, \tag{5}$$

$$\Pr[U(k) < T'_U | X_{k-n/2+1}, \dots X_k \sim P_0,\ X_{k+1}, \dots X_{k+n/2} \sim P_1] \leq \frac{\beta}{8}. \tag{6}$$

Condition (5) means that after taking a union bound over all the windows that do not contain k^*, the probability that ABOVETHRESHOLD raises the alarm on the window that does not contain the true change point k^* does not exceed $\beta/8$. Condition (6) means that on the window containing the true change-point k^* in the center of the window, ABOVETHRESHOLD will fail to raise the alarm with probability at most $\beta/8$.

It will be helpful to have high probability bounds that the test statistics $U(k)$ are close to their means. Using McDiarmid's Inequality (Theorem 10) we can obtain that for any $k > n$

$$\Pr[U(k) - \mathbb{E}[U(k)] > t] \leq \exp(-t^2 n/2), \tag{7}$$

$$\Pr[U(k) - \mathbb{E}[U(k)] < -t] \leq \exp(-t^2 n/2) \tag{8}$$

Using these bounds, we will first find T'_L. Note that Condition (5) on T'_L considers the setting where all points in the current window are drawn from P_0. Under this condition, $\mathbb{E}[U(k)] = 1/2$. Then by plugging in $t = T'_L - 1/2$ into Inequality (7), we get the following expression:

$$\Pr\left[U(k) \geq T'_L | X_{k-n/2+1}, \dots X_{k+n/2} \sim P_0\right] \leq \exp\left(-\frac{n}{2}\left(T'_L - \frac{1}{2}\right)^2\right)$$

Setting the right hand side of this to less than or equal to $\frac{\beta}{8(k^*-n/2)}$ and solving for T'_L gives the following lower bound, which satisfies Condition (5):

$$T'_L = \frac{1}{2} + \sqrt{\frac{2}{n}\log\left(\frac{8(k^* - n/2)}{\beta}\right)}.$$

Next we find the upper bound T_U. Note that Condition (6) on T'_U considers the setting where the first $n/2$ points in the window are drawn from P_0 and the remaining $n/2$ points are drawn from P_1. Under this condition, $\mathbb{E}[U(k)] = a$. Then plugging $t = a - T'_U$ in Inequality (8) and using Condition (6), we get the following bound:

$$\Pr[U(k) \leq T'_U | X_, \dots X_{n/2} \sim P_0,\ X_{n/2+1}, \dots X_n \sim P_1] \leq \exp\left(-(a - T'_U)^2 n/2\right) \leq \frac{\beta}{8}.$$

Solving this for T'_U gives the following Inequality which satisfies Condition (6):

$$T'_U \leq a - \sqrt{\frac{2}{n}\log\left(\frac{8}{\beta}\right)}.$$

We now return to account for the error from ABOVETHRESHOLD. To ensure that this error does not cause a window to be called before the true change-point and also does not skip the window with the true change-point, we require the following conditions to both hold with probability $\frac{\beta}{4}$

$$\text{For } T \geq T_L,\ \ U_k < T - \alpha' \text{ when } k < k^*$$
$$\text{For } T \leq T_U,\ \ U_{k^*} > T + \alpha'$$

Thus we obtain that the new interval for T is $[T_L, T_U]$, where $T_L = T'_L + \alpha'$, and $T_U = T'_U - \alpha'$. If both those conditions hold then for $\alpha' = \frac{32\log(8(k^*-n/2)/\beta)}{n\epsilon}$, ABOVETHRESHOLD will identify the window which contains the true change point with probability $(1 - \beta/4)$ by Theorem 9. Taking a union bound over the failure probabilities of Conditions (5) and (6), and the statement above, we can see that ONLINEPNCPD will call PNCPD on the right window except with small probability $\beta/2$.

Finally, we can use the accuracy guarantees of PNCPD to show that conditioned on raising an alarm in the correct window, we are likely to output an estimate $\hat{k}$ that is close to the true change-point k^*. Slightly more careful accounting

is needed here, because conditioning on raising an alarm and calling PNCPD, the data points in the chosen window are no longer distributed according to the change-point model. Let $W(k)$ denote the event that ONLINEPNCPD calls PNCPD$(\{x_{k-n/2+1+\gamma n}, \dots x_{k+n/2+\gamma n}\}, \epsilon/2, \gamma)$ on the window centered at k. Then

$$\begin{aligned}
\Pr\left[\left|\tilde{k}-k^*\right|>\alpha\right] &= \sum_{k>n/2} \Pr\left[W(k)\cap\{|\tilde{k}-k^*|>\alpha\}\right] \\
&\le \sum_{k\notin(k^*-n/2,k^*]} \Pr\left[W(k)\right] + \sum_{k\in(k^*-n/2,k^*]} \Pr\left[W(k)\cap\left\{\left|\tilde{k}-k^*\right|>\alpha\right\}\right] \\
&\le \frac{\beta}{2} + \frac{n}{2}\Pr\left[\text{PNCPD fails}\right] < \beta
\end{aligned}$$

To achieve the inequality above, the probability of PNCPD fails to report the change point within the α-window around k^* has to be bounded by β/n. Thus by Theorem 3 we set the error to be,

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

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

We have proved the theorem, but we should also show that the window $[T_L, T_U]$ is non-empty, and there exists a good range in which to choose the threshold T. The condition that $T_L < T_U$ is equivalent to,

$$a-\frac{1}{2} > \sqrt{\frac{2}{n}\log\left(\frac{8(k^*-n/2)}{\beta}\right)} + \sqrt{\frac{2}{n}\log\left(\frac{8}{\beta}\right)} + \frac{64\log(8(k^*-n/2)/\beta)}{n\epsilon}. \tag{9}$$

We can simplify Inequality (9) as,

$$\begin{aligned}
&\sqrt{\frac{2}{n}\log\left(\frac{8(k^*-n/2)}{\beta}\right)} + \sqrt{\frac{2}{n}\log\left(\frac{8}{\beta}\right)} + \frac{64\log(8(k^*-n/2)/\beta)}{n\epsilon} \\
&\quad< \sqrt{\frac{2}{n}\log\left(\frac{8k^*}{\beta}\right)} + \sqrt{\frac{2}{n}\log\left(\frac{8}{\beta}\right)} + \frac{64\log(8k^*/\beta)}{n\epsilon} < a-\frac{1}{2}.
\end{aligned}$$

Finally, solving the right hand side for n, we find the following bound on n that satisfies Inequality (9).

$$n > \frac{1}{(a-1/2)^2}\left(\sqrt{2\log\left(\frac{8k^*}{\beta}\right)} + \sqrt{2\log\left(\frac{8}{\beta}\right)} + \frac{64}{\epsilon}\log\left(\frac{8k^*}{\beta}\right)\right)^2.$$